

Lawrence Berkeley National Laboratory

LBL Publications

Title

Dynamic Transmission Line Switching Amid Wildfire-Prone Weather Under Decision-Dependent Uncertainty

Permalink

<https://escholarship.org/uc/item/6c76x6h3>

Journal

INFORMS Journal on Computing

ISSN

0899-1499

Authors

Estrada-Garcia, Juan-Alberto

Jiang, Ruiwei

Moreira, Alexandre

Publication Date

2026

DOI

10.1287/ijoc.2025.1210

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Dynamic Transmission Line Switching Amidst Wildfire-Prone Weather Under Decision-Dependent Uncertainty

Juan-Alberto Estrada-Garcia*

Ruiwei Jiang[†]

Alexandre Moreira[‡]

Abstract

During dry and windy seasons, environmental conditions significantly increase the risk of wildfires, exposing power grids to disruptions caused by transmission line failures. Wildfire propagation exacerbates grid vulnerability, potentially leading to prolonged power outages. To address this challenge, we propose a multi-stage optimization model that dynamically adjusts transmission grid topology in response to wildfire propagation, aiming to develop an optimal response policy. By accounting for decision-dependent uncertainty, where line survival probabilities depend on usage, we employ distributionally robust optimization to model uncertainty in line survival distributions. We adapt the stochastic nested decomposition algorithm and derive a deterministic upper bound for its finite convergence. To enhance computational efficiency, we exploit the Lagrangian dual problem structure for a faster generation of Lagrangian cuts. Using realistic data from the California transmission grid, we demonstrate the superior performance of dynamic response policies against two-stage alternatives through a comprehensive case study. In addition, we construct easy-to-implement policies that significantly reduce computational burden while maintaining good performance in real-time deployment.

Keywords: Line Switching, Wildfire, Multi-stage Optimization, Decision-dependent Uncertainty, Dual Dynamic Integer Programming

1 Introduction

Transmission grids are crucial infrastructures for societies worldwide. These grids occupy large physical spaces and are distributed in both urban and rural environments, interacting with risk factors to their continuous operation such as tall vegetation, buildings, and vehicles. Risk factors in environments can cause physical damages to key system components, therefore posing a major challenge for transmission grid operators (Rathor and Saxena, 2020). Among these risk factors, wildfires have raised major concerns due to their reported impacts on multiple regions of the world and have increased frequency in recent decades (Halofsky et al., 2020). Wind, temperature, and humidity have significant influence in wildfires ignition and propagation, but their occurrence is difficult to predict (Kondylatos et al., 2022). Wildfires pose significant challenges to transmission grids, because they not only break down infrastructure (e.g., towers and poles; Dian et al. (2019)) but also decrease the conductor ampacity (capacity) of transmission lines (Choobineh and Mohagheghi, 2016), creating outages and cascading impacts. An example is the 2017 Thomas wildfire in California, where increased temperatures, ash, and fire decreased the operative capabilities of power lines. As a result, the lines were forced to shut down, while other lines increased power flow beyond their nominal capacity, causing power outages (Al Saeed and Nazaripouya, 2022). Likewise,

*Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI. juanest@umich.edu

[†]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI. ruiwei@umich.edu

[‡]Lawrence Berkeley National Laboratory, Berkeley, CA. amoreira@lbl.gov

forest wildfires have caused a series of accidents related to power system in China. In February 2010, the Guizhou province power grid experienced 44 line-tripping accidents, most of which were attributed to phase-to-ground faults in wildfire-prone conditions (Pu et al., 2015).

These incidents result in substantial repair costs and widespread power outages, leaving communities vulnerable during emergencies. The increasing frequency and intensity of wildfires due to climate change require robust strategies to protect power systems and ensure uninterrupted service.

One approach to mitigating the wildfire impacts is transmission line switching, which reconfigures the topology of transmission grids and reroutes the electricity flow around wildfire-prone regions (Moreira et al., 2024). Line switching is considered an important measure in response to events in power systems. For example, Nagarajan et al. (2016) suggests installing switches on transmission lines to enhance grid resiliency, Bayani and Manshadi (2023) applies line switching to hedge against wildfire events, and the California utility San Diego Gas & Electric installed line switches to better mitigate wildfire impacts (Udren et al., 2022). When combined with other planning and restoration strategies, line switching has been shown to significantly enhance the resilience of power grids to the threats of extreme events including wildfires (Abdelmalak and Benidris, 2022; Wang et al., 2022).

Despite its wide applications, the modeling of line switching in wildfire-prone weather poses challenges. First, as wildfires decrease transmission line capacities, the probability of a line failure depends not only on environmental factors (temperature, humidity, wildfires in the vicinity, etc.) but also on the magnitude of power flow on the line, rendering the realization of line failures associated with *decision-dependent uncertainty* (DDU; Muhs et al. (2020); Moreira et al. (2024)). Stochastic programs with DDU are computationally intractable (Gupta and Grossmann, 2011; Li and Grossmann, 2021; Ryu and Jiang, 2025), mainly because (i) an accurate model of DDU demands estimating the distribution of uncertainty with respect to *all* possible decisions, and (ii) DDU usually results in a nonconvex optimization model that is difficult to solve, especially for large-scale engineering systems such as transmission grids. Second, the dynamic and stochastic nature of wildfires adds complexity to decision making, as operators must consider the evolving threats. Traditional two-stage models (see, e.g., Huang et al. (2017); Piancó et al. (2024)) to this problem can provide feasible switching plans, but fall short in capturing the dynamic states of wildfire propagation. In other words, two-stage line switching plans cannot be adjusted according to the most up-to-date wildfire states and hence are suboptimal.

To address these modeling challenges, in this paper we propose a multi-stage distributionally robust optimization (DRO) model. Specifically, we model the (line failure) DDU by adopting an ambiguity set of probability distributions, wherein the probability of a line failure depends parametrically on the magnitude of power flow passing through the line. This modeling choice waives the (burdensome) need to estimate the line failure probability for each power flow magnitude and, in addition, alleviates the potential misspecification of the line failure probabilities. This framework ensures that the ensuing line switching policy remains effective across a range of plausible scenarios, enhancing the robustness and reliability of power system operations under wildfire threats. Furthermore, we characterize the wildfire propagation using a scenario tree, which models the dynamic range and severity of the wildfire and accounts for its temporal and spatial variability. To mitigate the computational challenges arising from the large number of decision and state variables, we adapt the stochastic nested decomposition (SND) algorithm (Zou et al., 2018, 2019; Yu and Shen, 2022), derive deterministic upper bound for its finite convergence, and produce tight Lagrangian cuts by reusing past Lagrangian multipliers.

Our contributions in this work are fourfold. Firstly, we propose one of the first dynamic topology optimization models amidst wildfire risks as a multi-stage DRO formulation with DDU. Secondly, to solve this problem, we extend the SND algorithm to solve our DRO model with DDU. We leverage the binary state variables to compute a deterministic upper bound for the optimal cost-to-go with provable convergence. In addition, we produce tight Lagrangian cuts by reusing past Lagrangian multipliers, which speeds up the computation of SND significantly. Thirdly, leveraging the optimal dynamic policy produced by our

DRO model, we construct two easy-to-implement policies that alleviate the demand of solving large-scale optimization formulations in real-time deployment. Finally, we demonstrate the performance of our model and solution approaches through a realistic transmission grid in California and its wildfire data. We compare the solutions and policies obtained by our proposed multi-stage framework to a two-stage benchmark in terms of line switching decisions and their performance on reducing load shedding.

The remainder of the paper is organized as follows. Section 2 reviews relevant research and positions our work in the context of recent advancements. Section 3 presents the multi-stage DRO model with DDU. Section 4 extends the SND algorithm to solve the DRO model. Section 5 presents the numerical results, before we discuss and conclude in Section 6. All technical proofs are relegated to the appendix and additional numerical results are reported in an online repository (Estrada-Garcia et al., 2025).

2 Literature Review

With its long history of study, power system optimization problems (such as unit commitment and line switching) continue being challenging to solve, especially when considering large-scale systems and stochastic elements (Mohseni-Bonab et al., 2022). In particular, transmission grids are vulnerable to failures caused by random events such as natural disasters and wildfires (Sayarshad and Ghorbanloo, 2023). As a response, various planning or operational strategies, such as distributed generation (Mohagheghi and Rebennack, 2015), line switching (Fisher et al., 2008), and de-energizing (Yang et al., 2024a), have been proposed to enhance the grid resiliency. In this context, two-stage models have been widely applied, e.g., with first-stage decisions pertaining to line switching and network configuration before any failures take place and second-stage decisions pertaining to post-failure energy re-dispatch and load shedding (see, e.g., Nguyen et al., 2020; Mohseni-Bonab et al., 2022; Jalilpoor et al., 2022).

The performance of stochastic optimization approaches can suffer when uncertainty cannot be accurately modeled. To remedy this, two-stage distributionally robust extensions of these approaches have been proposed to increase the degree of robustness (see, e.g., Zhang et al., 2017; Moreira et al., 2024; Piancó et al., 2024). In contrast to these works, we consider a multi-stage model to adjust the grid topology dynamically, based on the most up-to-date wildfire propagation states.

In Table 1, we position our work with respect to the relevant state-of-the-art. To the best of our knowledge, multi-stage models that hedge against wildfire risks have been scarce to date. Within this literature, our work is most related to (Yang et al., 2024a,b), which considered stochastic optimization models for de-energizing grid components and achieving grid resiliency. Notably, Yang et al. (2024a,b) considered “endogenous” fires, which originate from random component failures but are independent of power system operations. For example, in their setting, a de-energized component can still fail, ignite, and propagate a fire. This renders their endogenous fires *de facto* decision-independent uncertainty (DIU). In addition, (Zou et al., 2018) considered a multi-stage stochastic optimization model with DIU for dynamic unit commitment. In contrast to Yang et al. (2024a,b); Zou et al. (2018), our model adopts DRO for modeling the line failure DDU and it also allows full flexibility for topology reconfiguration (i.e., both opening and closing lines).

Extending the seminal work on stochastic dual dynamic programming (Pereira and Pinto, 1991), Zou et al. (2019) proposed the stochastic dual dynamic integer programming (SDDiP) algorithm to obtain optimal mixed-integer dynamic policies and applied it to multi-stage unit commitment with demand uncertainty (Zou et al., 2018). Most applications of SDDiP assume stage-wise independent uncertainties for more efficient computation. In contrast, as we model an evolving wildfire process that samples the next wildfire state based on the current state, our model waives this assumption and addresses a more general, stage-wise *dependent* scenario tree. Accordingly, we adopt the SND algorithm (Zou et al., 2019), which generalizes SDDiP.

The convergence of the SND algorithm relies on (i) statistical upper bounds of the optimal value and (ii) cutting planes that iteratively refine the lower approximation of the cost-to-go function in each stage. For

Table 1: Comparison of relevant works in literature.

Work	Application	Uncertainty	Type	Horizon	Objective
Mohagheghi and Rebennack (2015)	Distributed generation	Failure	DIU	Two-stage	RO
Zou et al. (2018)	Unit commitment	Demand	DIU	Multi-stage	SO
Yang et al. (2024a)	Component de-energizing	Failure	DIU	Two-stage	SO
Yang et al. (2024b)	Component de-energizing	Failure	DIU	Multi-stage	SO
Nguyen et al. (2020)	Distributed generation	Failure	DIU	Two-stage	SO
Hosseini and Parvania (2020)	Line switching	–	–	Single-stage	DET
This work	Line switching	Failure	DDU	Multi-stage	DRO

DIU: Decision-independent uncertainty, DET: Deterministic, RO: Robust, SO: Stochastic, DRO: Distributionally

robust optimization.

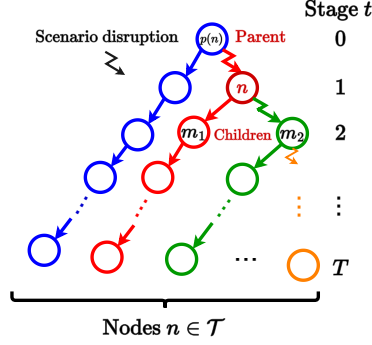


Figure 1: Scenario tree $\mathcal{T}$ of a general discrete stochastic process.

(i), statistical upper bounds in Zou et al. (2019) become inapplicable for our DRO model with DDU. As an alternative, we propose deterministic upper bounds, enabling an accurate evaluation of the optimality gap. To improve (ii), existing works have proposed various approaches to generate stronger Lagrangian cuts for binary state variables (see Yang et al. (2024b); Chen and Luedtke (2022) and the references therein). We leverage this body of work to derive cuts for general (mixed-integer) state variables. In addition, we propose an algorithm to reuse Lagrangian multipliers in past SND iterations and generate tight Lagrangian cuts faster.

3 Distributionally Robust Dynamic Line Switching

We formulate a model to prescribe a dynamic line switching policy for a transmission grid. We describe the scenario tree in Section 3.1 and present the model in Section 3.2. We derive a deterministic representation for this model and characterize the worst-case failure distribution in Section 3.3.

3.1 Scenario tree

We model the wildfire propagation using a scenario tree $\mathcal{T}$, whose nodes are organized in T levels/stages $\mathcal{S}_t$ such that node 1 denotes the root node, $\mathcal{T} = \bigcup_{t=1}^T \mathcal{S}_t$, $\mathcal{S}_r \cap \mathcal{S}_t = \emptyset$ whenever $r \neq t$, and each node $n \in \mathcal{S}_t$ represents a possible wildfire state in time unit t (see Fig. 1). In addition, each non-leaf node $n \in \mathcal{T}$ branches out to a set $C(n)$ of children nodes, and for each $m \in C(n)$, p_{nm} represents the (conditional) probability of branching to node m from n . Conversely, for each non-root node $n \in \mathcal{T}$, we denote by $p(n)$ its parent node, i.e., $n \in C(p(n))$. By convention, we define $p(1) := 0$ and $C(n) := \emptyset$ for all leaf nodes n , and we call the path of wildfire state realizations from the root node to a leaf node n a “scenario” and denote it by $\Pi(n)$.

The scenario tree $\mathcal{T}$ is sufficiently general for modeling any finite and discrete stochastic process and for

approximating a continuous one. For example, to model the wildfire propagation in a region, we can divide the region into a finite set Σ of cells and denote the state of each cell $\sigma \in \Sigma$ in time unit t using an integer $w_{\sigma t}^{\text{fire}}$, with $w_{\sigma t}^{\text{fire}} = 0, 1, 2$ representing the cell being unburned, burning, and burned, respectively. This way, the state of the wildfire across all cells in time t can be encoded by a vector $w_t^{\text{fire}} := \{w_{\sigma t}^{\text{fire}} : \forall \sigma \in \Sigma\}$ and $\mathcal{T}$ can model the wildfire propagation by appropriately designating p_{nm} based on $\mathbb{P}(w_{t+1}^{\text{fire}} | w_t^{\text{fire}})$. Other data-driven approaches have also been proposed to construct scenario trees (Heitsch and Römisch, 2009; Oliveira et al., 2015; Yang et al., 2024b). In this paper, we generate scenario trees based on historical wildfire perimeter data (see Appendix A.2 for details).

3.2 A multi-stage DRO model with DDU

We start by following Hedman et al. (2010) to formulate a deterministic line switching model for every node $n \in \mathcal{T}$. For ease of exposition, we summarize the nomenclature for all sets, parameters, and decision variables in Table 6 and the detailed model in formulation (12) of Appendix A.1. This model takes in as parameters the current line availability states $\tilde{\mathbf{a}}_n$, plus the line switching decisions $\mathbf{z}_{p(n)}$, the power flows $(\mathbf{o}_{p(n)}, \mathbf{f}_{p(n)})$, and the power generation amounts $\mathbf{p}_{p(n)}$ in the parent node $p(n)$. Then, it seeks to minimize the generator production cost plus the load shedding cost, subject to phase angle limits, flow balance for each bus, and transmission line and generator capacity. We define a vector $\mathbf{x}_n := [\mathbf{z}_n, \mathbf{o}_n, \mathbf{f}_n, \mathbf{p}_n]$ to group variables representing switching decisions, power flows, and power generation amounts; and we define another vector $\mathbf{y}_n := [\theta_n, \Delta_n]$ to group variables representing voltage angles and load shedding amounts. Vectors $\mathbf{x}_n$ and $\mathbf{y}_n$ are different in that $\mathbf{x}_n$ (*inter-stage* variables) can be passed from node n to its children, while $\mathbf{y}_n$ (*intra-stage* variables) only affects the operations within node n . Using these abstract notations, we express the deterministic line switching model in a compact form:

$$\min_{\mathbf{x}_n, \mathbf{y}_n} \{f_n(\mathbf{x}_n, \mathbf{y}_n, \tilde{\mathbf{a}}_n) : A_n \mathbf{x}_n + W_n \mathbf{y}_n + C_n \mathbf{x}_{p(n)} + D_n \tilde{\mathbf{a}}_n \geq h_n\},$$

where f_n represents the objective function, and matrices A_n, W_n, C_n, D_n and right-hand side vector h_n encode the coefficients of constraints (see Appendix A.1 for nomenclature in Table 6 and a detailed formulation in (12)).

Then, we propose a multi-stage DRO model to prescribe an optimal line switching policy under line failure DDU. We consider two sources of uncertainties: firstly, the scenario tree $\mathcal{T}$ models the (decision-independent) wildfire propagation uncertainty; and secondly, we model the (decision-dependent) line failure uncertainty through an ambiguity set, which depends on the latest wildfire state and the power flow on the line. Accordingly, the model is a dynamic program that, at each node $n \in \mathcal{T}$, seeks to balance the immediate cost $f_n(\mathbf{x}_n, \mathbf{y}_n, \tilde{\mathbf{a}}_n)$ and the expected cost-to-go with respect to the wildfire propagation distribution and the worst-case distribution of line failures. Specifically, the model evaluates $Q_1(\mathbf{x}_0, \mathbf{1})$ with a pre-specified system state $\mathbf{x}_0$ before the planning horizon starts and zero line failure ($\tilde{\mathbf{a}}_1 = \mathbf{1}$ almost surely), where

$$Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) := \min_{\mathbf{x}_n, \mathbf{y}_n} f_n(\mathbf{x}_n, \mathbf{y}_n, \tilde{\mathbf{a}}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] \quad (1a)$$

$$\text{s.t. } A_n \mathbf{x}_n + W_n \mathbf{y}_n + C_n \mathbf{x}_{p(n)} + D_n \tilde{\mathbf{a}}_n \geq h_n \quad (1b)$$

for all nodes $n \in \mathcal{T}$. We model the line failure DDU using a DRO model, as opposed to a classical stochastic program, to take advantage of existing parametric prediction of line failure probability as a function of power flow magnitude (see, e.g., Muhs et al. (2020); Píancó et al. (2024)) and to waive the need to estimate such probability for each possible magnitude $\mathbf{f}_n$. In particular, we follow Muhs et al. (2020); Píancó et al. (2024) to define a moment ambiguity set

$$\mathcal{P}_m(\mathbf{x}_n) = \{\mathbb{P} \in \mathcal{P}(\mathcal{A}_m) : \mathbb{E}_{\mathbb{P}} [\tilde{\mathbf{a}}_m] \leq \beta_m \mathbf{x}_n + \gamma_m\},$$

where the line survival probability is bounded by an affine function of $\mathbf{x}_n$ with constant matrix $\beta_m \in \mathbb{R}_{-}^{|\mathcal{L}| \times \dim(\mathbf{x}_n)}$ and vector $\gamma_m \in \mathbb{R}_{+}^{|\mathcal{L}|}$, where $\mathcal{L}$ denotes the set of transmission lines. Note that (β_m, γ_m) depends on m and hence can be calibrated based on the latest wildfire state in node m . In addition, although the definition of $\mathcal{P}_m(\mathbf{x}_n)$ is general and allows the line survival probability to depend on all components of $\mathbf{x}_n$, in this paper we focus on its (negative) correlation with the power flow magnitude $|f_{ln}|$ and the indicator o_{ln} for f_{ln} exceeding the nominal range. Finally, $\mathcal{P}(\mathcal{A}_m)$ denotes the set of all distributions supported on $\mathcal{A}_m$, which is defined through

$$\mathcal{A}_m := \left\{ \tilde{\mathbf{a}}_m \in \{0, 1\}^{|\mathcal{L}|} : e^\top \tilde{\mathbf{a}}_m \geq |\mathcal{L}| - K, \tilde{\mathbf{a}}_m \leq \tilde{\mathbf{a}}_{p(m)} \right\}.$$

In other words, we allow for at most K line failures and assume that a failed line during wildfires will remain dysfunctional throughout the planning horizon.

3.3 Deterministic representation and worst-case distribution

Formulation (1) is challenging to solve directly because the worst-case expectation $\sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)]$ embeds an infinite-dimensional optimization problem with respect to the probability distribution $\mathbb{P}$. We derive a deterministic and finite-dimensional representation.

Proposition 1. *For any fixed $m \in \mathcal{T}$ and $\mathbf{x}_n$, it holds that*

$$\sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] = \min_{\psi_m \geq 0, \phi_m} \psi_m^\top \beta_m \mathbf{x}_n + \psi_m^\top \gamma_m + \phi_m \quad (2a)$$

$$\text{s.t. } \phi_m \geq Q_m(\mathbf{x}_n, \mathbf{a}_m) - \mathbf{a}_m^\top \psi_m, \quad \forall \mathbf{a}_m \in \mathcal{A}_m. \quad (2b)$$

The deterministic representation (2) is a linear program with respect to (ψ_m, ϕ_m) . In addition, it produces a bilinear program when integrated back to formulation (1) (due to the product term $\psi_m^\top \beta_m \mathbf{x}_n$), which can be solved by commercial solvers (e.g., GUROBI) directly. On the other hand, constraints (2b) are exponential in number, which is computationally prohibitive if K is large. Nevertheless, we take advantage of the structure of $Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)$ to efficiently separate these constraints. Indeed, as shown in Section 4, the SND algorithm replaces $Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)$ with an (iteratively improved) lower approximation $\underline{Q}_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)$, which is the pointwise maximum of a set $\mathcal{H}_m$ of affine functions

$$\max_{h \in \mathcal{H}_m} \{(\pi^h)^\top \mathbf{x}_n + (\tau^h)^\top \tilde{\mathbf{a}}_m + \omega^h\}, \quad (3)$$

where $(\pi^h, \tau^h, \omega^h)$ are constants to be generated in the process of the SND algorithm. The next proposition shows that the separation of constraints (2b) is polynomial in $|\mathcal{H}_m|$.

Proposition 2. *If $Q_m(\mathbf{x}_n, \mathbf{a}_m)$ admits a representation in the form of (3), then Algorithm 1 solves the separation problem of constraints (2b), i.e.,*

$$\max_{\mathbf{a} \in \mathcal{A}_m} \{Q_m(\mathbf{x}_n, \mathbf{a}) - \mathbf{a}^\top \psi_m\} \quad (\text{SEP})$$

in time $O(|\mathcal{H}_m| |\mathcal{L}| \log(|\mathcal{L}|))$.

By Proposition 2, we can quickly detect if a given solution $(\hat{\psi}_m, \hat{\phi}_m)$ violates any of constraints (2b). Indeed, we can run Algorithm 1 to obtain an $\mathbf{a}^h$ and then check if $\hat{\phi}_m < Q_m(\mathbf{x}_n, \mathbf{a}^h) - (\mathbf{a}^h)^\top \hat{\psi}_m$ (violation of (2b) with respect to $\mathbf{a}^h$) or otherwise certify that $(\hat{\psi}_m, \hat{\phi}_m)$ satisfies (2b) with respect to all $\mathbf{a} \in \mathcal{A}_m$. Accordingly, we can solve formulation (1) using delayed constraint generation. Specifically, we first solve

Algorithm 1 Separation of constraints (2b)

```

1: for  $h = 1, \dots, |\mathcal{H}_m|$  do
2:   if  $\mathbf{1}^\top \tilde{\mathbf{a}}_{p(m)} = |\mathcal{L}| - K$  then
3:     Set  $\mathbf{a}^h = \tilde{\mathbf{a}}_{p(m)}$ ;
4:   else
5:     Set  $\mathbf{a}^h = \tilde{\mathbf{a}}_{p(m)}$ ;
6:     Find a permutation of set  $\Upsilon := \{j : (\tau^h - \psi_m)_j \leq 0\}$ , denoted by  $(1), \dots, (|\Upsilon|)$ , such that  $(\tau^h - \psi_m)_{(1)} \leq (\tau^h - \psi_m)_{(2)} \leq \dots \leq (\tau^h - \psi_m)_{(|\Upsilon|)}$ ;
7:     Set  $\mathbf{a}_{(j)}^h = 0$  for all  $j = 1, \dots, \min\{|\Upsilon|, \mathbf{1}^\top \tilde{\mathbf{a}}_{p(m)} - |\mathcal{L}| + K\}$ ;
8:   end if
9:   Set  $v^h := (\pi^h)^\top \mathbf{x}_n + (\tau^h - \psi_m)^\top \mathbf{a}^h + \omega^h$ ;
10: end for
11: Return the  $\mathbf{a}^h$  such that  $v^h \geq v^k$  for all  $k \in \mathcal{H}_m$ ;

```

a relaxation of (1) that ignores constraints (2b) to obtain an incumbent solution $(\hat{\mathbf{x}}_n, \hat{\mathbf{y}}_n, \hat{\psi}_m, \hat{\phi}_m)$. Then, we iteratively detect the violation of any of these constraints using Algorithm 1 and incorporate all violated constraints back into the relaxation till no violation can be found. At that point, the latest incumbent is then an optimal solution to (2). Let set $\mathcal{A}_m^*$ collect all $\mathbf{a}^h$ obtained from Algorithm 1 before an optimal solution is certified. The following proposition derives a worst-case probability distribution for $\tilde{\mathbf{a}}_m$ that attains $\sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)]$.

Proposition 3. For any fixed $m \in \mathcal{T}$ and $\mathbf{x}_n$, it holds that

$$\sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] = \min_{\psi_m \geq 0, \phi_m} \psi_m^\top \beta_m \mathbf{x}_n + \psi_m^\top \gamma_m + \phi_m \quad (4a)$$

$$\text{s.t. } \phi_m \geq Q_m(\mathbf{x}_n, \mathbf{a}_m) - \mathbf{a}_m^\top \psi_m, \quad \forall \mathbf{a}_m \in \mathcal{A}_m^*. \quad (4b)$$

In addition, let λ^h denote the dual optimal solution associated with constraints (4b) and define a probability distribution $\mathbb{P}^*$ with $\mathbb{P}^*\{\tilde{\mathbf{a}}_m = \mathbf{a}^h\} = \lambda^h$ for all $h \in [|\mathcal{A}_m^*|]$. Then, it holds that

$$\sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] = \mathbb{E}_{\mathbb{P}^*} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)].$$

4 A Stochastic Nested Decomposition (SND) Algorithm

We adapt the SND algorithm to solve the multi-stage DRO model with DDU. We present the SND algorithm in Section 4.1. Furthermore, we provide various computing enhancement strategies, including deterministic upper bounds and faster generation of Lagrangian cuts in Section 4.2.

4.1 Algorithm design

We adapt the SND algorithm for multi-stage stochastic integer program (Zou et al., 2019) to our DRO model. Our Algorithm 2 maintains and iteratively refines a lower approximation $\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ of the cost-to-go function $Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ through cutting planes. A main difference between Zou et al. (2019) and Algorithm 2, however, is that a part of our system state $\tilde{\mathbf{a}}_n$ is DDU. Consequently, the transition of system state is *decision-dependent*. Fortunately, we derived a worst-case distribution of $\tilde{\mathbf{a}}_n$ in Proposition 3 and can then sample from it for state transition.

The deterministic representation of formulation (1), by Proposition 1, is a bilinear program. Although commercial solvers (e.g., GUROBI) can solve bilinear programs directly, they solve MILPs more efficiently

Algorithm 2 SND algorithm for the DRO model with DDU

```

1: Initialization:  $LB \leftarrow -\infty, i \leftarrow 1$ ;
2: while (Stopping Criterion Not Satisfied) do
3:   Sample  $M$  scenarios  $\Omega^i = \{\omega_1^i, \dots, \omega_M^i\}$ ;
4:   [Forward pass]
5:   for  $k = 1, \dots, M$  do
6:     for  $n \in \omega_k^i$  do
7:       Solve formulation (1) with  $Q_m(\cdot, \cdot)$  replaced by  $\underline{Q}_m(\cdot, \cdot)$  and store solution  $(\hat{\mathbf{x}}_n^i, \hat{\mathbf{y}}_n^i)$ ;
8:       Sample  $\tilde{\mathbf{a}}_m^i$  from the worst-case distribution  $\mathbb{P}^*$ ;
9:     end for
10:  end for
11:  [Backward pass]
12:  for  $t = T - 1, \dots, 1$  do
13:    for  $n \in S_t$  do
14:      if  $n \in \omega_k^i$  for some  $k \in \{1, \dots, M\}$  then
15:        for  $m \in C(n)$  do
16:          Add cuts to strengthen the lower approximation  $\underline{Q}_m(\cdot, \cdot)$ ;
17:        end for
18:      end if
19:    end for
20:  end for
21:  [Lower bound]
22:  Evaluate  $Q_1(\mathbf{x}_0, \mathbf{1})$  using formulation (1) with  $n = 1$  and  $Q_m(\cdot, \cdot)$  replaced by  $\underline{Q}_m(\cdot, \cdot)$ ;
23:   $LB \leftarrow Q_1(\mathbf{x}_0, \mathbf{1})$  and  $i \leftarrow i + 1$ ;
24: end while

```

by far. In addition, lower approximations of $Q(\mathbf{x}_n, \tilde{\mathbf{a}}_m)$ by cutting planes are computationally easier to obtain with state variables $\mathbf{x}_n$ being binary, as opposed to being continuous or mixed-integer. As a result, in this paper we approximate all continuous state variables (power generation p_{gn} and power flow f_{ln}) using binary expansions (Owen and Mehrotra, 2002).

Specifically, for each continuous state variable $Z \in [L, U]$, we define auxiliary binary variables $\{z_e \in \{0, 1\}, e \in [E]\}$ and approximate

$$Z \approx L + s \sum_{e=1}^E 2^{e-1} z_e,$$

where s is a pre-specified approximation precision (e.g., $s = 10^{-1}, 10^{-2}$) and $E := \lfloor \log_2 \left(\frac{U-L}{s} \right) \rfloor + 1$. Under this approximation, $\mathbf{x}_n$ becomes (purely) binary variables. Consequently, the bilinear term $\boldsymbol{\psi}_m^\top \beta_m \mathbf{x}_n$ in the deterministic representation of formulation (1) can be linearized using standard McCormick inequalities and we can solve (1) as a MILP efficiently. In Theorem 3 of Appendix A.3, we show that the approximation error, in terms of the optimal value of our DRO model by applying the binary expansion, is *linear* in precision s . Consequently, for the rest of this section, we shall focus on solving the binary expansion approximation using a SND algorithm.

Throughout the algorithm, $\underline{Q}_n(\mathbf{x}_n, \tilde{\mathbf{a}}_m)$ is characterized by the pointwise maximum of affine functions (see, e.g., (3)) and we iteratively update $\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ by adding new cuts in the form

$$\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) \geq \boldsymbol{\pi}^\top \mathbf{x}_{p(n)} + \boldsymbol{\tau}^\top \tilde{\mathbf{a}}_n + \omega \quad (5)$$

to strengthen the lower approximation. In this paper, we follow Zou et al. (2019) and consider the following three families of cuts. To describe them, we rewrite (1) with respect to $\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ as

$$\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) := \min_{\substack{\mathbf{x}_n, \mathbf{y}_n, \\ \mathbf{r}_n, \mathbf{w}_n: (1b)}} f_n(\mathbf{x}_n, \mathbf{y}_n, \mathbf{w}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [\underline{Q}_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)]$$

$$\text{s.t. } \mathbf{r}_n = \mathbf{x}_{p(n)}, \quad (6a)$$

$$\mathbf{w}_n = \tilde{\mathbf{a}}_n, \quad (6b)$$

$$\mathbf{r}_n \in \{0, 1\}^{\dim(\mathbf{x}_{p(n)})}, \mathbf{w}_n \in \{0, 1\}^{\dim(\tilde{\mathbf{a}}_n)}, \quad (6c)$$

where constraints (6a)–(6b) make copies of the state variables $(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$. As both $\mathbf{x}_{p(n)}$ and $\tilde{\mathbf{a}}_n$ are binary and $\underline{Q}_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)$ admits a piecewise linear representation as in (3), formulation (6) is a mixed-binary linear program and strong duality holds when we relax constraints (6a)–(6a) in the Lagrangian manner. More specifically, $\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ equals the optimal value of the following (max-min) Lagrangian relaxation:

$$\begin{aligned} & \max_{\pi, \tau} \min_{\substack{\mathbf{x}_n, \mathbf{y}_n, \\ \mathbf{r}_n, \mathbf{w}_n: (1b), (6c)}} f_n(\mathbf{x}_n, \mathbf{y}_n, \mathbf{w}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [\underline{Q}_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] - \pi^\top (\mathbf{r}_n - \mathbf{x}_{p(n)}) \\ & \quad - \tau^\top (\mathbf{w}_n - \tilde{\mathbf{a}}_n) \\ & = \max_{\pi, \tau} \mathcal{R}_{\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n}(\pi, \tau) := \left\{ \pi^\top \mathbf{x}_{p(n)} + \tau^\top \tilde{\mathbf{a}}_n + \mathcal{L}_n(\pi, \tau) \right\}, \end{aligned} \quad (7)$$

where (π, τ) are dual variables associated with (6a)–(6b), respectively, and

$$\mathcal{L}_n(\pi, \tau) := \min_{\substack{\mathbf{x}_n, \mathbf{y}_n, \\ \mathbf{r}_n, \mathbf{w}_n: (1b), (6c)}} f_n(\mathbf{x}_n, \mathbf{y}_n, \mathbf{w}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [\underline{Q}_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] - \pi^\top \mathbf{r}_n - \tau^\top \mathbf{w}_n. \quad (8)$$

As (7) is an unconstrained convex program for (π, τ) , we can solve (7) by using, e.g., the subgradient algorithm. The strong duality of (7) gives rise to the validity and tightness of Lagrangian cuts.

Proposition 4 (Lagrangian Cuts). *At any binary $(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n)$, inequality (5) holds with*

$$(\pi, \tau) \in \arg \max_{\pi, \tau} \mathcal{R}_{\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n}(\pi, \tau), \quad \omega = \mathcal{L}_n(\pi, \tau).$$

In addition, the cut is tight in the sense that

$$\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) = \pi^\top \hat{\mathbf{x}}_{p(n)} + \tau^\top \hat{\mathbf{a}}_n + \omega. \quad (9)$$

We note that formulation (7) puts no restrictions on (π, τ) and so any (π, τ) produces a valid inequality. A convenient (but not necessarily optimal) choice of (π, τ) is the dual optimal solutions of the linear programming relaxation of (6), producing the strengthened Benders' cuts.

Proposition 5. (Strengthened Benders' Cuts) *For any binary $(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n)$, let (π, τ) represent dual optimal solutions associated with constraints (6a)–(6b) to formulation (6) with $\mathbf{x}_{p(n)} = \hat{\mathbf{x}}_{p(n)}$, $\tilde{\mathbf{a}}_n = \hat{\mathbf{a}}_n$, and no binary restrictions (6c). Then, inequality (5) holds with $\omega = \mathcal{L}_n(\pi, \tau)$.*

A third class of cuts extend the integer optimality cuts for two-stage stochastic integer program (Laporte and Louveaux, 1993), using a global lower bound L_n of $\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$. In computation, L_n can be found, e.g., by solving the LP relaxation of formulation (6).

Proposition 6. (Integer Optimality Cuts) *For any binary $(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n)$ and lower bound L_n of $\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$, inequality (5) holds with $\pi = (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(2\hat{\mathbf{x}}_{p(n)} - \mathbf{1})$, $\tau = (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(2\hat{\mathbf{a}}_n - \mathbf{1})$, and $\omega = \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(\mathbf{1}^\top \hat{\mathbf{x}}_{p(n)} + \mathbf{1}^\top \hat{\mathbf{a}}_n)$.*

We close this section by confirming that Algorithm 2 converges finitely to the global optimum of the DRO model $\underline{Q}_1(\mathbf{x}_0, \mathbf{1})$, which follows from Theorem 2 of Zou et al. (2019) and the tightness of the Lagrangian cuts (see Proposition 4) and the integer optimality cuts (see Proposition 7).

Theorem 1. *Suppose that Algorithm 2 adopts either Lagrangian cuts or integer optimality cuts to strengthen $\underline{Q}_m(\cdot, \cdot)$, then with probability one the forward pass solutions $\{(\hat{\mathbf{x}}_n^i, \hat{\mathbf{y}}_n^i) : n \in \mathcal{T}\}$ converge in a finite number of iterations to global optimal solutions to $\underline{Q}_1(\mathbf{x}_0, \mathbf{1})$.*

4.2 Computational strengthening strategies

We propose the following three strategies to strengthen Algorithm 2.

4.2.1 Deterministic upper bounds

In Algorithm 2, a stopping criterion can be a threshold on either the number of iterations or the improvement in LB. However, neither criterion certifies global optimality (or quantifies suboptimality) of solving the DRO model. As an alternative, we propose a deterministic upper bound UB for $Q_1(\mathbf{x}_0, \mathbf{1})$. Then, we can terminate Algorithm 2 when LB and UB are sufficiently close. To this end, similar to the lower approximations $\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$, we store an upper approximation $\overline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ for each $Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ and update it through

$$\overline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) := \min_{\mathbf{x}_n, \mathbf{y}_n: (1b)} f_n(\mathbf{x}_n, \mathbf{y}_n, \tilde{\mathbf{a}}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [\overline{Q}_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)], \quad (10)$$

for all $n \in \mathcal{T}$ and accordingly $\text{UB} = \overline{Q}_1(\mathbf{x}_0, \mathbf{1})$. We store each $\overline{Q}_n(\cdot, \cdot)$ as a boolean function: for the solutions $(\hat{\mathbf{x}}_{p(n)}^i, \hat{\mathbf{a}}_n^i)$ ever explored in forward passes of Algorithm 2, where $i \in [I]$ is the iteration index and I is the total number of iterations by far, $\overline{Q}_n(\hat{\mathbf{x}}_{p(n)}^i, \hat{\mathbf{a}}_n^i)$ equals the latest optimal value $\hat{u}_n^i$ of the above formulation; and otherwise it equals a large constant $\mathcal{M} \geq \max_{\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n} Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ (we discuss the initialization of $\mathcal{M}$ in Appendix A.6). We represent this function using an auxiliary binary variable $\delta_{p(n)}^i$, which equals 1 if and only if $\mathbf{x}_{p(n)} = \hat{\mathbf{x}}_{p(n)}^i$ for any $i \in [I]$, or otherwise $\delta_{p(n)}^{I+1} = 1$ if $\mathbf{x}_{p(n)}$ differs from all $\hat{\mathbf{x}}_{p(n)}^i$. Likewise, we define an auxiliary binary variable γ_n^i , which equals 1 if and only if $\tilde{\mathbf{a}}_n = \hat{\mathbf{a}}_n^i$, or otherwise $\gamma_n^{I+1} = 1$. This gives rise to the following reformulation by Proposition 1:

$$\begin{aligned} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [\overline{Q}_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] &= \min_{\boldsymbol{\psi}_m \geq 0, \phi_m} \boldsymbol{\psi}_m^\top \boldsymbol{\beta}_m \mathbf{x}_n + \boldsymbol{\psi}_m^\top \boldsymbol{\gamma}_m + \phi_m \\ \text{s.t. } \phi_m &\geq \sum_{i=1}^I \delta_n^i \gamma_m^i \hat{u}_m^i + \mathcal{M}(\delta_n^{I+1} + \gamma_m^{I+1} - \delta_n^{I+1} \gamma_m^{I+1}) \\ &\quad - \mathbf{a}_m^\top \boldsymbol{\psi}_m, \quad \forall (\mathbf{a}_m, \gamma_m) \in \overline{\mathcal{A}}_m, \end{aligned}$$

where $\overline{\mathcal{A}}_m := \{(\mathbf{a}_m, \gamma_m) : \mathbf{a}_m \in \mathcal{A}_m, \sum_{i=1}^{I+1} \gamma_m^i = 1, \gamma_m^i = 1 \Leftrightarrow \mathbf{a}_m = \hat{\mathbf{a}}_m^i, \forall i \in [I]\}$. Similar to the representation (2), the above formulation involves exponentially many constraints. Nevertheless, we can identify the violated constraints and incorporate them only when violated by solving the following MILP for fixed $(\mathbf{x}_n, \delta_n, \boldsymbol{\psi}_m)$:

$$\begin{aligned} \max_{\mathbf{a}_m \in \mathcal{A}_m, \gamma_m \geq 0} \quad & \sum_{i=1}^I \delta_n^i \gamma_m^i \hat{u}_m^i + \mathcal{M}(\delta_n^{I+1} + \gamma_m^{I+1} - \delta_n^{I+1} \gamma_m^{I+1}) - \mathbf{a}_m^\top \boldsymbol{\psi}_m \\ \text{s.t. } \quad & 1 - \|\mathbf{a}_m - \hat{\mathbf{a}}_m^i\|_1 \leq \gamma_m^i \leq 1 - \|\mathbf{a}_m - \hat{\mathbf{a}}_m^i\|_\infty, \quad \forall i \in [I], \end{aligned} \quad (11a)$$

$$\sum_{i=1}^{I+1} \gamma_m^i = 1, \quad (11b)$$

where constraints (11a)–(11b) ensure $(\mathbf{a}_m, \gamma_m) \in \overline{\mathcal{A}}_m$. In lines 13–19 of Algorithm 2, in addition to updating the lower approximation $\underline{Q}_m(\cdot, \cdot)$, we solve formulation (10) with respect to $(\hat{\mathbf{x}}_{p(n)}^i, \hat{\mathbf{a}}_n^i)$ obtained from the latest forward pass and update $\overline{Q}_n(\cdot, \cdot)$ with the ensuing optimal value $\hat{u}_n^i$. We close this section by confirming that this modification of Algorithm 2 yields a series of upper bounds UB, which converge finitely to $Q_1(\mathbf{x}_0, \mathbf{1})$ from above.

Theorem 2. *UB decreases and, with probability one, converges in a finite number of iterations to $Q_1(\mathbf{x}_0, \mathbf{1})$.*

4.2.2 Lagrangian cuts through continuous state variables

We approximate the continuous state variables of $\mathbf{x}_n$, including power flow f_{ln} and power generation p_{gn} , through binary expansion. This generates tight Lagrangian cuts by Proposition 4 but increases the dimension of $\mathbf{x}_n$ as well as the state space. As an alternative, we can generate the Lagrangian cuts through the continuous state variables without applying the binary expansion approximation. Specifically, we generate the Lagrangian cuts by solving formulation (7), but without the binary restrictions (6c) for variables (f_{ln}, p_{gn}) . This leads to weaker Lagrangian cuts because the Lagrangian relaxation of formulation (6) pertaining to continuous $\mathbf{x}_{p(n)}$ does not admit strong duality. Nevertheless, this decreases the dimension of $\mathbf{x}_{p(n)}$ and can speed up the convergence of the lower approximation $\underline{Q}_n(\cdot, \cdot)$. In implementation of Algorithm 2, we can start by incorporating Lagrangian cuts through continuous $\mathbf{x}_n$ and then switch to their binary expansion, e.g., when the Lagrangian cuts cease to improve LB. In Section 5.2 and Table 3, we demonstrate the effectiveness of this strategy.

4.2.3 Faster generation of Lagrangian cuts

The Lagrangian cuts (5) are valid and tight by Proposition 4. However, the search for the coefficients (π, τ) often involves a subgradient algorithm, which iteratively solves MILP (8) and is computationally heavy. In contrast, the integer optimality cuts specified in Proposition 6 are generally considered weaker than the Lagrangian cuts due to the large magnitude of its cut coefficients. Nevertheless, these coefficients admit closed-form expressions and so generating integer optimality cuts is much faster. Our numerical experiments in Table 4 confirm these observations as Algorithm 2 with integer optimality cuts (see I+SB), which incurs significantly more iterations, consistently converges faster than with Lagrangian cuts (see L+SB).

To seek a better trade-off between the strength and efficiency of generating Lagrangian cuts, we make two observations. First, although weaker in general, the integer optimality cut is in fact a *tight Lagrangian cut* at any binary $(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n)$.

Proposition 7. *For any binary $(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n)$, let $\pi = (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(2\hat{\mathbf{x}}_{p(n)} - \mathbf{1})$, $\tau = (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(2\hat{\mathbf{a}}_n - \mathbf{1})$, and $\omega = \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(\mathbf{1}^\top \hat{\mathbf{x}}_{p(n)} + \mathbf{1}^\top \hat{\mathbf{a}}_n)$ as in the integer optimality cut (see Proposition 6). Then, equality (9) holds. In addition, $(\pi, \tau) \in \arg \max_{\pi, \tau} \mathcal{R}_{\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n}(\pi, \tau)$ and $\omega = \mathcal{L}_n(\pi, \tau)$.*

Proposition 7 suggests that, at any binary $(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n)$, there exist multiple tight Lagrangian cuts in the sense of (9) and the integer optimality cut is one of them, but with steep slopes.

Second, the objective function $\mathcal{R}_{\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n}(\cdot, \cdot)$ of formulations (7) remains nearly unchanged in consecutive iterations of Algorithm 2. Indeed, in each iteration Algorithm 2 only updates the lower approximation $\underline{Q}_m(\cdot, \cdot)$ in formulation (8), while the rest of (8) remains the same. This suggests that, when searching for the coefficients (π, τ) of a Lagrangian cut through the subgradient algorithm, Algorithm 2 solves *nearly the same* formulation (7) in consecutive iterations, and it will be a waste of effort if we start each subgradient algorithm from scratch. In contrast, we propose using the integer optimality cut coefficients as an “anchor” and to revise past maximizers for (7) to accelerate the subgradient algorithm. We make this concrete in Algorithm 3. This algorithm starts from the most recent maximizer and checks its optimality to formulation (7) with respect to an updated (but similar) $\mathcal{R}_{\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n}(\cdot, \cdot)$. In case of optimal, we find a tight and strong Lagrangian cut in one shot; and if not optimal, we iteratively move towards the integer optimality cut coefficients to check other past maximizers. This procedure ends with a Lagrangian cut that is at least as strong as the integer optimality cut and we demonstrate its effectiveness in Section 5.2 (see Table 4).

Algorithm 3 Generation of Lagrangian cuts by reusing past cut coefficients

```
1: Input: iteration index  $i \leftarrow 0$ , iteration budget  $I_{\max}$ , past cut coefficients  $\Pi := \{\pi^h, \tau^h\}_{h=1}^H$ , most recent cut coefficients  $(\pi^H, \tau^H)$ , step size  $\alpha$ , neighborhood threshold  $\epsilon$ , binary  $(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n)$ ;
2: Compute the integer optimality cut coefficients  $(\pi_{\text{int}}, \tau_{\text{int}})$  and set the iterate  $(\bar{\pi}, \bar{\tau}) \leftarrow (\pi^H, \tau^H)$ ;
3: while  $i < I_{\max}$  do
4:   if  $\mathcal{R}_{\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n}(\bar{\pi}, \bar{\tau}) = \mathcal{R}_{\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n}(\pi_{\text{int}}, \tau_{\text{int}})$  then
5:     break;
6:   else
7:     Consider the convex combination  $(\pi, \tau) \leftarrow \alpha(\pi_{\text{int}}, \tau_{\text{int}}) + (1 - \alpha)(\bar{\pi}, \bar{\tau})$ ;
8:     Find a  $(\pi^{h^*}, \tau^{h^*}) \in \arg \min\{\|(\pi^h, \tau^h) - (\pi, \tau)\|_1 : (\pi^h, \tau^h) \in \Pi\}$ ;
9:     If  $\|(\pi^{h^*}, \tau^{h^*}) - (\pi, \tau)\|_1 \leq \epsilon$  then  $(\bar{\pi}, \bar{\tau}) \leftarrow (\pi^{h^*}, \tau^{h^*})$  and remove  $(\pi^{h^*}, \tau^{h^*})$  from  $\Pi$ ;
10:    else  $(\bar{\pi}, \bar{\tau}) \leftarrow (\pi, \tau)$ ;
11:     $i \leftarrow i + 1$ ;
12:   end if
13: end while
14: if  $\mathcal{R}_{\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n}(\bar{\pi}, \bar{\tau}) \neq \mathcal{R}_{\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n}(\pi_{\text{int}}, \tau_{\text{int}})$  then
15:    $(\bar{\pi}, \bar{\tau}) \leftarrow (\pi_{\text{int}}, \tau_{\text{int}})$ ;
16: end if
17: Return  $(\bar{\pi}, \bar{\tau})$ ;
```

5 Numerical Case Study

We present a numerical case study based on a California power grid. We introduce the case and experiment setup in Section 5.1, compare the effectiveness of different computational strategies in Section 5.2, report the performance of our dynamic line switching model in Section 5.3, and demonstrate the value of modeling DDU in Section 5.4. Finally, in Section 5.5 we construct two classes of easy-to-implement line switching policies and demonstrate their performance.

Table 2: Characteristics of transmission grid

Component	Characteristic	Min	Max	Average
Line	Rating (MW/h)	20.1	110.4	73.2
Bus	Load (MW/h)	0.0	14.2	3.5
Generator	Capacity (MW/h)	2.3	120.6	88.2
	Cost (\$/MWh)	29.5	54.0	32.5
Network	Total load (MW/h)	170.3	220.2	192.2

5.1 Experiment setup

We create a test instance based on a region in California and the transmission grid therein through the CATS dataset (Taylor et al., 2023). In Table 2, we report the ranges of the instance parameters. In addition, we generate a scenario tree for wildfire propagation based on historical wildfire perimeters (California Department of Forestry and Fire Protection, 2024) and detail the generation approach in Appendix A.2. Finally, we generate the bus load based on the CAISO 2019 load data (FERC (2022); see Figs. 7–8 in Appendix A.8 for a depiction of the instance data).

To calibrate the ambiguity set $\mathcal{P}_m(\mathbf{x}_n)$, we set a base value $\bar{\gamma}$ based on the fuel data in the region as well as transmission line characteristics, and a base value $\bar{\beta}$ based on transmission line characteristics such as nominal capacity and thermal rating. We conduct sensitivity analysis for γ and β in Section 5.4. We ran all experiments on an Intel Xeon CPU with 12 cores @ 3.4 GHz and 128 GB of memory. We solved all MILP

formulations using GUROBI 11 through Python 3.11.8. We generated the cutting planes in Algorithm 2 using lazy-constraints in GUROBI.

Table 3: Gap comparison between different representations of $\mathbf{x}$.

Representation of $\mathbf{x}$ in Lagrangian cuts	Best gap (%)	Iterations	Runtime (s)
Binary $\mathbf{f}$ and $\mathbf{p}$	74.3	24	391.2
Continuous $\mathbf{f}$ + Binary $\mathbf{p}$	82.9	43	254.3
Continuous $\mathbf{p}$ + Binary $\mathbf{f}$	60.3	35	301.2
Proposed	45.2	51	434.9

5.2 Comparisons of computing strategies

First, we evaluate the strategy of generating the Lagrangian cuts with respect to the continuous decision variables $(f_{ln}, p_{gp(n)})$, as opposed to their binary expansion approximations, as detailed in Section 4.2.2. For comparison, we implemented the benchmark strategies of generating Lagrangian cuts based on (i) binary expansion of $(f_{ln}, p_{gp(n)})$, (ii) continuous f_{ln} and binary expansion of $p_{gp(n)}$, and (iii) continuous $p_{gp(n)}$ and binary expansion of f_{ln} . We run Algorithm 2 on an instance with $T = 24$, $|\mathcal{S}_T| = 150$ using these strategies until no improvement of LB for 5 consecutive iterations and then report the gap between the final LB and the optimal value in Table 3. From this table, we observe that the proposed strategy achieved a significantly smaller gap than other benchmarks in slightly longer runtime. Hence, we stick to this strategy for the rest of numerical case studies.

Table 4: Comparison of time and iterations of different cut strategies to convergence

T	$ \mathcal{S}_T $	I			L			I+SB			L+SB			Algorithm 3+SB		
		1%	0.1%	#Iter	1%	0.1%	#Iter	1%	0.1%	#Iter	1%	0.1%	#Iter	1%	0.1%	#Iter
6	50	881	1446	252	909	1096	23	512	630	37	678	794	15	382	425	27
6	100	1840	3141	378	2377	2983	45	1017	1302	64	1487	1813	26	894	1023	43
6	150	3561	5737	593	2948	3492	45	2012	2431	92	3126	3597	47	1405	2231	69
12	50	2230	3306	365	1468	1600	28	840	934	51	1956	2071	33	682	765	44
12	100	2306	3512	339	2546	2852	48	1840	2102	83	2389	2599	36	1537	1729	58
12	150	6924	9722	766	7452	7694	77	3895	4102	124	8102	8126	82	3402	3624	101
24	50	2076	3304	458	2078	2431	37	840	1480	73	2617	2975	45	920	1252	81
24	100	5491	7933	813	5999	6373	95	2957	3204	120	5041	5202	57	2498	2758	111
24	150	11656	17596	1137	10567	11729	75	5203	7023	147	10079	10868	68	5810	6203	126

Columns 1% and 0.1% report the time in seconds to reach the corresponding optimality gap.

Second, we evaluate the effectiveness of different cutting planes, as well as their combinations, for lower approximating $Q_m(\mathbf{x}_n, \mathbf{a}_m)$. Specifically, we consider three types of cuts: integer optimality cuts (I), Lagrangian cuts (L), and strengthened Benders cuts (SB) in the backward pass of Algorithm 2. We tested (I) and (L) by themselves as they are tight cuts that can guarantee convergence. Additionally, we incorporated (SB) to accelerate convergence and tested the combinations (I+SB) and (L+SB). We report the runtime and the number of iterations for Algorithm 2 converging to an optimality gap of 1% and 0.1%, respectively, in Table 4. Putting the last strategy (Algorithm 3+SB) aside, we observe that the per-iteration runtime for (I) is the shortest, yet more iterations are required. This suggests that, although the (I) cuts are cheap to generate, their strength is relatively weaker. In contrast, (L) achieved the same optimality gaps within much less number of iterations but similar or even longer runtime than (I). This indicates that Lagrangian cuts are stronger but take longer to obtain. It makes sense because producing each Lagrangian cut involves a subgradient algorithm and solving a series of MILP. In addition, incorporating (SB) significantly accelerated

the convergence, with shorter runtime and less iterations. This indicates that (SB) cuts are good complements for both (I) and (L) cuts. In particular, the combination (I+SB) constantly outperforms other strategies on different test instances.

Third, we evaluate Algorithm 3, which reuses dual multipliers from earlier iterations to generate Lagrangian cuts. We invoked this strategy every 5 iterations of Algorithm 2 with $\epsilon = 0.3$ and $I_{\max} = 15$. From Table 4, we observe that the proposed strategy significantly shortened the runtime. For example, it cuts the runtime of (L+SB) in half in most instances to achieve an optimality gap of 1% or 0.1%. Even when compared to the best-performing strategy (I+SB) from the last comparison, Algorithm 3 was able to further shorten the runtime by around 15%. In addition, this strategy produced more iterations than the combination (L+SB), suggesting that Algorithm 3 is indeed computationally cheaper than the subgradient algorithm.

5.3 Value of dynamic line switching

We evaluate the value of adopting a dynamic line switching policy (MS), as opposed to an alternative two-stage line switching plan (TS) or simply no switching (NS). Here, TS refers to (preventively) reconfiguring the transmission grid at the root node of $\mathcal{T}$ and then sticking to the grid topology throughout (Yang et al., 2024a; Hosseini and Parvania, 2020), while other operations (power generation, phase angle, etc.) can still be dynamically adjusted (see Appendix A.5 for a detailed model). In addition, NS is a special case of TS without initial grid reconfiguration.

In Fig. 2, we compare the average load shedding costs and average operational costs of MS, TS, and NS across all 24 hours of simulation. We compute the average at stage t by considering the costs of every node in the set $\mathcal{S}_t$ and weighing them by their probability of occurrence. From Fig. 2a, we observe that TS and NS resulted in drastically higher load loss than MS. This demonstrates the value of adopting a dynamic line switching policy during wildfires. To demonstrate the flexibility of MS, we depict Fig. 3 and compare the switching decisions of MS and TS in two wildfire scenarios (denoted A and B). We observe that, while TS cannot adapt to these distinct wildfire propagation scenarios (it opened the same line), MS reconfigured the grid and rerouted the power flow accordingly. On the other hand, Fig. 2b shows that MS also incurred a higher operational cost than TS and NS (for extended comparisons, see the online repository Estrada-Garcia et al. (2025) of this paper).

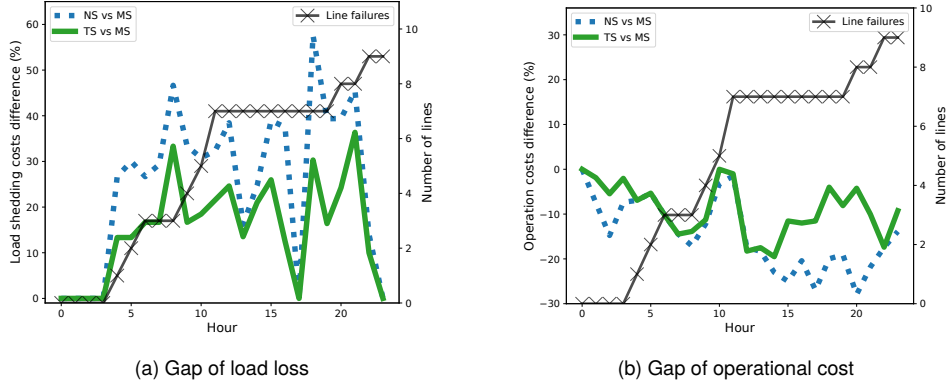


Figure 2: Average metric gaps between MS, TS, and NS

5.4 Impact of DDU and DIU parameters

The ambiguity set $\mathcal{P}_m(\mathbf{x}_n)$ models the line survival probability using a DDU component using parameter β and a DIU component using parameter γ . To examine the impacts of DDU and DIU parameters, we

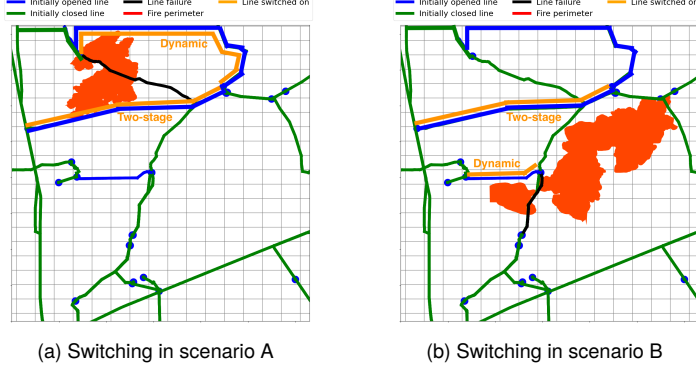


Figure 3: Comparison of switching decisions between MS and TS

visualize the optimal value of the MS model as a function of β and γ in Fig. 4. From Fig. 4, we observe that the impacts of the DDU parameter β are significantly larger than the DIU counterpart. This suggests that ignoring DDU can undermine the effectiveness of the MS model, demonstrating the value of modeling DDU in our study (for a complete set of sensitivity analyses, see the online repository [Estrada-Garcia et al. \(2025\)](#) of this paper).

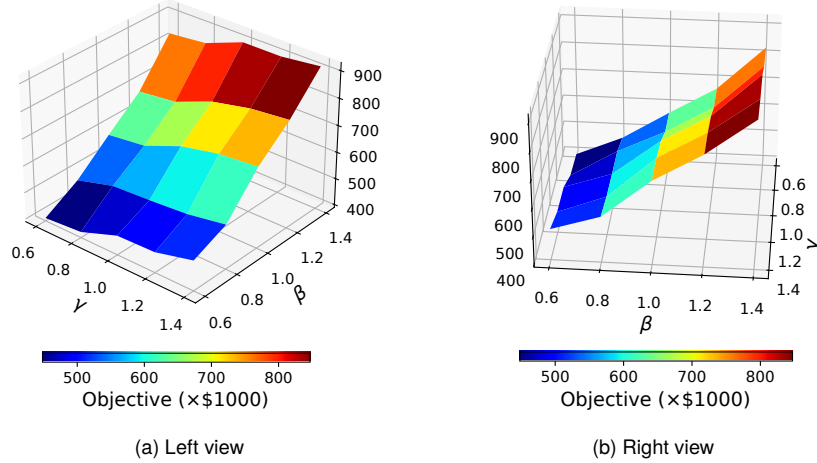


Figure 4: Sensitivity of the MS optimal value in the DDU parameter β and DIU parameter γ

5.5 Easy-to-implement policies

We produce three line switching policies based on the final value function approximations $\underline{Q}_m(\cdot, \cdot)$ obtained from the SND algorithm. Given a time unit t , the wildfire state w_t^{fire} , and the grid state $(\mathbf{x}_{t-1}, \tilde{\mathbf{a}}_t)$, we first project w_t^{fire} to a node $n \in \mathcal{S}_t$ encoding the closest wildfire state. Then, we produce the following three policies based on $\underline{Q}_m(\cdot, \cdot)$.

Dynamic policy solves formulation (1) pertaining to node n and state $(\mathbf{x}_{t-1}, \tilde{\mathbf{a}}_t)$ with $Q_m(\cdot, \cdot)$ replaced by $\underline{Q}_m(\cdot, \cdot)$, and applies the ensuing solution $(\mathbf{x}_n, \mathbf{y}_n)$.

Topology policy is the same with the dynamic policy except that, when solving formulation (1), it shrinks the feasible region of line switching to a set of grid topologies we collect from the final iteration of the SND algorithm. We detail the generation of this topology set in Algorithm 4 in Appendix A.7.

Mapping policy seeks to establish a deterministic look-up table that maps to the switching decision z_{lt} from the current state $z_{l,t-1}$ of line l , as well as the availability $\tilde{\mathbf{a}}_t^{\mathcal{L}(l)}$ of the redundant lines in $\mathcal{L}(l) := \{j \in \mathcal{L} : \text{line } l \text{ and } j \text{ feeds to the same bus}\}$. We sample the scenario tree in multiple replications and, in each replication, retrieve the decision for z_{lt} from the dynamic policy. Then, we round the average of these z_{lt} among all replications to either zero or one. We detail this approach in Algorithm 5 in Appendix A.7.

From a computing perspective, dynamic policy demands solving an MILP (e.g., formulation (1)) for each time unit t in its implementation. This is cheaper than solving the entire multi-stage DRO model (with a receding horizon, from time t to T), but can still be time-consuming with a large-scale transmission grid. As an alternative, the topology policy only chooses among a select set of topologies. We can implement it much more efficiently and interpret it more easily. Mapping policy completely waives the need of solving any optimization problems. We note that, although the (offline) preparation of the look-up tables demands resolving MILPs, the (online) implementation of these tables consumes no time. We evaluate the out-of-sample performance of these policies. To this end, we shall refer to the scenario tree we use in Algorithm 2 as the training (in-sample) tree and perform out-of-sample simulations in the following two new trees:

- We regenerate a scenario tree using the same transition probabilities as in the training tree and refer to it as the out-of-sample tree.
- We generate another scenario tree with a more hazardous transition probabilities, with faster and/or larger propagation of the wildfire. We refer to it as the stress-test tree.

When implementing the above three policies, we project each new scenario (unobserved in the training tree) of the out-of-sample tree and stress-test tree to the closest scenario in the training tree with respect to 1-norm. We compare the performance of the topology policy and the mapping policy with the TS and the NS approaches in Table 5, which reports the average and standard deviation of the performance gap (between these approaches and the dynamic policy) across 10 replications. From this table, we observe that the topology and mapping policies outperform both TS and NS significantly. This suggests that even the simplified versions of the dynamic policy retain an operational advantage of adaptability and outperform the non-adaptive policies (for additional results of this comparison, see the online repository Estrada-Garcia et al. (2025) of this paper).

Table 5: Performance comparison of various policies in out-of-sample simulation.

Approach vs Dynamic	Metric (%)	Average (Standard Deviation)		
		Training	Out-of-sample	Stress-test
TS	Objective	15.83	15.75 (1.03)	24.58 (4.79)
	Operation	-9.54	-9.97 (2.63)	-8.07 (4.53)
	Load shedding	26.01	26.11 (3.29)	37.77 (5.61)
NS	Objective	30.13	30.33 (0.82)	40.41 (4.22)
	Operation	-11.73	-11.32 (1.84)	-10.15 (2.78)
	Load shedding	47.92	47.11 (2.06)	60.83 (5.67)
Topology	Objective	5.78	5.14 (0.89)	8.74 (3.62)
	Operation	-7.79	-9.08 (0.82)	-7.57 (3.04)
	Load shedding	11.23	10.88 (0.95)	15.33 (2.40)
Mapping	Objective	4.35	4.12 (1.23)	10.47 (3.74)
	Operation	-0.35	-1.38 (1.52)	-3.19 (2.17)
	Load shedding	6.23	6.34 (2.01)	16.02 (3.84)

6 Conclusion

We proposed a multi-stage DRO model with DDU to address the dynamic line switching of a transmission grid amidst wildfire propagation. We extended the SND algorithm proposed in Zou et al. (2018) to solve this model and proposed strategies to enhance its computational performance, including deterministic upper bounds and faster generation of Lagrangian cuts. In addition, to facilitate online deployment of the proposed model, we proposed policies that are cheaper to implement and easier to interpret. We demonstrated the proposed model and computational strengthening strategies in a case study using a real-world transmission system and wildfire data.

References

- Michael Abdelmalak and Mohammed Benidris. Enhancing power system operational resilience against wildfires. *IEEE Transactions on Industry Applications*, 58(2):1611–1621, 2022.
- Qasim Al Saeed and Hamidreza Nazarpouya. Impact of wildfires on power systems. In *2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, pages 1–5. IEEE, 2022.
- Reza Bayani and Saeed D Manshadi. Resilient expansion planning of electricity grid under prolonged wildfire risk. *IEEE Transactions on Smart Grid*, 14(5):3719–3731, 2023.
- Charles E Blair and Robert G Jeroslow. The value function of a mixed integer program: I. *Discrete Mathematics*, 19(2):121–138, 1977.
- California Department of Forestry and Fire Protection. Historic fire perimeters. Data set, California Department of Forestry and Fire Protection, May 2024.
- Rui Chen and James Luedtke. On generating lagrangian cuts for two-stage stochastic integer programs. *INFORMS Journal on Computing*, 34(4):2332–2349, 2022.
- Moein Choobineh and Salman Mohagheghi. Power grid vulnerability assessment against wildfires using probabilistic progression estimation model. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2016.
- Songyi Dian, Peng Cheng, Qiang Ye, Jirong Wu, Ruisen Luo, Chen Wang, Dafeng Hui, Ning Zhou, Dong Zou, Qin Yu, et al. Integrating wildfires propagation prediction into early warning of electrical transmission line outages. *IEEE Access*, 7:27586–27603, 2019.
- Juan-Alberto Estrada-Garcia, Ruiwei Jiang, and Alexandre Moreira. Dynamic line-switching in transmission system operation amidst wildfire-prone weather under decision-dependent uncertainty, 2025. Available for download at <https://github.com/JestradaG/DynamicLineSwitching>.
- FERC. Form 1 - electric utility annual report. Technical report, Federal Energy Regulatory Commission (FERC), March 2022.
- Emily B Fisher, Richard P O’Neill, and Michael C Ferris. Optimal transmission switching. *IEEE Transactions on Power Systems*, 23(3):1346–1355, 2008.
- Vijay Gupta and Ignacio E Grossmann. Solution strategies for multistage stochastic programming with endogenous uncertainties. *Computers & Chemical Engineering*, 35(11):2235–2247, 2011.

- Jessica E Halofsky, David L Peterson, and Brian J Harvey. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the pacific northwest, usa. *Fire Ecology*, 16(1):1–26, 2020.
- Kory W Hedman, Michael C Ferris, Richard P O'Neill, Emily Bartholomew Fisher, and Shmuel S Oren. Co-optimization of generation unit commitment and transmission switching with n-1 reliability. *IEEE Transactions on Power Systems*, 25(2):1052–1063, 2010.
- Holger Heitsch and Werner Römis. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118:371–406, 2009.
- Mohammad Mehdi Hosseini and Masood Parvania. Computationally efficient formulations for fault isolation and service restoration in distribution systems. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2020.
- Yuping Huang, Panos M Pardalos, and Qipeng P Zheng. *Electrical power unit commitment: deterministic and two-stage stochastic programming models and algorithms*. Springer, 2017.
- Kamran Jalilpoor, Arman Oshnoei, Behnam Mohammadi-Ivatloo, and Amjad Anvari-Moghaddam. Network hardening and optimal placement of microgrids to improve transmission system resilience: A two-stage linear program. *Reliability Engineering & System Safety*, 224:108536, 2022.
- Spyros Kondylatos, Ioannis Prapas, Michele Ronco, Ioannis Papoutsis, Gustau Camps-Valls, María Piles, Miguel-Ángel Fernández-Torres, and Nuno Carvalhais. Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters*, 49(17):e2022GL099368, 2022.
- Gilbert Laporte and François V Louveaux. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13(3):133–142, 1993.
- Can Li and Ignacio E Grossmann. A review of stochastic programming methods for optimization of process systems under uncertainty. *Frontiers in Chemical Engineering*, 2:622241, 2021.
- Salman Mohagheghi and Steffen Rebennack. Optimal resilient power grid operation during the course of a progressing wildfire. *International Journal of Electrical Power & Energy Systems*, 73:843–852, 2015.
- Seyed Masoud Mohseni-Bonab, Innocent Kamwa, Abbas Rabiee, and CY Chung. Stochastic optimal transmission switching: A novel approach to enhance power grid security margins through vulnerability mitigation under renewables uncertainties. *Applied Energy*, 305:117851, 2022.
- Alexandre Moreira, Felipe Piancó, Bruno Fanzeres, Alexandre Street, Ruiwei Jiang, Chaoyue Zhao, and Miguel Heleno. Distribution system operation amidst wildfire-prone climate conditions under decision-dependent line availability uncertainty. *IEEE Transactions on Power Systems*, 2024.
- John W Muhs, Masood Parvania, Hieu T Nguyen, and John A Palmer. Characterizing probability of wildfire ignition caused by power distribution lines. *IEEE Transactions on Power Delivery*, 36(6):3681–3688, 2020.
- Harsha Nagarajan, Emre Yamangil, Russell Bent, Pascal Van Hentenryck, and Scott Backhaus. Optimal resilient transmission grid design. In *2016 Power Systems Computation Conference (PSCC)*, pages 1–7. IEEE, 2016.
- Hieu Trung Nguyen, John Muhs, and Masood Parvania. Preparatory operation of automated distribution systems for resilience enhancement of critical loads. *IEEE Transactions on Power Delivery*, 36(4): 2354–2362, 2020.

- Fernando Luiz Cyrino Oliveira, Reinaldo Castro Souza, and André Luís Marques Marcato. A time series model for building scenarios trees applied to stochastic optimisation. *International Journal of Electrical Power & Energy Systems*, 67:315–323, 2015.
- Jonathan H Owen and Sanjay Mehrotra. On the value of binary expansions for general mixed-integer linear programs. *Operations Research*, 50(5):810–819, 2002.
- Mario VF Pereira and Leontina MVG Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.
- Felipe Piancó, Alexandre Moreira, Bruno Fanzeres, Ruiwei Jiang, Chaoyue Zhao, and Miguel Heleno. Decision-dependent uncertainty-aware distribution system planning under wildfire risk. *arXiv preprint arXiv:2405.04350*, 2024.
- Zi Pu, Jiang Ruan, Dao Huang, Tian Wu, and Peng Li. Study on the breakdown characteristics of the transmission line gap under forest fire conditions. *International Transactions on Electrical Energy Systems*, 25(11):2731–2744, 2015.
- Sumit K Rathor and Dipti Saxena. Energy management system for smart grid: An overview and key issues. *International Journal of Energy Research*, 44(6):4067–4109, 2020.
- Minseok Ryu and Ruiwei Jiang. Nurse staffing under absenteeism: A distributionally robust optimization approach. *Manufacturing & Service Operations Management*, 2025.
- Hamid R Sayarshad and Romina Ghorbanloo. Evaluating the resilience of electrical power line outages caused by wildfires. *Reliability Engineering & System Safety*, 240:109588, 2023.
- Sofia Taylor, Aditya Rangarajan, Noah Rhodes, Jonathan Snodgrass, Bernie Lesieutre, and Line A Roald. California test system (cats): A geographically accurate test system based on the california grid. *IEEE Transactions on Energy Markets, Policy and Regulation*, 2023.
- Eric A Udren, Chris Bolton, Dan Dietmeyer, Tariq Rahman, and Sergio Flores-Castro. Managing wildfire risks: Protection system technical developments combined with operational advances to improve public safety. *IEEE Power and Energy Magazine*, 20(1):64–77, 2022.
- Ke Wang, Yixun Xue, Qinglai Guo, Mohammad Shahidehpour, Quan Zhou, Bin Wang, and Hongbin Sun. A coordinated reconfiguration strategy for multi-stage resilience enhancement in integrated power distribution and heating networks. *IEEE Transactions on Smart Grid*, 14(4):2709–2722, 2022.
- Hanbin Yang, Noah Rhodes, Haoxiang Yang, Line Roald, and Lewis Ntamo. Multi-period power system risk minimization under wildfire disruptions. *IEEE Transactions on Power Systems*, 2024a.
- Hanbin Yang, Haoxiang Yang, Noah Rhodes, Line Roald, and Lewis Ntamo. Multistage stochastic program for mitigating power system risks under wildfire disruptions. *Electric Power Systems Research*, 234:110773, 2024b.
- Xian Yu and Siqian Shen. Multistage distributionally robust mixed-integer programming with decision-dependent moment-based ambiguity sets. *Mathematical Programming*, 196(1):1025–1064, 2022.
- Yiling Zhang, Siqian Shen, Bowen Li, and Johanna L Mathieu. Two-stage distributionally robust optimal power flow with flexible loads. In *2017 IEEE Manchester PowerTech*, pages 1–6. IEEE, 2017.
- Jikai Zou, Shabbir Ahmed, and Xu Andy Sun. Multistage stochastic unit commitment using stochastic dual dynamic integer programming. *IEEE Transactions on Power Systems*, 34(3):1814–1823, 2018.

A Appendix to the paper

A.1 Nomenclature and formulation

We present a detailed formulation for the deterministic nodal line switching model as follows, whose nomenclature is summarized in Table 6.

$$\min_{\mathbf{x}_n, \mathbf{y}_n} \sum_{g \in \mathcal{G}_n} C_g p_{gn} + \sum_{b \in \mathcal{N}} C^l \Delta_{bn} \quad (12a)$$

$$\text{s.t. } \bar{\theta}^{\min} \leq \theta_{bn} \leq \bar{\theta}^{\max}, \forall b \in \mathcal{N}, \quad (12b)$$

$$\sum_{l \in l(b, \cdot)} f_{ln} - \sum_{l \in l(\cdot, b)} f_{ln} + \sum_{g \in \mathcal{G}(b)} p_{gn} + \Delta_{bn} = d_{bn}, \forall b \in \mathcal{N}, \quad (12c)$$

$$\bar{F}_l^{\min} z_{ln} \leq f_{ln} \leq \bar{F}_l^{\max} z_{ln}, \forall l \in \mathcal{L}, \quad (12d)$$

$$z_{ln} \leq a_{ln}, \forall l \in \mathcal{L}, \quad (12e)$$

$$B_l(\theta_{bn} - \theta_{b'n}) - f_{ln} + (2 - z_{ln} - a_{ln})\mathcal{M}_l \geq 0, \forall l \in \mathcal{L}, (b, b') \in \mathcal{N}, \quad (12f)$$

$$B_l(\theta_{bn} - \theta_{b'n}) - f_{ln} - (2 - z_{ln} - a_{ln})\mathcal{M}_l \leq 0, \forall l \in \mathcal{L}, (b, b') \in \mathcal{N}, \quad (12g)$$

$$P_g^{\min} \leq p_{gn} \leq P_g^{\max}, \forall g \in \mathcal{G}, \quad (12h)$$

$$p_{gn} - p_{gp(n)} \leq R_g^+, \forall g \in \mathcal{G}, \quad (12i)$$

$$p_{gp(n)} - p_{gn} \leq R_g^-, \forall g \in \mathcal{G}, \quad (12j)$$

$$f_{ln} - (\bar{F}_{ln}^{\max} - \bar{F}_{ln}^{\text{nom}})o_{ln} \leq \bar{F}_{ln}^{\text{nom}}, \forall l \in \mathcal{L}, \quad (12k)$$

$$f_{ln} + (\bar{F}_{ln}^{\max} - \bar{F}_{ln}^{\text{nom}})o_{ln} \geq -\bar{F}_{ln}^{\text{nom}}, \forall l \in \mathcal{L}, \quad (12l)$$

$$p_{gn} \geq 0, \forall g \in \mathcal{G}, \quad (12m)$$

$$\Delta_{bn} \geq 0, \forall b \in \mathcal{N}, \quad (12n)$$

$$z_{ln}, o_{ln} \in \{0, 1\}, \forall l \in \mathcal{L}. \quad (12o)$$

In formulation (12), the objective function (12a) seeks to minimize the generator production cost plus the load shedding cost, constraints (12b) describe limits on the phase angles, constraints (12c) account for the flow balance for each bus, constraints (12d) describe the transmission line capacity limits, depending on if the line is closed ($z_{ln} = 1$) or open ($z_{ln} = 0$), constraints (12e) make sure that we can close a line only if it is available ($a_{ln} = 1$) in the first place, constraints (12f)–(12g) enforce the relationship between power flows and phase angles using DC power flow when the line is both available and closed, with $\mathcal{M}_l := |B_l(\bar{\theta}^{\max} - \bar{\theta}^{\min})|$ denoting a sufficiently large constant, constraints (12h) describe the generator minimum and maximum generation amounts, constraints (12i)–(12j) describe the generator ramp-rate limits, and constraints (12k)–(12l) determine whether the power flow along each transmission line is within its nominal capacity ($o_{ln} = 0$) or not ($o_{ln} = 1$).

A.2 Generation of the scenario tree

We employ a data-driven approach to generate the scenario tree in the case study. We begin by discretizing the geographical region in our instance by considering a rectangular grid, such that we can have xx cells.

Table 6: A summary of sets, parameters, and variables.

Sets	
$C(n)$	Set of children nodes of node n .
$\mathcal{G}$	Set of generators, indexed by g .
$g(b)$	Set of generators at bus $b \in \mathcal{N}$.
$\mathcal{L}$	Set of transmission lines, indexed by l .
$l(b, \cdot)$	Set of transmission lines with $b \in \mathcal{N}$ as the “from” bus.
$l(\cdot, b)$	Set of transmission lines with $b \in \mathcal{N}$ as the “to” bus.
$\mathcal{N}$	Set of buses, indexed by b .
$\Pi(n)$	Set of nodes in the scenario path leading to node n .
$\mathcal{S}_t$	Set of nodes in stage t of $\mathcal{T}$.
$\mathcal{T}$	Scenario tree for the wildfire propagation stochastic process.
Parameters	
a_{ln}	Availability of transmission line $l \in \mathcal{L}$ at node $n \in \mathcal{T}$ (1 if available, 0 otherwise).
B_l	Susceptance of transmission line $l \in \mathcal{L}$.
β_{ln}	Sensitivity of the survival probability to the active power flow f_{ln} of line $l \in \mathcal{L}$ at node $n \in \mathcal{T}$.
C_g	Unit generation cost of generator $g \in \mathcal{G}$.
C_b^L	Unit cost of load shedding at bus $b \in \mathcal{N}$.
d_{bn}	Real power load at bus $b \in \mathcal{N}$ for node $n \in \mathcal{T}$.
$\bar{F}_l^{\max}, \bar{F}_l^{\min}$	Maximum/minimum ratings of transmission line $l \in \mathcal{L}$.
$\bar{F}_l^{\text{nom}}$	Nominal transmission rating of line $l \in \mathcal{L}$.
γ_{ln}	Nominal survival probability of line $l \in \mathcal{L}$ at node $n \in \mathcal{T}$.
K	Maximum number of line failures throughout the planning horizon.
$P_g^{\max}, P_g^{\min}$	Maximum/minimum generation amounts of generator $g \in \mathcal{G}$.
p_{nm}	Transition probability from node n to node $m \in C(n)$ in the scenario tree $\mathcal{T}$.
T	Number of decision-making stages.
$\bar{\theta}^{\max}, \bar{\theta}^{\min}$	Maximum/minimum bus voltage angle.
Decision variables	
o_{ln}	Indicator variable for f_{ln} exceeding the nominal rating $\bar{F}_l^{\text{nom}}$ of line l (1 if exceeding, 0 if not exceeding).
p_{gn}	Real power supply from generator $g \in \mathcal{G}$ at bus $b \in \mathcal{N}$ and node $n \in \mathcal{T}$.
f_{ln}	Real power flow on line $l \in \mathcal{L}$ at node $n \in \mathcal{T}$.
θ_{bn}	Voltage angle at bus b and node $n \in \mathcal{T}$.
Δ_{bn}	Amount of load shedding at bus b and node $n \in \mathcal{T}$.
$\mathbf{x}_n$	$= [\mathbf{z}_n, \mathbf{o}_n, \mathbf{f}_n, \mathbf{p}_n]$ Vector of inter-stage decision variables at node $n \in \mathcal{T}$.
$\mathbf{y}_n$	$= [\theta_n, \Delta_n]$ Vector of intra-stage decision variables at node $n \in \mathcal{T}$.
z_{ln}	Binary variable indicating a switching action of line l at node $n \in \mathcal{T}$ (1 if closing, 0 if opening).

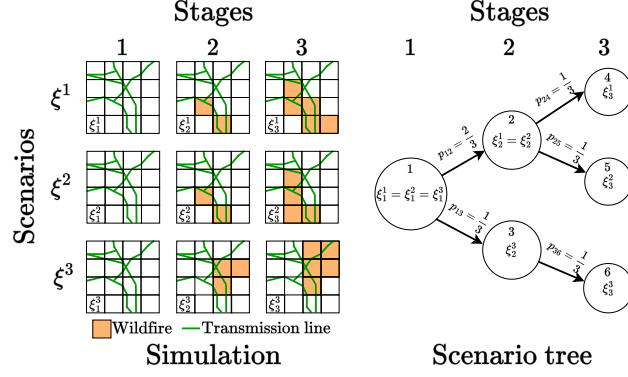


Figure 5: Scenario tree generation process

We construct the scenario tree in a backwards fashion. We begin by having as many terminal nodes in the last stage, as we have historical fire perimeters in the region. Then, we consider a cellular automaton process, with basic propagation rules considering neighborhood fire propagation over the 24-hour period we simulate. We note that we do not consider dynamic wind conditions, nor more complex propagation models, but our framework can admit scenario trees that are generated by any other valid method. Then, with the backward-simulated scenarios, we compare the state of the cells starting from an initial state without fires, and merge scenarios as long as they have the same wildfire state, accumulating their probabilities, only branching whenever the fires change as they are simulated. In Figure 5, we show the conceptual process of generating the scenario tree for the case study. Given that we assume a one-direction relationship, where wildfire impact line availability, we reduce the number of scenarios by only differentiating branching on them, i.e., differentiating between scenarios, if the wildfire impacts cells that are associated with lines that can fail, further reducing the scenario tree size.

A.3 Approximation error of the binary expansion approximation

Theorem 3. Assume that $f_n(\mathbf{x}_n, \mathbf{y}_n, \tilde{\mathbf{a}}_n)$ is Lipschitz continuous in $(\mathbf{x}_n, \mathbf{y}_n)$, i.e.,

$$|f_n(\mathbf{x}_n, \mathbf{y}_n, \tilde{\mathbf{a}}_n) - f_n(\mathbf{x}'_n, \mathbf{y}'_n, \tilde{\mathbf{a}}_n)| \leq L \|(\mathbf{x}_n, \mathbf{y}_n) - (\mathbf{x}'_n, \mathbf{y}'_n)\|_\infty, \quad \forall \mathbf{x}_n, \mathbf{x}'_n, \mathbf{y}_n, \mathbf{y}'_n, \tilde{\mathbf{a}}_n, n \in \mathcal{T}$$

for a constant $L \geq 0$. Let $z^* := Q_1(\mathbf{x}_0, \mathbf{1})$ without the binary expansion approximation and z_{bin}^* represent the optimal value with the approximation. Then, it holds that $z^* \leq z_{bin}^* \leq z^* + Ks$ for a constant $K \geq 0$, where K depends only on L and the parameters for defining $Q_1(\mathbf{x}_0, \mathbf{1})$.

Proof of Theorem 3. The binary expansion approximation on $\mathbf{x}_n$ is equivalent to incorporating a new constraint $\mathbf{x}_n \in S(\mathbf{x}_n) := \{z \in \mathbb{R}^{\dim(\mathbf{x}_n)} : z_n \in \{L_n, L_n + s, \dots, \min(U_n, L_n + (2^{E_n} - 1)s)\}, \forall n \in [N]\}$, where L, U are lower and upper bounds of $\mathbf{x}_n$ and $E_n = \lfloor \log_2 \left(\frac{U_n - L_n}{s} \right) \rfloor + 1$. Then, the binary expansion is a conservative approximation of the multi-stage DRO model and so $z^* \leq z_{bin}^*$. In addition, we note that $S(\mathbf{x}_n)$ is an s -net for the hypercube $[L, U]$. That is, for any $\mathbf{x}_n \in [L, U]$, there exists a $z \in S(\mathbf{x}_n)$ such that $\|\mathbf{x}_n - z\|_\infty \leq s$.

To prove the second inequality, we use mathematical induction. More specifically, we will focus on a further conservative approximation that approximates both $\mathbf{x}_n$ and $\mathbf{y}_n$ through binary expansion. First (**base case**), for each leaf node n and any fixed $\mathbf{x}_{p(n)}$ and $\tilde{\mathbf{a}}_n$, recall that

$$\begin{aligned} Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) &= \min_{\mathbf{x}_n, \mathbf{y}_n} f_n(\mathbf{x}_n, \mathbf{y}_n, \tilde{\mathbf{a}}_n) \\ \text{s.t.} \quad &A_n \mathbf{x}_n + W_n \mathbf{y}_n + C_n \mathbf{x}_{p(n)} + D_n \tilde{\mathbf{a}}_n \geq h_n. \end{aligned}$$

We let $Q_n^x(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ denote the optimal value of this formulation after adding the constraint $\mathbf{x}_n \in S(\mathbf{x}_n)$. Now suppose that we apply the binary expansion on both $\mathbf{x}_n$ and $\mathbf{y}_n$, that is, add the constraints $\mathbf{x}_n \in S(\mathbf{x}_n)$ and $\mathbf{y}_n \in S(\mathbf{y}_n)$ to the above formulation and denote by $Q_n^{x,y}(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ the ensuing optimal value. Then, $Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) \leq Q_n^x(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) \leq Q_n^{x,y}(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$. Likewise, we have $z_{\text{bin}}^* \equiv Q_1^x(\mathbf{x}_0, \mathbf{1}) \leq Q_1^{x,y}(\mathbf{x}_0, \mathbf{1})$.

But $S(\mathbf{x}_n) \times S(\mathbf{y}_n)$ forms an s -net. Then, for an optimal solution $(\mathbf{x}_n^*, \mathbf{y}_n^*)$ to the above formulation, there exists an $(\bar{\mathbf{x}}_n, \bar{\mathbf{y}}_n) \in S(\mathbf{x}_n) \times S(\mathbf{y}_n)$ such that

$$\begin{cases} A_n \bar{\mathbf{x}}_n + W_n \bar{\mathbf{y}}_n + C_n \mathbf{x}_{p(n)} + D_n \tilde{\mathbf{a}}_n \geq h_n \\ \|(\bar{\mathbf{x}}_n, \bar{\mathbf{y}}_n) - (\mathbf{x}_n^*, \mathbf{y}_n^*)\|_\infty \leq s. \end{cases}$$

It follows that

$$\begin{aligned} Q_n^{x,y}(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) &\leq f_n(\bar{\mathbf{x}}_n, \bar{\mathbf{y}}_n, \tilde{\mathbf{a}}_n) \\ &\leq f_n(\mathbf{x}_n^*, \mathbf{y}_n^*, \tilde{\mathbf{a}}_n) + L \|(\bar{\mathbf{x}}_n, \bar{\mathbf{y}}_n) - (\mathbf{x}_n^*, \mathbf{y}_n^*)\|_\infty \\ &\leq Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) + Ls, \end{aligned}$$

where the first inequality follows from the sub-optimality of $(\bar{\mathbf{x}}_n, \bar{\mathbf{y}}_n)$ and the second inequality follows from the Lipschitz continuity of f_n . Furthermore, since f_n is linear and decision variables $(\mathbf{x}_n, \mathbf{y}_n)$ are binary in the formulation of $Q_n^{x,y}(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$, the proximity result of binary integer program (Theorem 2.2 in [Blair and Jeroslow \(1977\)](#)) yields, for some $E_n \geq 0$

$$|Q_n^{x,y}(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) - Q_n^{x,y}(\mathbf{x}'_{p(n)}, \tilde{\mathbf{a}}_n)| \leq E_n \|\mathbf{x}_{p(n)} - \mathbf{x}'_{p(n)}\|_\infty, \quad \forall \mathbf{x}_{p(n)}, \mathbf{x}'_{p(n)}.$$

Second (**induction step**), for each node $n \in \mathcal{S}_t$ with $t \leq T-1$ and any fixed $\mathbf{x}_{p(n)}$ and $\tilde{\mathbf{a}}_n$, recall that

$$\begin{aligned} Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) &= \min_{\mathbf{x}_n, \mathbf{y}_n} f_n(\mathbf{x}_n, \mathbf{y}_n, \tilde{\mathbf{a}}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] \\ \text{s.t.} \quad &A_n \mathbf{x}_n + W_n \mathbf{y}_n + C_n \mathbf{x}_{p(n)} + D_n \tilde{\mathbf{a}}_n \geq h_n \end{aligned}$$

and assume that, for all $m \in C(n)$, $\mathbf{x}_n$, $\mathbf{x}'_n$, and $\tilde{\mathbf{a}}_m$,

1. $Q_m^{x,y}(\mathbf{x}_n, \tilde{\mathbf{a}}_m) \leq Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m) + L_m s$ for some $L_m \geq 0$;
2. $|Q_m^{x,y}(\mathbf{x}_n, \tilde{\mathbf{a}}_m) - Q_m^{x,y}(\mathbf{x}'_n, \tilde{\mathbf{a}}_m)| \leq E_m \|\mathbf{x}_n - \mathbf{x}'_n\|_\infty$ for some $E_m \geq 0$.

Then, for fixed $\mathbf{x}_n$ and $\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)$, assumption (i) implies that

$$\mathbb{E}_{\mathbb{P}} [Q_m^{x,y}(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] \leq \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] + L_m s.$$

Driving both sides to supremum with respect to $\mathbb{P}$ and noting that $\sum_{m \in C(n)} p_{nm} \equiv 1$, we have

$$\sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m^{x,y}(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] \leq \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] + \left(\sum_{m \in C(n)} p_{nm} L_m \right) s.$$

In addition, assumption (ii) implies that, for any $\mathbf{x}_n$, $\mathbf{x}'_n$, and $m \in C(n)$,

$$\left| \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m^{x,y}(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] - \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}'_n)} \mathbb{E}_{\mathbb{P}} [Q_m^{x,y}(\mathbf{x}'_n, \tilde{\mathbf{a}}_m)] \right|$$

$$\begin{aligned}
&\leq \left| \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m^{\mathbf{x},\mathbf{y}}(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] - \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}'_n)} \mathbb{E}_{\mathbb{P}} [Q_m^{\mathbf{x},\mathbf{y}}(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] \right| + E_m \|\mathbf{x}_n - \mathbf{x}'_n\|_{\infty} \\
&\leq (H_m + E_m) \|\mathbf{x}_n - \mathbf{x}'_n\|_{\infty}
\end{aligned}$$

for some $H_m \geq 0$, where the second inequality follows from the Hoffman's Lemma. Then, for an optimal solution $(\mathbf{x}_n^*, \mathbf{y}_n^*)$ to the above formulation, there exists a feasible $(\bar{\mathbf{x}}_n, \bar{\mathbf{y}}_n) \in S(\mathbf{x}_n) \times S(\mathbf{y}_n)$ such that $\|(\bar{\mathbf{x}}_n, \bar{\mathbf{y}}_n) - (\mathbf{x}_n^*, \mathbf{y}_n^*)\|_{\infty} \leq s$. It follows that

$$\begin{aligned}
Q_n^{\mathbf{x},\mathbf{y}}(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) &\leq f_n(\bar{\mathbf{x}}_n, \bar{\mathbf{y}}_n, \tilde{\mathbf{a}}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\bar{\mathbf{x}}_n)} \mathbb{E}_{\mathbb{P}} [Q_m^{\mathbf{x},\mathbf{y}}(\bar{\mathbf{x}}_n, \tilde{\mathbf{a}}_m)] \\
&\leq f_n(\mathbf{x}_n^*, \mathbf{y}_n^*, \tilde{\mathbf{a}}_n) + Ls + \sum_{m \in C(n)} p_{nm} \left(\sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n^*)} \mathbb{E}_{\mathbb{P}} [Q_m^{\mathbf{x},\mathbf{y}}(\mathbf{x}_n^*, \tilde{\mathbf{a}}_m)] + (H_m + E_m)s \right) \\
&\leq f_n(\mathbf{x}_n^*, \mathbf{y}_n^*, \tilde{\mathbf{a}}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n^*)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n^*, \tilde{\mathbf{a}}_m)] + (L + H_m + E_m)s \\
&\leq Q_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) + (L + H_m + E_m)s.
\end{aligned}$$

Finally, since function $\sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m^{\mathbf{x},\mathbf{y}}(\mathbf{x}_n, \tilde{\mathbf{a}}_m)]$ is defined on a binary domain due to the binary expansion of $(\mathbf{x}_n, \mathbf{y}_n)$, its epigraph admits a polyhedral description. Then, the formulation of $Q_n^{\mathbf{x},\mathbf{y}}(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n)$ can be written as a binary integer program. Hence, we can once again invoke the proximity result (Theorem 2.2 in [Blair and Jeroslow \(1977\)](#)) to obtain

$$|Q_n^{\mathbf{x},\mathbf{y}}(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) - Q_n^{\mathbf{x},\mathbf{y}}(\mathbf{x}'_{p(n)}, \tilde{\mathbf{a}}_n)| \leq E_n \|\mathbf{x}_{p(n)} - \mathbf{x}'_{p(n)}\|_{\infty}, \quad \forall \mathbf{x}_{p(n)}, \mathbf{x}'_{p(n)}$$

for some $E_n \geq 0$. This completes the proof. $\square$

Theorem 3 implies that the binary expansion approximation with a precision s of the multi-stage DRO model converges linearly to the true optimal value as s decreases towards zero.

A.4 Technical Proofs

Proof of Proposition 1. We express the worst-case expectation with respect to $\mathbb{P}$ as an optimization formulation:

$$\max_{\mathbf{p}_{\mathbf{a}_m}} \sum_{\mathbf{a}_m \in \mathcal{A}_m} p_{\mathbf{a}_m} Q_m(\mathbf{x}_n, \mathbf{a}_m) \quad (13a)$$

$$\text{s.t.} \quad \sum_{\mathbf{a}_m \in \mathcal{A}_m} p_{\mathbf{a}_m} \mathbf{a}_m \leq \beta_m \mathbf{x}_n + \gamma_m \quad [\psi_m] \quad (13b)$$

$$\sum_{\mathbf{a}_m \in \mathcal{A}_m} p_{\mathbf{a}_m} = 1 \quad [\phi_m] \quad (13c)$$

$$p_{\mathbf{a}_m} \geq 0, \quad \forall \mathbf{a}_m \in \mathcal{A}_m. \quad (13d)$$

In the above formulation, we encode $\mathbb{P}$ using a finite-dimensional vector $p_{\mathbf{a}_m}$ because $\tilde{\mathbf{a}}_m$ is a discrete random vector with a finite support $\mathcal{A}_m \subseteq \{0, 1\}^{|\mathcal{L}|}$. This renders formulation (13) a finite-dimensional linear program and the conclusion follows from the standard dual formulation, where ψ_m and ϕ_m are dual variables associated with constraints (13c)–(13d), respectively. $\square$

Proof of Proposition 2. By construction, we recast (SEP) as

$$\begin{aligned} \max_{\mathbf{a} \in \mathcal{A}_m} \{Q_m(\mathbf{x}_n, \mathbf{a}) - \mathbf{a}^\top \boldsymbol{\psi}_m\} &= \max_{\mathbf{a} \in \mathcal{A}_m} \left\{ \max_{h \in \mathcal{H}_m} \{(\pi^h)^\top \mathbf{x}_n + (\tau^h)^\top \mathbf{a} + \omega^h\} - \mathbf{a}^\top \boldsymbol{\psi}_m \right\} \\ &= \max_{h \in \mathcal{H}_m} \max_{\mathbf{a} \in \mathcal{A}_m} \{(\pi^h)^\top \mathbf{x}_n + (\tau^h - \boldsymbol{\psi}_m)^\top \mathbf{a} + \omega^h\}. \end{aligned}$$

For each $h \in \mathcal{H}_m$, the inner maximization problem with respect to $\mathbf{a}$ is a linear integer program with an individual cardinality constraint. Hence, it can be solved by sorting the entries $(\tau^h - \boldsymbol{\psi}_m)$, as done in each for loop of Algorithm 1. Therefore, Algorithm 1 correctly evaluates the optimal value of (SEP). Finally, since Algorithm 1 conducts $|\mathcal{H}_m|$ for loops and each loop sorts at most $|\mathcal{L}|$ entries, the running time is $O(|\mathcal{H}_m||\mathcal{L}|\log(|\mathcal{L}|))$. $\square$

Proof of Proposition 3. First, formulations (2) and (4) have the same optimal value by construction of the delayed constraint generation. The equality in (4a) follows. Second, we take the dual of formulation (4) to obtain

$$\begin{aligned} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)] &= \max_{\lambda \geq 0} \sum_{\mathbf{a}^h \in \mathcal{A}_m^*} \lambda^h Q_m(\mathbf{x}_n, \mathbf{a}^h) \\ \text{s.t.} \quad &\sum_{\mathbf{a}^h \in \mathcal{A}_m^*} \lambda^h \mathbf{a}^h \leq \beta_m \mathbf{x}_n + \gamma_m, \\ &\sum_{\mathbf{a}^h \in \mathcal{A}_m^*} \lambda^h = 1, \end{aligned}$$

The last two constraints show that a dual optimal solution λ^h defines a probability distribution $\mathbb{P}^*$ for $\tilde{\mathbf{a}}_m$ supported on $\mathcal{A}_m^* \subseteq \mathcal{A}_m$ and the first constraint shows that $\mathbb{P}^* \in \mathcal{P}_m(\mathbf{x}_n)$. Finally, we complete the proof by noticing that

$$\sum_{\mathbf{a}^h \in \mathcal{A}_m^*} \lambda^h Q_m(\mathbf{x}_n, \mathbf{a}^h) = \mathbb{E}_{\mathbb{P}^*} [Q_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)].$$

$\square$

Proof of Proposition 6. It is clear that

$$\underline{Q}_n(\mathbf{x}_{p(n)}, \tilde{\mathbf{a}}_n) \geq \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - \left(\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n \right) (\|\mathbf{x}_{p(n)} - \hat{\mathbf{x}}_{p(n)}\|_1 + \|\tilde{\mathbf{a}}_n - \hat{\mathbf{a}}_n\|_1).$$

The conclusion follows from noticing that $\|\mathbf{x}_{p(n)} - \hat{\mathbf{x}}_{p(n)}\|_1 = \mathbf{1}^\top \mathbf{x}_{p(n)} + \mathbf{1}^\top \hat{\mathbf{x}}_{p(n)} - 2\hat{\mathbf{x}}_{p(n)}^\top \mathbf{x}_{p(n)}$ and $\|\tilde{\mathbf{a}}_n - \hat{\mathbf{a}}_n\|_1 = \mathbf{1}^\top \tilde{\mathbf{a}}_n + \mathbf{1}^\top \hat{\mathbf{a}}_n - 2\hat{\mathbf{a}}_n^\top \tilde{\mathbf{a}}_n$. $\square$

Proof of Proposition 7. First, equality (9) holds for the integer optimality cut because

$$\begin{aligned} &\pi^\top \hat{\mathbf{x}}_{p(n)} + \tau^\top \hat{\mathbf{a}}_n + \omega \\ &= (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(2\hat{\mathbf{x}}_{p(n)} - \mathbf{1})^\top \hat{\mathbf{x}}_{p(n)} + (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(2\hat{\mathbf{a}}_n - \mathbf{1})^\top \hat{\mathbf{a}}_n + \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) \\ &\quad - (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(\mathbf{1}^\top \hat{\mathbf{x}}_{p(n)} + \mathbf{1}^\top \hat{\mathbf{a}}_n) \\ &= 2(\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(\hat{\mathbf{x}}_{p(n)} - \mathbf{1})^\top \hat{\mathbf{x}}_{p(n)} + 2(\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(\hat{\mathbf{a}}_n - \mathbf{1})^\top \hat{\mathbf{a}}_n + \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) \\ &= \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n), \end{aligned}$$

where the last equality is because $(x - \mathbf{1})^\top x = 0$ for any binary vector x . Second, the validity of the Lagrangian cut implies that $\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) \geq \pi^\top \hat{\mathbf{x}}_{p(n)} + \tau^\top \hat{\mathbf{a}}_n + \mathcal{L}_n(\pi, \tau)$ for all (π, τ) . But this inequality holds with

equality for the choice of (π, τ) in the integer optimality cut, implying that $(\pi, \tau) \in \operatorname{argmax}_{\pi, \tau} \{\pi^\top \hat{\mathbf{x}}_{p(n)} + \tau^\top \hat{\mathbf{a}}_n + \mathcal{L}_n(\pi, \tau)\}$. Finally, for notation brevity we denote

$$\bar{f}_n(\mathbf{x}_n, \mathbf{y}_n, \mathbf{w}_n) := f_n(\mathbf{x}_n, \mathbf{y}_n, \mathbf{w}_n) + \sum_{m \in C(n)} p_{nm} \sup_{\mathbb{P} \in \mathcal{P}_m(\mathbf{x}_n)} \mathbb{E}_{\mathbb{P}} [\underline{Q}_m(\mathbf{x}_n, \tilde{\mathbf{a}}_m)]$$

. Then,

$$\begin{aligned} & \mathcal{L}_n(\pi, \tau) + \pi^\top \hat{\mathbf{x}}_{p(n)} + \tau^\top \hat{\mathbf{a}}_n \\ = & \min_{\substack{\mathbf{x}_n, \mathbf{y}_n, \\ \mathbf{r}_n, \mathbf{w}_n: (1b), (6c)}} \bar{f}_n(\mathbf{x}_n, \mathbf{y}_n, \mathbf{w}_n) - \pi^\top (\mathbf{r}_n - \hat{\mathbf{x}}_{p(n)}) - \tau^\top (\mathbf{w}_n - \hat{\mathbf{a}}_n) \\ = & \min_{\substack{\mathbf{x}_n, \mathbf{y}_n, \\ \mathbf{r}_n, \mathbf{w}_n: (1b), (6c)}} \bar{f}_n(\mathbf{x}_n, \mathbf{y}_n, \mathbf{w}_n) - (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(2\hat{\mathbf{x}}_{p(n)} - \mathbf{1})^\top (\mathbf{r}_n - \hat{\mathbf{x}}_{p(n)}) \\ & - (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n)(2\hat{\mathbf{a}}_n - \mathbf{1})^\top (\mathbf{w}_n - \hat{\mathbf{a}}_n) \\ = & \min_{\substack{\mathbf{x}_n, \mathbf{y}_n, \\ \mathbf{r}_n, \mathbf{w}_n: (1b), (6c)}} \bar{f}_n(\mathbf{x}_n, \mathbf{y}_n, \mathbf{w}_n) + (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n) \|\hat{\mathbf{x}}_{p(n)} - \mathbf{r}_n\|_1 \\ & + (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n) \|\hat{\mathbf{a}}_n - \mathbf{w}_n\|_1 \\ = & \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n), \end{aligned}$$

where the second equality follows from the definition of π and τ , the third equality is because $(2x - \mathbf{1})^\top (y - x) = 2x^\top y - \mathbf{1}^\top x - \mathbf{1}^\top y = -\|x - y\|_1$ whenever x, y are binary vectors, and the last equality is because the objective function is at least $L_n + (\underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - L_n) = \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n)$ if either $\mathbf{w}_n \neq \hat{\mathbf{a}}_n$ or $\mathbf{r}_n \neq \hat{\mathbf{x}}_{p(n)}$, implying that $\mathbf{w}_n = \hat{\mathbf{a}}_n$ and $\mathbf{r}_n = \hat{\mathbf{x}}_{p(n)}$ at optimum. It follows that $\mathcal{L}_n(\pi, \tau) = \underline{Q}_n(\hat{\mathbf{x}}_{p(n)}, \hat{\mathbf{a}}_n) - \pi^\top \hat{\mathbf{x}}_{p(n)} - \tau^\top \hat{\mathbf{a}}_n = \omega$, which finishes the proof. $\square$

A.5 Two-Stage Line Switching Model

Within the SND algorithm, the complete set of variables is decided at every stage $t \in [T]$. We can consider a more general decision setup, where a set of variables is decided at the root node (first stage), and implemented in the later stages, with other variables being decided at every stage. For exposition purposes, we will consider a problem with two-stage line switching decisions and dynamic power generation and power flow. Thus, we will only consider a set of line switching decisions for each stage $t \in [T]$ that will be used at nodes in the stage node set $\mathcal{S}_t$.

We can modify the current formulation by extending the state variable vector $\mathbf{x}$ for the root node at stage $t = 1$ to considering all switching decisions in the planning horizon.

Given that we obtaining dual information of every stage switching, we propagate the dual information beyond subsequent stages to plan the switching for all stages at the first stage. We therefore, can consider the use of additional fishing variables and constraints for this propagation. In this way, in the forward pass, at the root node, we obtain the switching decisions for next stages, and when solving the sampled paths, we fix all switching decisions beyond one later stage, as done in the standard SND algorithm. At the backward step, at any node $n \in \mathcal{T} \setminus \{1\}$, we solve the problem:

$$\begin{aligned} & \min_{\mathbf{x}_n, \mathbf{s}_t, \mathbf{y}_n, \psi_m, \phi_m, \mathbf{r}, \mathbf{w}} f_n(\mathbf{x}_n, \mathbf{y}_n) + \sum_{m \in C(n)} q_{nm} (\psi_m^\top \beta_m \mathbf{x}_n + \psi_m^\top \gamma_m + \phi_m) \\ & \text{s.t.} \quad A_n \mathbf{x}_n + W_n \mathbf{y}_n + C_n \mathbf{x}_{p(n)} + D_n \tilde{\mathbf{a}}_n \geq h_n \\ & \quad \mathbf{r}_n = \mathbf{x}_n \\ & \quad \mathbf{s}_i = \mathbf{z}_i, \forall i = t, \dots, T \end{aligned} \tag{14a} \tag{14b}$$

$$\mathbf{w}_n = \tilde{\mathbf{a}}_n \quad (14c)$$

$$\phi_m \geq \max_{\mathbf{a}} \left\{ (\pi_m^h)^\top \mathbf{x}_n + \sum_{i=t}^T (\eta_m^h)^\top \mathbf{s}_i + (\tau_m^h - \psi_m)^\top \mathbf{a} + \omega_m^h \right\}, \forall h \in \mathcal{H}. \quad (14d)$$

A.6 Initial Upper Bound Computation

Algorithm 2 requires a big- $\mathcal{M}$ coefficient that serves as an upper bound to have a valid deterministic upper bound via vertex enumeration. We can consider the worst-case realization of the line survival vector as a tighter initial upper bound. Having all failure-prone lines to fail is the worst-case realization for the uncertainty. During the first iteration of Algorithm 2, we run a forward pass by setting all failure-prone lines in $\mathbf{a}$, unavailable. We note that we construct a scenario tree that encodes the nodal information of the ambiguity set, but does not define the values of the matrices A, W, C, D or vector h , which are only stage-dependent. Thus, if we consider the summation of the objective over all stages in this deterministic forward pass, we can set it as the big- $\mathcal{M}$ for all nodes in the corresponding stage. Furthermore, we can employ these trial points to generate an initial set of hyperplanes to initialize the outer approximation of the cost-to-go function for all nodes in the corresponding stage. Our numerical experiments use this procedure to compute this worst-case upper bound during the first iteration of the algorithm.

A.7 Construction of easy-to-implement policies

Topology policy For each time unit $t \in [T]$, the *Topology* policy chooses a transmission grid topology from a set of topologies. Algorithm 4 describes how to obtain the set of topologies through the SND algorithm.

Algorithm 4 Construction of topology sets for the *Topology* policy

- 1: **Input:** optimal solutions $\{\mathbf{z}_n^*, n \in \mathcal{T}\}$ of the SND algorithm, $\text{Top}_t \leftarrow \emptyset$ for all $t \in [T]$;
 - 2: **for** $t = 1, \dots, T$ **do**
 - 3: Sort the nodes $\mathcal{S}_t$ in a decreasing order of their probabilities of occurrence and denote the permutation as $< 1 >, \dots, < |\mathcal{S}_t| >$;
 - 4: **for** $n = 1, \dots, |\mathcal{S}_t|$ **do**
 - 5: **if** $\text{Top}_t = \emptyset$ or $\min_{\mathbf{z} \in \text{Top}_t} \{\|\mathbf{z} - \mathbf{z}_{< n >}^*\|_1 / \|\mathbf{z}\|_1\} \geq 0.05$ **then**
 - 6: $\text{Top}_t \leftarrow \text{Top}_t \cup \{\mathbf{z}_{< n >}^*\}$;
 - 7: **end if**
 - 8: **end for**
 - 9: **end for**
 - 10: Return Top_t for all $t \in [T]$;
-

Intuitively, Algorithm 4 greedily collects a subset of optimal grid topologies from the SND algorithm. We require that these topologies are sufficiently different from each other to obtain a diverse candidate set for the *Topology* policy. Figure 6 depicts a set of three topologies used in this policy for the grid in Figure 3.

Mapping policy For each time unit $t \in [T]$ and for each switchable line $l \in \mathcal{L}$, the *Mapping* policy seeks to establish a deterministic look-up table mapping to the switching decision z_{lt} from the current state $z_{l,t-1}$ of line l , as well as the availability $\tilde{\mathbf{a}}_t^{\mathcal{L}(l)}$ of the redundant lines in $\mathcal{L}(l)$. We establish the *Mapping* policy by sampling the scenario tree in multiple replications and, in each replication, retrieving the *Dynamic* policy for z_{lt} under different scenarios but the same $(z_{l,t-1}, \tilde{\mathbf{a}}_t^{\mathcal{L}(l)})$ input. Then, we round the average of these z_{lt} to either zero or one. We detail this approach in Algorithm 5.

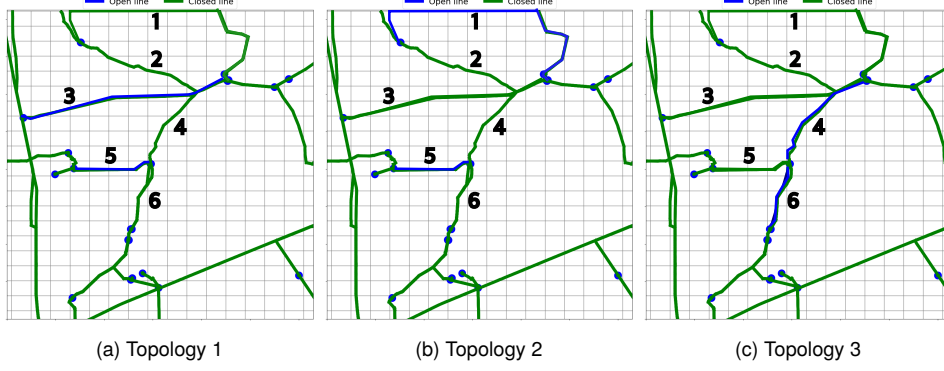


Figure 6: Example topologies for subregion in Figure 3

Algorithm 5 Mapping policy for transmission line l

- 1: Input: set of redundant lines $\mathcal{L}(l)$, number of replications I ;
 - 2: **for** $i = 1, \dots, I$ **do**
 - 3: Sample a scenario ω^i from $\mathcal{T}$;
 - 4: **for** $t = 1, \dots, T$ **do**
 - 5: Solve formulation (1) pertaining to node $n \in \mathcal{S}_t \cap \omega^i$ and state $(\mathbf{x}_{t-1}^i, \tilde{\mathbf{a}}_t^i)$ with $Q_m(\cdot, \cdot)$ replaced by $\underline{Q}_m(\cdot, \cdot)$ and store solution $(\mathbf{x}_t^i, \mathbf{y}_t^i)$;
 - 6: Sample $\tilde{\mathbf{a}}_{t+1}^i$ from the worst-case distribution $\mathbb{P}^*$;
 - 7: **for** all $S \in \{0, 1\} \times 2^{\mathcal{L}(l)}$ **do**
 - 8: Construct a state $(\mathbf{x}'_{t-1}, \tilde{\mathbf{a}}'_t)$ by replacing the entries of $(z_{l,t-1}, \tilde{\mathbf{a}}^{\mathcal{L}(l)})$ in $(\mathbf{x}_{t-1}^i, \tilde{\mathbf{a}}_t^i)$ by S ;
 - 9: Solve formulation (1) pertaining to $(\mathbf{x}'_{t-1}, \tilde{\mathbf{a}}'_t)$ with $Q_m(\cdot, \cdot)$ replaced by $\underline{Q}_m(\cdot, \cdot)$ and store the line switching decision $z_{lt}^i(S)$;
 - 10: **end for**
 - 11: **end for**
 - 12: **end for**
 - 13: Return mapping function $\text{Map}_{l,t} : \{0, 1\} \times 2^{\mathcal{L}(l)} \rightarrow \{0, 1\}$ with $\text{Map}_{l,t}(S) := \text{Round}(\sum_{i=1}^I z_{lt}^i(S)/I)$ for all $t \in [T]$;
-

A.8 Instance details

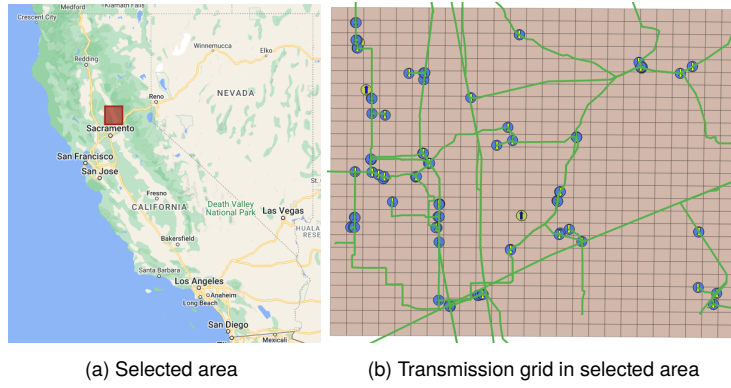


Figure 7: Area and topology of test instance . Buses are marked in blue circles, generator in green circles, and transmission lines are marked in green.

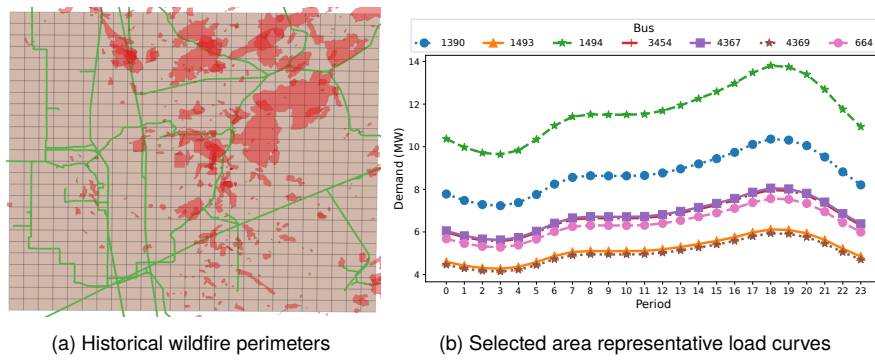


Figure 8: Wildfire and load data for selected area