PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DATA-DRIVEN OPTIMIZATION FOR SEISMIC-RESILIENT POWER NETWORK PLANNING

ALFREDO ONETO SCHIAPPACASSE

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
ÁLVARO LORCA GÁLVEZ

Santiago de Chile, Julio 2021

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DATA-DRIVEN OPTIMIZATION FOR SEISMIC-RESILIENT POWER NETWORK PLANNING

ALFREDO ONETO SCHIAPPACASSE

Members of the Committee:
ÁLVARO LORCA GÁLVEZ
MATÍAS NEGRETE-PINCETIC
BENJAMÍN MALUENDA PHILIPPI
GONZALO YAÑEZ CARRIZO

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Santiago de Chile, Julio 2021

Grateful for all the people who
accompanied me on this journey.

ACKNOWLEDGMENTS

Throughout the research of this thesis, I have received tremendous support from many people. I would first like to thank my supervisor, Ph.D. Álvaro Lorca, whose outstanding dedication and vast expertise have been crucial. I would like to acknowledge the collaboration of Ph.D. Elisa Ferrario and M.Sc. Alan Poulos, for their valuable guidance and assistance concerning seismic resilience studies. I would also like to thank Ph.D. Juan Carlos De La Llera and Ph.D. Matías Negrete-Pincetic for their wise council. Finally, I would like to thank my family and friends for keeping me motivated until the very last day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Many regions of the planet are exposed to seismic risks that can have devastating consequences in electric power systems. These systems' crucial role in modern society makes the evaluation and planning for their safe and reliable operation paramount. In this context, this thesis develops a novel data-driven optimization framework to assess the power network seismic resilience and to plan investments for improving its operational cost under contingencies. Under a robust optimization scheme, an earthquake attacker-defender model finds worst-case network contingencies based on the statistical properties of seismic scenarios generated with a state-of-the-art seismic engineering method and representing the recovery process of electric power systems' components. Additionally, data-driven stochastic-robust optimization is employed in a two-stage seismic-resilient power network planning model, which decides transmission line expansions and siting and sizing energy storage systems in the first stage, while the second stage makes optimum operational decisions. Further, this model leverages distributional information of multiple seismic sources and yields a bi-level optimization structure, which has an outer stochastic programming optimization problem that minimizes the expected objective values of the respective earthquake attacker-defender problem for each seismic source based on investment decisions. Subsequently, extensive computational experiments on a 281-node representation of the Chilean electric power system provide meaningful insights for seismic-resilient planning and demonstrate the efficiency of the solution approach.

Keywords: Operations research in energy, Data-driven optimization, Power system resilience, Seismic hazard.

RESUMEN

Muchas regiones del planeta están expuestas a riesgos sísmicos que pueden tener consecuencias devastadoras en los sistemas de energía eléctrica. El papel crucial de estos sistemas en la sociedad moderna hace que la evaluación y la planificación de su funcionamiento seguro y fiable sean primordiales. En este contexto, este trabajo desarrolla un novedoso marco de optimización basado en datos para evaluar la resiliencia sísmica de la red eléctrica y planificar las inversiones para mejorar su coste operativo en caso de contingencias. Bajo un esquema de optimización robusta, un modelo atacante-defensor de terremotos encuentra las peores contingencias de la red basándose en las propiedades estadísticas de los escenarios sísmicos generados con un método de ingeniería sísmica de última generación y representando el proceso de recuperación de los componentes de los sistemas de energía eléctrica. Además, se emplea la optimización estocástica basada en datos en un modelo de dos etapas para la planificación sísmico-resiliente de redes eléctricas, que decide las expansiones de las líneas de transmisión y la ubicación y el tamaño de los sistemas de almacenamiento de energía en la primera etapa, mientras que en la segunda etapa se toman decisiones operativas óptimas. Además, este modelo aprovecha la información distribucional de múltiples fuentes sísmicas y produce una estructura de optimización de dos niveles, que tiene un problema de optimización de programación estocástica externa que minimiza los valores objetivos esperados del respectivo problema de atacante-defensor contra los terremotos para cada fuente sísmica basada en las decisiones de inversión. Posteriormente, extensos experimentos computacionales en una representación de 281 nodos del sistema de energía eléctrica chileno proporcionan información significativa para la planificación sísmica-resiliente y demuestran la eficiencia del enfoque de solución.

Palabras Clave: Investigación operativa en energía, Optimización basada en datos, Resiliencia de sistemas de potencia, Amenaza sísmica.

1. INTRODUCTION

Natural disasters can potentially devastate the energy supply infrastructure, resulting in severe economic losses (Dilley, 2005; Castillo, 2014). Among high-impact and low-probability (HILP) events, earthquakes are considered one of the gravest, given their low predictability and damaging effects on power network systems (Y. Wang et al., 2016). For instance, the earthquake in the Chinese Wenchuan county on May 18 of 2008 damaged approximately 900 substations and 270 transmission lines (Eidinger, 2009). Moreover, the 8.8 Mw earthquake in Chile that occurred on February 27 of 2010 caused an immediate blackout in the Chilean Central Interconnected System, which, at that time, provided electricity to over 93% of the country's population (Araneda et al., 2010). Furthermore, on March 11, 2011, the Tohoku earthquake in Japan damaged 70 transformers, 14 power plants, 42 transmission towers in conjunction with other facilities (Shumuta, 2011).

Many countries have their power network systems exposed to significant seismic threats, so their seismic resilience capabilities are of concern. In this work, *resilience* is understood as "the ability of a system to resist the effects of a disruptive force and to reduce performance deviations" (Nan & Sansavini, 2017). Then, it is paramount to have a planning framework capable of assessing the uncertain nature of such hazards to find investments that optimizes the seismic resilience of power systems in areas prone to suffering grievous losses caused by earthquakes. In this regard, the objective of this work is to develop a data-driven optimization planning scheme of this kind. In what follows, we present a brief literature survey of previous works related to this article.

1.1. Seismic modeling

A comprehensive assessment of the seismic-induced damage and recovery of power network components must include a ground-shake intensity characterization and the evaluation of the fragility and restoration of vulnerable components (Espinoza et al., 2020).

Monte Carlo Simulations are suitable to perform this type of assessment (Poulos et al., 2017). For simulations, earthquake sources and their magnitudes' distribution have to be identified, as explained in Baker (2015). Then, the ground motion intensities can be computed via a Ground Motion Prediction Model. Further, some techniques have been developed to sample these scenarios efficiently (e.g., Jayaram & Baker (2010)). Further, to determine a component's functionality after an earthquake, given an intensity measure of ground motion, independent fragility and restoration curves are used (FEMA, 2013).

1.2. Resilient operation and planning of power networks

Resilience evaluation and enhancement of power network systems performance under components' failure have been vastly researched (Koç et al., 2014; Babaei et al., 2020; Yuan et al., 2014; Bagheri et al., 2019; Yan et al., 2020). On the one hand, resilient planning models are used to prepare those systems to withstand contingencies better. On the other hand, operational models serve to determine emergency response actions and spot worst-case disruptions (Tapia et al., 2021). In the remainder of this section, we briefly introduce the reader to planning and operational models of power network systems and present some applications for resilience decision-making.

In what concerns energy policy design and investments, expansion planning models play a crucial role in optimizing their consequences in the evolution and future performance of power systems (Gacitua et al., 2018). Consequently, a wide variety of such models are found in the power systems literature (Velloso et al., 2020; Maluenda et al., 2018; Verástegui et al., 2019), including some related models of seismic-resilient planning. Romero et al. (2013) developed a two-stage stochastic optimization modeling for expansion planning to mitigate seismic risk, involving transmission lines capacity and power generation enlargements. Lagos et al. (2019) presented another two-stage stochastic optimization formulation of seismic-resilient expansion planning, which considers generation and transmission investments and substation hardening. Nazemi et al. (2019) introduced a

two-stage stochastic optimization model for energy storage planning for seismic resiliency enhancement of distribution networks. P. Zhao et al. (2020) proposed a distributionally robust optimization model, based on moment information of damage contingencies, for the planning of hardening investments for transmission lines and gas pipelines against earthquakes.

The economical and reliable operation of power systems is a crucial task, and the literature about these problems is wide-ranging (Q. Wang et al., 2013; Lorca & Sun, 2016; Zhang et al., 2016; Duan et al., 2018; Zheng & Chen, 2020). Among those problems, the optimal power flow problem is fundamental, given that it minimizes operational costs deciding optimal power dispatch actions subject to several constraints imposed by engineering limits and physical laws (Frank & Rebennack, 2016). Further, when the optimal power flow problem is embedded in a bi-level interdiction framework with a max-min structure, characteristic of robust optimization, the problem is known as an attacker-defender model (AD), which aims to find the worst-case realization of disruptive contingencies such as terrorist events, hurricanes, or wildfires (Salmeron et al., 2004; Yuan et al., 2016; Tapia et al., 2021). Motivated by this, in this work we consider a data-driven seismic set for an attacker-defender interdiction model, which can determine the worst earthquake scenarios out of the seismic nature under which the power network is exposed; thus, sheds light on the vulnerabilities of the system when faced to this type of hazard.

1.3. Proposed framework and contributions

Notwithstanding recent advancements, some relevant gaps remain in the literature regarding assessing and improving the seismic resilience of power networks. For instance, existing seismic resilience assessment frameworks usually do not explicitly model the recovery of components, and none of them accounts for the most critical potential earthquake-induced simultaneous failures of various components. Regarding seismic-resilient planning models, they often rely on hardening approaches, which means that

it is possible to enhance an asset to make it indestructible or much harder to damage, potentially neglecting the intrinsically fragile nature of different infrastructure types. These planning schemes tend to be inefficient for large-scale networks because of their computational burden or overcome this difficulty by employing a small number of earthquake scenarios, ignoring a detailed representation of the uncertainty representation of the hazard. Further, no seismic-resilient planning framework has yet leveraged the distributional information about seismic sources.

The underlying seismic-related contingency uncertainty poses notable challenges when modeling for resilience assessment and planning for the robustness of power networks; hence, suitable techniques for managing this problem are needed. For managing uncertainty in decision-making models, the data-driven optimization paradigm has proved to be befitting because it relies on uncertainty models formulated based on data, as they systematically harness information from realizations of uncertain parameters for decision processes (Ning & You, 2019).

The present thesis aims to fill the gaps previously mentioned by proposing a novel data-driven optimization framework for seismic-resilience assessment and planning. The main contributions of this work are summarized in the following paragraphs.

The development of a data-driven earthquake attacker-defender model, which accounts for temporal and spatial correlation of seismic-induced damage of components. This max-min optimization structure finds the worst simultaneous failures produced by seismic hazards, an unanswered question thus far. Moreover, as the problem handles high-dimensional random vectors composed of correlated variables, we define an uncertainty set accounting for moment information of the data while maintaining computational tractability. In this regard, previous solutions have been proposed (Ning & You, 2018a; Shang & You, 2018). However, uncertainty set construction for high-dimensional data is still an open problem. This work develops a novel polyhedral distance-based uncertainty set with the following

characteristics: it incorporates information of the first and second moment of the empirical distribution of data, provides the means of performing dimensionality reduction with a guarantee of explained variance percentage, has an adjustable parameter for the set's size which is directly obtained as a function of the covered training data samples, and includes clustering techniques for reducing the number of points required to define the set. Further, computational experiments confirm its efficiency. All these qualities make the uncertainty set widely applicable.

A seismic-resilient power network planning model is proposed based on data-driven stochastic-robust optimization (DDSRO), which decides optimum investments in energy storage systems and transmission lines expansion. The DDSRO model yields a computationally tractable two-stage structure with stochastic programming nested in the outer level to optimize the expected objective over adaptive robust optimization problems for each label nested in the inner problem (Yue & You, 2016; Ning & You, 2018b). When the uncertainty data of a process is categorically labeled, such as earthquake contingency scenarios of power networks according to their seismic source, the DDSRO framework can be used as a way of leveraging the probability distribution of seismic sources while capturing the structure and complexity of uncertainty data within the same source employing uncertainty sets. Thus, the first stage decides optimum investment decisions to minimize the weighted average over the objective of the respective earthquake attacker-defender problem for each seismic source, where each weight corresponds to the probability of observing an earthquake of the given seismic source. To the best of our knowledge, this novel modeling scheme is the first to harness distributional information of multiple seismic sources in a seismic-resilient planning model.

In extensive computational experiments, based on a 281-node representation of the Chilean power network, we analyze the solutions and computational performance for both the earthquake attacker-defender model and the DDSRO planning model. A methodology for efficiently solving these large-scale optimization problems is proposed. The results

obtained provide meaningful insights. On the one hand, the earthquake attacker-defender model, with its adjustable-size uncertainty set, identifies worst-case contingencies for every seismic source and different risk levels. Further, we show that planners can map these solutions with geographical information systems, so decision-makers can easily visualize critical parts of the power network. On the other hand, we solved the DDSRO model for several combinations of budgets and conservativeness levels, and the results exhibit the following characteristics under our case study conditions:

- Unfulfilled power demand diminishes with investments from 2.3% up to 10.5% compared to the no-investment solution with other budgets for the same risk level.
- All solutions show that investment in transmission lines is preferred over energy storage systems.
- Different investment configurations share no obvious preferable assets. Hence, the capacity of our framework for handling different risk levels proves to be helpful because there are no trivial one-fits-all solutions.
- Optimum investment decisions tend to be more diversified while the risk level increases, as a diversification index attests.

The remainder of the thesis is organized as follows. Section 2 summarizes the nomenclature to be used. Section 3 presents the seismic uncertainty modeling approach adopted in this work. Section 4 develops the optimization models that constitute our seismic-resilience power network planning framework. Section 5 defines a solution methodology for the optimization models. Section 6 shows a Chilean case study. Finally, Section 7 presents concluding remarks and future works.

2. NOMENCLATURE

The nomenclature of the optimization models is provided within this section.

2.1. Sets and Indexes

$b \in \mathcal{B}$**:** Index and set of substations.

$d \in \mathcal{D}$**:** Index and set of representative days.

$i \in \mathcal{G}$**:** Index and set of power generation units.

$\mathcal{G}^F \subset \mathcal{G}$**:** Set of power units excluding solar and wind technologies.

$\mathcal{G}^V \subset \mathcal{G}$**:** Set of wind and solar power units.

$h \in \mathcal{H}$**:** Index and set of time blocks of the representative day.

$k \in \mathcal{K}$**:** Index and set of points for uncertainty modeling.

$j \in \mathcal{L}$**:** Index and set of transmission lines.

$\mathcal{L}^C \subset \mathcal{L}$**:** Set of transmission lines candidate to expansion.

$\mathcal{L} \setminus \mathcal{L}^C \subset \mathcal{L}$**:** Set of transmission lines non-candidate to expansion

$q \in \mathcal{Q}$**:** Index and set of dimensions of uncertainty points.

$s \in \mathcal{S}$**:** Index and set of seismic sources.

$t \in \mathcal{T}$**:** Index and set of recovery stages.

2.2. Parameters

2.2.1. Operational parameters

B_j**:** Susceptance of line j.

D_{bdh}**:** Power demand at substation b, representative day d and time block h.

$\overline{f}_j$**:** Installed maximum flow capacity of line j.

H**:** Number of time blocks describing representative days.

$\overline{\theta}$**,** $\underline{\theta}$**:** Maximum and minimum phase angle.

p_d^D**:** Weight of representative day d.

$\overline{P_i^f}$**:** Maximum power capacity of non-solar or wind units.

p^H**:** Duration of time blocks.

$\overline{P_{dhi}^v}$**:** Maximum power that a wind or solar unit can produce according to a representative profile.

η^{cha}**,** η^{dis}**:** Charging and discharging efficiencies of energy storage systems.

τ_t**:** Duration of recovery stage.

2.2.2. Earthquake attacker parameters

a_q^k**:** Data point k at entry q.

β^k**:** Weight of point k.

$\overline{d}$**,** $\underline{d}$**:** Shape parameters of an uncertainty set.

δ**:** Lines disconnection threshold.

Δ**:** Size parameter of an uncertainty set (depends on external parameter γ).

$\boldsymbol{\mu}$**:** Centroid of an uncertainty set.

$\boldsymbol{W}^{-1}$**:** Inverse of the whitening matrix.

2.2.3. Investment parameters

c_j^l**,** c_b^p**:** Investment costs in line expansions and energy storage systems.

ρ**:** Technical relation between storage and capacity of an energy storage system.

2.3. Decision Variables

2.3.1. Operational variables

P_{bdht}^{cha}**:** Power charged to the energy storage system.

P^{dis}_{bdht}: Power discharged to the energy storage system.

P^{g}_{dhit}: Power generated by the given unit.

P^{LS}_{bdht}: Load shedding at the given substation.

f_{dhjt}: Power flow at line j.

θ_{bdht}: Voltage angle at substation b.

w_{bdht}: Energy stored in the energy storage system at bus b.

2.3.2. Earthquake attacker variables

d_{kq}: Distance variable for uncertain point.

φ_{bt}: Substation functionality derate.

ω_q: Entry at dimension q for generated uncertainty point.

ζ_{jt}: Line capacity derate

z_{jt}: Connection state of the given line.

2.3.3. Investment variables

$\overline{E_b}$: Energy storage capacity.

$\overline{P_b}$: Power capacity of the energy storage system.

x^l_j: Power flow capacity expansion of the line.

3. SEISMIC UNCERTAINTY MODELING

The main objective of this work is to determine seismic-resilient investment plans for the power network. Therefore, it is essential to characterize the nature of seismic risk appropriately. To achieve an adequate representation of the seismic hazards to which the power network is exposed, we generate an extensive set of seismic scenarios that feed the optimization models developed in this thesis with realizations of uncertain damage parameters. We use a simulation scheme that allows us to determine the damage condition of the network's components from the occurrence of an earthquake until the complete recovery of the network. This scheme is summarized in two steps: 1) generation of seismic scenarios and 2) computation of components' damage and recovery through time. Section 3.1 presents step 1) of the framework, and Section 3.2 describes step 2).

3.1. Generation of earthquake scenarios

The seismic nature of a given area can be characterized by its seismic sources and a catalog of simulated earthquake scenarios. A seismic source can be a fault, a typically planar surface, or an areal region where earthquakes may occur anywhere, and for which it is possible to identify its distribution of magnitudes and source-to-site distances associated with its earthquakes Baker (2015). To clarify what a seismic source is, we show the example of Chilean subduction seismic sources in Figure 3.1, based on the zonation from Poulos et al. (2019), with antecedents in Leyton et al. (2009); Martin. (1990). In our context, these scenarios are defined by peak ground acceleration (PGA) at the location of each fragile component of the network. To generate synthetic earthquake scenarios, we adopt the methodology previously presented in Poulos et al. (2017). This part briefly revisits the methodology.

Figure 3.1. Chilean subduction seismic sources.

This methodology calibrates an earthquake recurrence model for each seismic source capable of producing damaging ground motions, based on an exhaustive catalog of historical earthquakes. Recurrence models represent main shocks as homogeneous Poisson processes and enable the estimation of seismic hazard as the annual frequency of exceedance for some intensity measure threshold. For further descriptions of the calibration of these types of models, see Poulos et al. (2019). In addition, to efficiently sample earthquakes of high moment magnitudes, importance sampling is used, a technique for sampling random variables through an alternate density function (Owen & Zhou, 2000).

To estimate PGA at the locations of interest, given the location and the moment magnitude (Mw) of an earthquake, a ground-motion prediction model (GMPM) is required, such as the one presented in Parker et al. (2020) for subduction-type seismicity. Further, since

power networks are spatially distributed systems, the correlations of ground motions must be considered. For this purpose, we use the methodology developed in Goda & Atkinson (2010).

The procedure to generate J seismic scenarios is as follows:

- Based on a recurrence model and sampling techniques, simulate magnitudes and locations of seismic events to build a synthetic earthquake catalog. The sample weight is indicated in a parameter α^j, with $\sum_{j\in\mathcal{J}} \alpha^j = 1$, where $\mathcal{J} = \{1, \ldots, J\}$. A label parameter l^j for the j-th scenario shows its seismic source, which can take a value from a given set $\mathcal{S}$.
- Then, using the GMPM, compute the PGA at each fragile component's location for every sampled earthquake. In addition, the PGA vectors, their labels, and their weights, are stored as a collection of the form $\{\boldsymbol{PGA}^j, \alpha^j, l^j\}_{j\in\mathcal{J}}$, where $\boldsymbol{PGA}^j$ is the vector with the PGA at each site in the j-th scenario.

3.2. Component vulnerability assessment

Following the seismic risk assessment methodology developed in FEMA (2013), the recovery process of each vulnerable component is determined through functionality recovery functions for all the seismic scenarios previously obtained. In brief, this methodology states that the available capacity of a vulnerable component in a given recovery stage is computed as follows:

$$FA(e, p, t) = \sum_{i=1}^{n} FR(e, d_i, t) FD(e, d_i, p). \tag{1a}$$

The available capacity is $FA(e, p, t)$ presented in (1a), for which its arguments are: e, the component's infrastructure type; p, the PGA at the component's site; and t, the recovery stage of interest. Considering that after the occurrence of an earthquake, a component

can be in one of the possible discrete damage states d_i for $i \in \{1, \ldots, n\}$, the probability of a component of being in a damage state d_i is $FD(e, d_i, p)$, and the probability of recovery of a component under that damage state is $FR(e, d_i, t)$. These two functions are defined below:

$$FR(c, d_i, t) = \mathbb{P}(R \leq t | DS = d_i, CT = e) \tag{1b}$$

$$FD(c, d_i, p) = \mathbb{P}(DS = d_i | PGA = p, CT = e) \tag{1c}$$

The equations (1a)-(1c) are defined through conditional probabilities, where four random variables describe the conditions of the component under study: R, the recovery time [1]; DS, the damage state after the earthquake; CT, the infrastructure type; and PGA, the suffered peak ground acceleration. The equation (1b) is computed with restoration curves, and (1c) with fragility curves. For the sake of simplicity, we omit further details. See FEMA (2013) for further descriptions of this procedure.

The focus of this work is placed on power systems' resilience since earthquakes can produce widespread contingencies. Hence, the general vulnerability assessment model presented in (1a) - (1c) provides the means for evaluating seismic impacts in power systems' infrastructure and representing their recovery processes. Moreover, to perform this evaluation, vulnerable components have to be identified. In this work, we model substations as vulnerable components. In contrast, all other components are considered indestructible. This assumption avoids some non-linear constraints in optimization models, and recent experience of a high-impact earthquake justifies it given that circuit towers and power plants did not place the energy supply at risk, in opposition to the substations (Rudnick et al., 2011) [2].

[1] In this context, it is assumed that the decision-maker cannot control this variable.

[2] They observed that transmission lines resisted well to strong ground motions. In addition, while generation units were damaged, they did not risk the connectivity of the system.

The recovery process of the network is modeled through a set of recovery stages as $\mathcal{T} = \{1, \ldots, T\}$ (Romero et al., 2013). Then, from the seismic scenarios $\{\boldsymbol{PGA}^j, \alpha^j, l^j\}_{j\in\mathcal{J}}$, we can compute damage scenarios with the following equation:

$$1 - \varphi_{bt}^j = FA(e_b, PGA_b^j, t) \quad \forall b \in \mathcal{B}, t \in \mathcal{T}, j \in \mathcal{J} \tag{2}$$

In equation (2), sets $\mathcal{B}$, and $\mathcal{J}$ represent the set of substations, and the set of scenarios, respectively. The parameter e_b is the infrastructure type of the substation b (i.e., substations with anchored or unanchored parts); PGA_b^j is the peak ground acceleration in the location of the substation b under the j-th seismic scenario [3]. Further, the parameter φ_{bt}^t indicates the capacity derate of substation b in recovery stage t under the j-th scenario. As operational models of electric power systems do not explicitly present a capacity value for substations, the derate is imposed on the capacity values of the elements attached to substations: generating units, transmission lines, and energy storage systems; as done by Lagos et al. (2019).

Once each value of φ_{bt}^j is obtained with (2), the seismic-induced damage scenarios are stacked in a collection of the form $\{\boldsymbol{\varphi}^j, \alpha^j, l^j\}_{j\in\mathcal{J}}$. These scenarios are employed in the data-driven optimization schemes developed in the next section.

[3]Note that estimates of local soil conditions of components may distort the PGA results significantly.

4. OPTIMIZATION MODELS

This Section develops novel optimization models for seismic-resilient power network planning. Section 4.1 describes the operational model for the power network under a given seismic-induced damage scenario and fixed investment decisions. Section 4.2 develops an *earthquake attacker-defender model* that yields a data-driven worst-case earthquake scenario under a given investment plan, using a distance-based uncertainty set for robust optimization, particularly suited for modeling high-dimensional uncertainty. Then, Section 4.3 builds a data-driven stochastic robust optimization planning model, which decides the optimal investment in transmission lines' capacity expansion and siting and sizing of energy storage systems incorporating information from multiple seismic sources.

4.1. Operational model

A crucial problem for power systems consists of finding optimal operational decisions to minimize overall operating costs. The system's operator computes the best decisions for the economic functioning of the system by solving what is known as the optimal power flow problem. These decisions involve some control variables (e.g., power injection in a node), whereas state variables that depend on the control variables value (e.g., voltage angles), while their relationships are defined through electrical laws and engineering limits (Frank & Rebennack, 2016). In this work, the optimal power flow is used for modeling the operation of the system after an earthquake, considering the damaged state of infrastructure and capturing notions of the recovery process of the network.

The operational model computes the minimum operational cost under certain investment decisions $\boldsymbol{x}$ and a given vector of damage $\boldsymbol{\xi}$, where $\boldsymbol{x} = (\overline{\boldsymbol{E}}, \overline{\boldsymbol{P}}, \boldsymbol{x}^l)$ and $\boldsymbol{\xi} = (\boldsymbol{\varphi}, \boldsymbol{\zeta}, \boldsymbol{z})$. The operation horizon consists of several recovery stages, where a set of representative days define the operational conditions (i.e., maximum renewable generation and demand) for each stage and where each representative day has a set of chronologically

ordered time points. Hence, we consider the following linear formulation for the system's operation, which chooses the optimal operation decisions $\boldsymbol{y} = (\boldsymbol{P}^{cha}, \boldsymbol{P}^{dis}, \boldsymbol{P}^{g}, \boldsymbol{P}^{LS}, \boldsymbol{f}, \boldsymbol{\theta}, \boldsymbol{w})$,

$$Q(\boldsymbol{x}, \boldsymbol{\xi}) = \min_{\boldsymbol{y}} \sum_{t \in \mathcal{T}} \tau_t \Big(\sum_{d \in \mathcal{D}} p_d^D \Big(\sum_{h \in \mathcal{H}} \frac{L(\boldsymbol{y}_{dht})}{H} \Big) \Big) \tag{3a}$$

$$\text{s.t.} \Big\{ 0 \le P^g_{dhit} \le \overline{P_i^f}(1 - \varphi_{b(i)t}) \quad \forall i \in \mathcal{G}^F \tag{3b}$$

$$0 \le P^g_{dhit} \le \overline{P^v_{dhi}}(1 - \varphi_{b(i)t}) \quad \forall i \in \mathcal{G}^V \tag{3c}$$

$$0 \le P^{LS}_{bdht} \le D_{bdh} \quad \forall b \in \mathcal{B} \tag{3d}$$

$$f_{dhjt} \ge -(1 - \zeta_{jt})\overline{f_j} \quad \forall j \in \mathcal{L} \setminus \mathcal{L}^C \tag{3e}$$

$$f_{dhjt} \le (1 - \zeta_{jt})\overline{f_j} \quad \forall j \in \mathcal{L} \setminus \mathcal{L}^C \tag{3f}$$

$$f_{dhjt} \ge -(1 - \zeta_{jt})(\overline{f_j} + x^l_j) \quad \forall j \in \mathcal{L}^C \tag{3g}$$

$$f_{dhjt} \le (1 - \zeta_{jt})(\overline{f_j} + x^l_j) \quad \forall j \in \mathcal{L}^C \tag{3h}$$

$$f_{dhjt} \ge -(1 - z_{jt})\overline{f_j} \quad \forall j \in \mathcal{L} \setminus \mathcal{L}^C \tag{3i}$$

$$f_{dhjt} \le (1 - z_{jt})\overline{f_j} \quad \forall j \in \mathcal{L} \setminus \mathcal{L}^C \tag{3j}$$

$$f_{dhjt} \ge -(1 - z_{jt})(\overline{f_j} + x^l_j) \quad \forall j \in \mathcal{L}^C \tag{3k}$$

$$f_{dhjt} \le (1 - z_{jt})(\overline{f_j} + x^l_j) \quad \forall j \in \mathcal{L}^C \tag{3l}$$

$$M z_{jt} \ge B_j(\theta_{b^-(j),h,d,t} - \theta_{b^+(j),h,d,t}) - f_{dhjt} \quad \forall j \in \mathcal{L} \tag{3m}$$

$$-M z_{jt} \le B_j(\theta_{b^-(j),h,d,t} - \theta_{b^+(j),h,d,t}) - f_{dhjt} \quad \forall j \in \mathcal{L} \tag{3n}$$

$$\underline{\theta} \le \theta_{bdht} \le \overline{\theta} \quad \forall b \in \mathcal{B} \tag{3o}$$

$$0 \le w_{bdht} \le \overline{E_b}(1 - \varphi_{bt}) \quad \forall b \in \mathcal{B} \tag{3p}$$

$$0 \le P^{cha}_{bdht} \le \overline{P_b}(1 - \varphi_{bt}) \quad \forall b \in \mathcal{B} \tag{3q}$$

$$0 \le P^{dis}_{bdht} \le \overline{P_b}(1 - \varphi_{bt}) \quad \forall b \in \mathcal{B} \tag{3r}$$

$$D_{bdh} - P^{LS}_{bdht} = \sum_{i \in \mathcal{G}(b)} P^g_{dhit} + P^{dis}_{bdht} - P^{cha}_{bdht} \tag{3s}$$

$$+ \sum_{j|b^+(j)=b} f_{dhjt} - \sum_{j|b^-(j)=b} f_{dhjt} \quad \forall b \in \mathcal{B}\Big\},$$

$$\forall d \in \mathcal{D}, h \in \mathcal{H}, t \in \mathcal{T}$$

$$\Big\{ w_{bdht} = w_{b,d,h-1,t} + p^H \left(P^{cha}_{b,d,h-1,t} \eta^{cha} - \frac{P^{dis}_{b,d,h-1,t}}{\eta^{dis}} \right) \tag{3t}$$

$$\forall h \in \mathcal{H} \setminus \{1\}$$

$$w_{b,d,h=1,t} = w_{b,d,h=H,t} + p^H \left(P^{cha}_{b,d,h=H,t} \eta^{cha} - \frac{P^{dis}_{b,d,h=H,t}}{\eta^{dis}} \right) \Big\}, \tag{3u}$$

$$\forall b \in \mathcal{B}, d \in \mathcal{D}, t \in \mathcal{T}$$

The objective function (3a) is linear over the operational variables and accounts for the operational cost over the evaluation's horizon. The loss function $L(\cdot)$ is chosen by the decision-maker. Parameter p^D_d represents the weight of the representative day d, H corresponds to the number of time points in each day, and τ_t is the time length in hours of the recovery stage t. Constraint (3b) establishes the minimum and maximum generation capacity for non-renewable units and (3c) for renewable units (with a given maximum generation profile). The inequalities (3d) limit the minimum and maximum load shedding up to the node demand level and the minimum as zero. Constraints (3e)-(3h) limit power flow considering partial damage, and (3i)-(3l) limit power flow considering the binary state that determines whether a line is connected or not. Equations (3m)-(3n) relate power flow variables with angle variables when such lines are connected, using a Big-M formulation [1]. Constraints (3o)-(3r) model the allowable range for voltage angles, energy storage, charging power, and discharging power. Load balance is ensured through (3s). Finally, (3t) imposes energy balance for energy storage systems for each time point except the first one, while (3u) relates the last time point storage level with the first one, under the assumption

[1] In this case, the Big-M formulation means that if variable z_{jt} takes the value of 0, the right-hand side of the constraints is equal to 0. If z_{jt} is equal to 1, the constant M is sufficiently large that the constraints become redundant. Thus, this formulation only relates voltage angles with flow variables if the condition $z_{jt} = 0$ is fulfilled.

that those systems are perfectly optimized for each given day of operation (Gan et al., 2019). Note that the modeling scheme of the optimal power flow is done through the DC approximation, usually adopted in the literature of performance assessment and planning power networks under seismic risk (Romero et al., 2013; Espinoza et al., 2020; P. Zhao et al., 2020).

4.2. The earthquake attacker-defender problem

In this part, we employ the previous operational model to build the earthquake attacker-defender problem, which represents the uncertain nature of vector $\boldsymbol{\xi}$ under a fixed investment decision $\boldsymbol{x}$. The purpose of the earthquake attacker-defender problem is to find the realization of $\boldsymbol{\xi}$ which results in the highest optimal system's cost within the operational horizon.

The earthquake attacker-defender problem can be represented as the following bi-level optimization problem:

$$\max_{\boldsymbol{\xi}\in\mathcal{U}} Q(\boldsymbol{x},\boldsymbol{\xi}) \tag{4}$$

The max-min problem (4) considers the worst-case realization of the uncertain parameters and the adaptive actions of the system. The objective function of (4) is a vectorial compact form of (3a). Random parameters $\boldsymbol{\xi}$ are chosen from $\mathcal{U}$, and consequently, this type of set is known as uncertainty sets (Bertsimas & Sim, 2004). Moreover, the set $\mathcal{Y}(\boldsymbol{x},\boldsymbol{\xi})$ is defined by operational restrictions (3b)-(3u).

The uncertainty set $\mathcal{U}$ alluded in (4) is defined as the intersection of the uncertainty set with structural constraints, namely $\mathcal{U}^{PH}$, and the data-driven distance-based uncertainty set that contains data about substations' damage, termed $\mathcal{U}^{DD}$; i.e., $\mathcal{U} = \{\boldsymbol{\xi} : \boldsymbol{\xi} \in \mathcal{U}^{DD}, \boldsymbol{\xi} \in \mathcal{U}^{PH}\}$. In the coming sections, we cover the definitions of $\mathcal{U}^{PH}$ and $\mathcal{U}^{DD}$.

4.2.1. Structural constraints' uncertainty set

To represent the substations' damage, some physical constraints must be incorporated. First of all, each substation has a damage indicator which is bounded between 0 and 1. Moreover, substations maintain their level of damage or reduce it over successive recovery stages. These two aspects are expressed as follows:

$$0 \leq \varphi_{bt} \leq 1 \qquad \forall b \in \mathcal{B}, t \in \mathcal{T} \tag{5a}$$

$$\varphi_{bt} \leq \varphi_{b,t-1} \qquad \forall b \in \mathcal{B}, t \in \mathcal{T} \setminus \{0\} \tag{5b}$$

where, (5a) bounds damage levels, and (5b) imposes recovery considerations.

Based on the substations' damage indicators, we model the transmission lines' damage states as follows. For any given line, its maximum power flow capacity is derated depending on the damage indicators of the two connecting substations, through the following equation:

$$\zeta_{jt} = 1 - \big(1 - \varphi_{b^-(j)t}\big)\big(1 - \varphi_{b^+(j)t}\big) \quad \forall j \in \mathcal{L}, t \in \mathcal{T}.$$

However, due to the bilinear nature of the above expression, we use McCormick convex envelopes (Mccormick, 1976), replacing the above expression with the following linear relaxation:

$$\zeta_{jt} \geq \varphi_{b^+(j)t} \qquad \forall j \in \mathcal{L}, t \in \mathcal{T} \tag{5c}$$

$$\zeta_{jt} \geq \varphi_{b^-(j)t} \qquad \forall j \in \mathcal{L}, t \in \mathcal{T} \tag{5d}$$

$$\zeta_{jt} \leq \varphi_{b^-(j)t} + \varphi_{b^+(j)t} \qquad \forall j \in \mathcal{L}, t \in \mathcal{T} \tag{5e}$$

$$\zeta_{jt} \leq 1 \qquad \forall j \in \mathcal{L}, t \in \mathcal{T} \tag{5f}$$

Note that the use of a linear relaxation through the above expressions yields a conservative approach for the overall attacked-defender model, in the sense that the feasible space for the attacker is enlarged.

Further, our model considers the possibility of completely disconnecting any given transmission line, changing the topology of the power network, for any line whose damage state surpasses a given threshold. This is expressed through the equations below:

$$\delta + z_{jt} \geq \zeta_{jt} \quad \forall j \in \mathcal{L}, t \in \mathcal{T} \tag{5g}$$

$$\zeta_{jt} \geq \delta z_{jt} \quad \forall j \in \mathcal{L}, t \in \mathcal{T} \tag{5h}$$

$$z_{jt} \in \{0, 1\} \quad \forall j \in \mathcal{L}, t \in \mathcal{T} \tag{5i}$$

Whereby, the structural constraints' uncertainty set is:

$$\mathcal{U}^{PH} = \left\{ \quad \boldsymbol{\xi} \quad \middle| \; (5a) - (5i) \right\} \tag{5j}$$

4.2.2. Data-driven distance-based uncertainty set

We propose a novel methodology to construct an uncertainty set directly based on data samples. This methodology is used to incorporate the information of substations' damage scenarios. The procedure for constructing the polyhedral uncertainty set is summarized in the next steps:

- *Step 1.* Gather J weighted data samples $\{\boldsymbol{\varphi}^j, \alpha^j\}_{j=1}^J$, where $\boldsymbol{\varphi}^j$ is j-th data point and α^j is the point's weight.
  The uncertainty data points are stacked into a data matrix $\boldsymbol{X} = [\boldsymbol{\varphi}^1, \ldots, \boldsymbol{\varphi}^J] \in \mathbb{R}^{m \times J}$. Each column represents a data point in m-dimensional space, totaling a number of J columns. In addition, we define the vector $\boldsymbol{\alpha} = [\alpha^1, \ldots, \alpha^J]^\top \in \mathbb{R}^J$, which contains the normalized weights of samples, i.e., $\boldsymbol{\alpha}^\top \mathbf{1} = 1$. Note that $\boldsymbol{\alpha}$ could be given by $\alpha^j = 1/J \quad \forall j$.

- *Step 2.* Obtain the mean vector $\boldsymbol{\mu}$, the centered-to-zero data matrix $\hat{\boldsymbol{X}}$, and the unbiased covariance matrix $\boldsymbol{\Sigma}$.

  The mean vector of the sample is

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\alpha} \tag{6a}$$

  We compute the centered-to-zero data matrix, termed $\hat{\boldsymbol{X}}$

$$\hat{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{\mu}\mathbf{1}^\top \tag{6b}$$

  Then, the unbiased weighted covariance matrix is obtained through the next equality (GSL, 2007):

$$\boldsymbol{\Sigma} = \frac{\sum_{j=1}^{J} \alpha^j}{\left(\sum_{j=1}^{J} \alpha^j\right)^2 - \sum_{j=1}^{J} \alpha^{j\,2}} \left(\sum_{j=1}^{J} \alpha_j \hat{\boldsymbol{X}}^{j^\top} \hat{\boldsymbol{X}}^j\right) \tag{6c}$$

  where $\hat{\boldsymbol{X}}^j$ is the j-th column of $\hat{\boldsymbol{X}}$.

- *Step 3.* Calculate the principal components of the covariance matrix using singular value decomposition.

  The singular value decomposition of the covariance matrix $\boldsymbol{\Sigma}$ gives us the eigenvalues and the eigenvectors associated with it. The decomposition $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ is constituted by the information contained in $\boldsymbol{\Lambda}$, which is the diagonal matrix $\text{diag}\{\lambda_1, \ldots, \lambda_m\}$ where $\lambda_i \quad \forall i \in \{1, \ldots, m\}$ are the eigenvalues, and $\boldsymbol{V}$ is the square matrix with the respective eigenvectors $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m] \in \mathbb{R}^{m \times m}$. For this work, we assume, with no loss of generality, that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$.

- *Step 4.* Apply whitening to each centered sample point.

  It is desirable to apply a linear rescaling to the centered-to-zero data matrix to eliminate cross-correlations and attain a unit variance over each dimension; hence, the resulting covariance matrix corresponds to the identity matrix. The identity covariance matrix is obtained through a whitening transformation (Bishop et al., 1995). The resulting rescaled data matrix is then said to be isotropic

(Brubaker & Vempala, 2008). Moreover, one can decide to represent the data in a q-dimensional subspace (where $q \leq m$), defined by a principal subspace, and the quality of the q-dimensional approximation is computed as the proportion of the total variance of the original data set that is accounted for, calculated as $\sum_{i=1}^{q} \lambda_i / \sum_{i=1}^{m} \lambda_i$ (Jolliffe & Cadima, 2016). Note that dimensionality reduction could be wanted to lighten the computational burden of treating the robust problem.

Let $\boldsymbol{\Lambda}_q^{-1/2}$ the diagonal matrix diag$\{\lambda_1^{-1/2}, \ldots, \lambda_q^{-1/2}\}$ be the $q \times q$ the diagonal matrix with the first (largest) q eigenvalues in a decreasing order, $\boldsymbol{V}_q$ the $m \times q$ matrix resulting from retaining the corresponding eigenvectors, and $\boldsymbol{W} = \boldsymbol{\Lambda}_q^{-1/2} \boldsymbol{V}_q^{\top}$. Then, the whitened zero-centered samples are:

$$\boldsymbol{Y} = \boldsymbol{W} \hat{\boldsymbol{X}} \tag{6d}$$

where $\boldsymbol{Y} \in \mathbb{R}^{q \times J}$ is a lower-dimensional and rescaled representation of the original data, which has an identity covariance matrix and a zero mean vector.

- *Step 5.* Compute the two shape parameters $\overline{d}$ and $\underline{d}$.

  We want to construct a polyhedron with adjustable size, depending on the level of conservatism of the decision-maker. For this regard, we propose a set of this form:

$$\left\{ \quad \boldsymbol{y} \in \mathbb{R}^q \quad \middle| \quad \sum_{j=1}^{J} \alpha^j \|\boldsymbol{Y}^j - \boldsymbol{y}\|_1 \leq \phi \right\} \tag{6e}$$

  where $\boldsymbol{Y}^j$ denotes the j-th column of the matrix $\boldsymbol{Y}$. To properly choose the value of ϕ, two shape parameters are used, namely $\overline{d}$ and $\underline{d}$ (for which $\overline{d} \geq \underline{d}$). Then, ϕ results from a linear combination between them, i.e, $\phi = \underline{d}(1-\Delta) + \Delta \overline{d}$, with $\Delta \geq 0$. These parameters are defined bellow:

$$\underline{d} = \sum_{j=1}^{J} \alpha^j \|\boldsymbol{Y}^j\|_1 \tag{6f}$$

$$\overline{d} = \max_{i=1,\dots,J} \sum_{j=1}^{J} \alpha^j \|\boldsymbol{Y}^j - \boldsymbol{Y}^i\|_1 \tag{6g}$$

so, if $\Delta = 0$ the size of the polyhedron is the minimum size that fulfills the condition that the centroid is included, and $\Delta = 1$ yields the smallest size definable by the set which includes all data samples.

- *Step 6 (Optional).* Aggregation of whitened samples through clustering.
  If the number J is too large we suggest reducing the total points that define the uncertainty set using some clustering algorithm. We recommend the usage of the `K-means++` algorithm (Arthur & Vassilvitskii, 2007), which has been experimentally shown that outperforms the most widely used partitional clustering algorithm termed `K-means` algorithm (Celebi et al., 2013), and it is implemented in scikit-learn package in Python language (Pedregosa et al., 2011). The resulting K centers are stacked in a matrix $\boldsymbol{A} = [a^1, \dots, a^K] \in \mathbb{R}^{q \times K}$, and a vector $\boldsymbol{\beta}$ of weights is associated to them accordingly, given $\boldsymbol{\beta} = [\beta^1, \dots, \beta^K]^\top \in \mathbb{R}^K$, where $\boldsymbol{\beta}^\top \mathbf{1} = 1$.
- *Step 7.* Define the data-driven distance-based polyhedron.
  The polyhedron based in K q-dimensional points is as follows,

$$\mathcal{U}^{DD} \left\{ \boldsymbol{\xi} \;\middle|\; \begin{array}{ll} \displaystyle\sum_{k\in\mathcal{K}} \sum_{q\in\mathcal{Q}} \beta^k d_{kq} \leq \underline{d}(1-\Delta) + \Delta\overline{d} & \\ d_{kq} \geq \omega_q - a_q^k, & k \in \mathcal{K}, q \in \mathcal{Q} \\ d_{kq} \geq a_q^k - \omega_q, & k \in \mathcal{K}, q \in \mathcal{Q} \\ \boldsymbol{\varphi} = W^{-1}\boldsymbol{\omega} + \boldsymbol{\mu} & \end{array} \right\} \tag{6h}$$

  where $\boldsymbol{W}^{-1} = \boldsymbol{V}_q \boldsymbol{\Lambda}_q^{1/2}$, $\mathcal{K} = \{1, \dots, K\}$, $\mathcal{Q} = \{1, \dots, q\}$, and $\boldsymbol{\omega} = [\omega_1, \dots, \omega_q]^\top$. Note that $\|\boldsymbol{\omega} - \boldsymbol{a}\|_1$ has been replaced by an equivalent linear formulation, introducing auxiliary variables d_{kq}.
- *Step 8.* Choose polyhedron's size with γ value.

The value of Δ is chosen according to the total weight included from the N whitened q-dimensional samples $\boldsymbol{Y}$ in the polyhedron. Let $\psi_j = \sum_{k\in\mathcal{K}} \beta^k \|a^k - y^j\|_1$, for $j \in \{1,\ldots,J\}$. Then we compute $\Delta_j = (\psi_j - \underline{d})/(\overline{d} - \underline{d}) \quad \forall j \in \{1,\ldots,J\}$. Without loss of generality, we assume that the original samples are ordered such that $\Delta_1 \leq \Delta_2 \leq \ldots \leq \Delta_J$. By construction, the weight included from samples $\boldsymbol{Y}$ in a polyhedron with size parameter Δ_j is $\sum_{i=1}^{j} \alpha_i$. Hence, for a given value of confidence γ, Δ is set to Δ_{j^*}, where $j^* = \text{argmin}_{j\in\{1,\ldots,J\}}\{\sum_{i=1}^{j} \alpha_i : \sum_{i=1}^{j} \alpha_i \geq \gamma\}$.

4.3. Data-driven stochastic-robust power network planning model

We propose a two-stage power network planning model that determines the optimum investment decisions for a given budget to improve the power system's resilience under an uncertain seismic hazard. Two types of investments are considered: 1) transmission lines' capacity expansion and 2) siting and sizing energy storage systems. Further, we incorporate information about the multiple earthquake sources to which the power system is exposed to make data-driven decisions. This is achieved by adopting a data-driven stochastic-robust optimization scheme (DDSRO) (Yue & You, 2016; Ning & You, 2018b). The DDSRO scheme leverages the probability distribution of seismic sources while capturing complex uncertainty data structures within the same source employing uncertainty sets. Hence, the first stage decides investment decisions to minimize the expected value of objectives of the earthquake attacker-defender problem over the seismic source variable. In this way, the proposed model is defined in compact form as follows:

$$\min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_S \Big[\max_{\boldsymbol{\xi}\in\mathcal{U}_S} Q(\boldsymbol{x}, \boldsymbol{\xi}) \Big], \tag{7}$$

where $\boldsymbol{x}$ represents investment decisions and $\mathcal{X}$ its associated feasible space. Also, S is a random variable defined over the set of seismic sources, i.e., $\mathbb{P}(S \in \mathcal{S}) = 1$, with $\mathcal{S} = \{1,\ldots,C\}$. The uncertainty set corresponding to the seismic source S is $\mathcal{U}_S$. In

addition, the investment decisions and constraints $\boldsymbol{x} = (\overline{\boldsymbol{E}}, \overline{\boldsymbol{P}}, \boldsymbol{x}^l)$ and $\mathcal{X}$ are defined as:

$$\mathcal{X} = \Big\{ \boldsymbol{x} \;\Big|\; \sum_{j \in \mathcal{L}^C} c_j^l x_j^l + \sum_{b \in \mathcal{B}} c_b^p \overline{P_b} = I \tag{8a}$$

$$\overline{P_b} = \frac{\overline{E_b}}{\rho} \quad \forall b \in \mathcal{B} \tag{8b}$$

$$x_j^l \geq 0 \quad \forall j \in \mathcal{L}^C \tag{8c}$$

$$\overline{E_b} \geq 0 \quad \forall b \in \mathcal{B} \Big\}. \tag{8d}$$

Here, (8a) is a resiliency budget constraint (Romero et al., 2013; Lagos et al., 2019), (8b) defines the technical relation between power and storage capacities for energy storage systems (Dvorkin et al., 2017), and (8c)-(8d) are non-negativity constraints.

4.3.1. Seismic sources uncertainty modeling

Through maximum likelihood estimation, one can derive the probability of different labels from labeled uncertainty data (Ning & You, 2018b). From a collection of J weighted labeled damage samples $\{\boldsymbol{\varphi}^j, \alpha^j, l^j\}_{\mathcal{J}}$, with $\mathcal{J} = \{1, \ldots, J\}$, we obtain the probability of each seismic sources as

$$p_s = \sum_{j \in \mathcal{J}} \alpha^j \mathbb{I}(l^j = s), \tag{9}$$

where p_s is the probability assigned to seismic source s, l^j is the label of the j-th data point, and $\mathbb{I}(\cdot)$ is an indicator function, i.e., $\mathbb{I}(l^j = s)$ is 1 if $l^j = s$ holds and 0 if not.

From the J weighted and labeled data samples, we separate labeled data $\{\boldsymbol{\varphi}_s^i, \hat{\alpha}_s^i\}_{i \in \mathcal{J}_s}$ $\forall s \in \mathcal{S}$, with $\mathcal{J}_s = \{j \in \mathcal{J} : l^j = s\}$, where $\hat{\alpha}_s^i = \alpha^i / \sum_{j \in \mathcal{J}_s} \alpha^j \quad \forall i \in \mathcal{J}_s$. Note that $\sum_{i \in \mathcal{J}_s} \hat{\alpha}_s^i = 1 \quad \forall s \in \mathcal{S}$. Then, for each seismic source s, we define a data-driven

distance-based uncertainty set $\mathcal{U}_s^{DD}$ over the collection $\{\boldsymbol{\varphi}_s^i, \hat{\alpha}_s^i\}_{i \in \mathcal{J}_s}$, following the procedure presented in (4.2.2). Therefore, the uncertainty set used for each seismic source $s \in \mathcal{S}$ is defined as $\mathcal{U}_s = \mathcal{U}_s^{DD} \cap \mathcal{U}^{PH}$, with $\mathcal{U}^{PH}$ previously shown in (4.2.1).

5. SOLUTION METHOD

The structure of the DDSRO problem (7) yields a convenient reformulation making use of a few auxiliary variables:

$$\min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\vartheta}} \quad \sum_{s\in\mathcal{S}} p_s \vartheta_s \tag{10a}$$

$$\text{s.t.} \quad \vartheta_s \geq Q(\boldsymbol{x},\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \mathcal{U}_s, s \in \mathcal{S}. \tag{10b}$$

The variables ϑ_s represent the worst-case second-stage cost for each seismic source. Constraint (10b) ensures that each objective value of the seismic source s is that of the worst realization of the uncertain vector $\boldsymbol{\xi}$ within the uncertainty set $\mathcal{U}_s$.

The formulation presented in (10a)-(10b) considers an infinite number of constraints, because there are infinite points $\boldsymbol{\xi}$ contained in each uncertainty set $\mathcal{U}_s \quad \forall s \in \mathcal{S}$, which are all bounded mixed integer polyhedral sets. To solve this problem, we adopt the column-and-constraint generation algorithm (CCG), an effective solution for dealing with problems of this structure (L. Zhao & Zeng, 2012; Zeng & Zhao, 2013).

5.1. Column and constraint generation method

Bounded mixed-integer sets defined by linear constraints have a finite number of extreme points (L. Zhao & Zeng, 2012). Then, let $\text{Ext}\big[\mathcal{U}_s\big] \subset \mathcal{U}_s \quad \forall s \in \mathcal{S}$ be the sets containing every extreme point of their respective uncertainty set. The problem presented in (10a)-(10b) is equivalent to solving the following problem:

$$\min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\vartheta}} \quad \sum_{s\in\mathcal{S}} p_s \vartheta_s \tag{11a}$$

$$\text{s.t.} \quad \vartheta_s \geq Q(\boldsymbol{x},\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \text{Ext}\big[\mathcal{U}_s\big], s \in \mathcal{S} \tag{11b}$$

We define $\mathcal{R}_s \subset \text{Ext}\big[\mathcal{U}_s\big] \quad \forall s \in \mathcal{S}$, a subset containing a partial enumeration of the extreme points. Then, a valid relaxation of the problem (11a)-(11b) is:

$$\min_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\vartheta}} \quad \sum_{s\in\mathcal{S}} p_s \vartheta_s \tag{12a}$$

$$\text{s.t.} \quad \vartheta_s \geq Q(\boldsymbol{x}, \boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \mathcal{R}_s, \forall s \in \mathcal{S} \tag{12b}$$

The CCG algorithm consists of the iterative addition of non-trivial scenarios to the subsets $\mathcal{R}_s \quad \forall s \in \mathcal{S}$, until the optimal solution of the relaxed problem remains invariant. To add scenarios to each $\mathcal{R}_s$, the earthquake attacker-defender problem (4) must be solved for each seismic source s under a given first stage solution $\boldsymbol{x}$:

$$F_s(\boldsymbol{x}) = \max_{\boldsymbol{\xi}\in\mathcal{U}_s} \min_{\boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x},\boldsymbol{\xi})} \boldsymbol{e}^\top \boldsymbol{y} \quad \forall s \in \mathcal{S} \tag{13}$$

The CCG algorithm for the DDSRO problem, presented in Algorithm 1, converges in a finite number of iterations for bounded mixed integer sets with linear constraints given by $\mathcal{U}_s \quad \forall s \in \mathcal{S}$ and polyhedral $\mathcal{Y}(\boldsymbol{x}, \boldsymbol{\xi})$, as it is in this case (L. Zhao & Zeng, 2012).

In the next section, we discuss a solution method for the earthquake attacker-defender problem (13).

Algorithm 1 Column and Constraint Generation Method

Set $r \leftarrow 0$, $\mathcal{R}_s \leftarrow \emptyset \quad \forall s \in \mathcal{S}$
repeat
  $(\boldsymbol{x}, \boldsymbol{\vartheta}) \leftarrow$ optimal solution of the master problem (12a)-(12b)
  for s in $\mathcal{S}$ **do**
    Evaluate $F_s(\boldsymbol{x}) : \boldsymbol{\xi}_{s,r+1} \leftarrow$ optimal solution of (13)
    $\mathcal{R}_s \leftarrow \mathcal{R}_s \cup \{\boldsymbol{\xi}_{s,r+1}\}$
  end for
  $r \leftarrow r + 1$
until $\sum_{s\in\mathcal{S}} p_s F_s(\boldsymbol{x}) \leq \sum_{s\in\mathcal{S}} p_s \vartheta_s$
return $(\boldsymbol{x}, \boldsymbol{\vartheta})$

5.2. Alternating direction method

The inner level problem of the earthquake attacker-defender model (13) is a linear minimization problem which is feasible for any possible attack, in fact, problem (3a)-(3u) always admits the solution given by $\boldsymbol{P}^{LS} = \boldsymbol{D}$ and all other operational variables set to $\mathbf{0}$. Thus, through strong duality, an equivalent linear maximization problem can be obtained. A compact formulation of the problem for a given s seismic source and a fixed investment decision $\boldsymbol{x}$ is as follows:

$$\max_{\boldsymbol{\xi},\boldsymbol{\pi}} \quad \boldsymbol{\xi}^\top E(\boldsymbol{x})\boldsymbol{\pi} + f(\boldsymbol{x})^\top \boldsymbol{\pi} \tag{14a}$$

$$\text{s.t.} \quad \boldsymbol{\xi} \in \mathcal{U}_S \tag{14b}$$

$$\boldsymbol{\pi} \in \mathcal{P}, \tag{14c}$$

where $E(\boldsymbol{x})$ and $f(\boldsymbol{x})$ are a matrix and a vector of the corresponding dimensions and with entries defined by affine functions over the investment decision $\boldsymbol{x}$. The $\boldsymbol{\pi}$ variables and the polyhedral set $\mathcal{P}$ represent the dual variables and dual constraints of the inner minimization problem, respectively. Note that the objective function (14a) is bilinear in variables $\boldsymbol{\pi}$ and $\boldsymbol{\xi}$, and the constraints of these two types of variables are disjoint. See Appendix A for a detailed formulation of this problem.

To solve the problem (14a)-(14c), we propose the adoption of an iterative method for disjoint bilinear programs named the alternating direction (AD) method which approximately solves the problem (Konno, 1976; Sun & Lorca, 2015; Lorca & Sun, 2017). This algorithm optimizes over $\boldsymbol{\pi}$ with $\boldsymbol{\xi}$ fixed, to then fix the computed $\boldsymbol{\pi}$ to optimize over $\boldsymbol{\xi}$, and alternates until convergence. The formal definition of the method is presented in Algorithm 2. Note that in this case, optimizing over $\boldsymbol{\xi}$ accounts for solving a mixed-integer linear optimization problem and optimizing over $\boldsymbol{\pi}$ accounts for solving a linear optimization problem.

Algorithm 2 Alternating Direction Method

1: Choose an initial $\boldsymbol{\xi}$
2: **repeat**
3: $\quad \boldsymbol{\pi} \leftarrow \arg\max_{\boldsymbol{\pi} \in \mathcal{P}} \{\boldsymbol{\xi}^\top E(\boldsymbol{x})\boldsymbol{\pi} + f(\boldsymbol{x})^\top \boldsymbol{\pi}\}$ with objective ν
4: $\quad \boldsymbol{\xi} \leftarrow \arg\max_{\boldsymbol{\xi} \in \mathcal{U}} \{\boldsymbol{\xi}^\top E(\boldsymbol{x})\boldsymbol{\pi} + f(\boldsymbol{x})^\top \boldsymbol{\pi}\}$ with objective ν'
5: **until** $\nu = \nu'$ or tolerance criterion is met.
6: **return** $(\nu, \boldsymbol{\xi})$

6. CASE STUDY: CHILEAN ELECTRIC POWER SYSTEM

This Section presents exhaustive computational experiments to clarify how the earthquake attacker-defender model can provide meaningful insights about the seismic risk under which the power network is exposed and the support that the data-driven robust optimization planning model can contribute to making investments that enhance the seismic resilience of the power network.

We set all computational experiments on the Chilean power network representation employed by the system operator, consisting of 281 buses, 365 transmission lines, 362 conventional power units, 192 solar power units and 49 wind power units. Moreover, we employed the renewable generation and load profiles from Verástegui et al. (2019). In addition, the loss function in (3a) was defined as $L(\boldsymbol{y_{dht}}) = \sum_{b \in \boldsymbol{B}} P^{LS}_{bdht}$, so the overall objective of the operational model is to minimize the total load shedding over the operation horizon, as previously done in (Lagos et al., 2019). The investment costs in batteries and transmission line expansions are based on the information provided by the Chilean Ministry of Energy (MinEnergia, 2020), as well as the set of candidate transmission lines, which are the ones that operate in the voltage range between 220 kV and 500 kV.

Concerning seismic risk, a collection of 20000 seismic-induced damage scenarios were considered in this study, simulating earthquakes of 5 Mw (moment magnitude) or higher. The peak ground acceleration vectors, their seismic source, and their weight were obtained using the Chilean seismic recurrence model developed in Poulos et al. (2019) and the ground motion prediction model of Parker et al. (2020). Further, the number of seismic sources [1] considered is seven, as shown in Figure 3.1. These sources are constituted of three subduction interface zones (zones 1–3), and four subduction intraslab zones (zones 4–7) Poulos et al. (2017). Then, the available capacity of all substations at different recovery stages was determined using the technical specifications for anchored substations

[1] The definition of seismic source is presented in Section 3.1.

(or those with high seismic performance) provided in FEMA (2013). Furthermore, in the optimization models, we considered a one-month operation horizon, divided into recovery stages based on the discretized restoration process description for substations shown in FEMA (2013). The details of these stages are covered in each experiment.

Uncertainty sets were built using dimensionality reduction, with the minimum dimensions which explained 90% or more of the data variance. Clustering was performed onto the whitened data samples, and we defined the number of clusters as one order of magnitude less than the original number of data points for each set (dividing by 10). For the damage threshold that implies disconnection of transmission lines (see Section 4.2.1), we fixed $\delta = 70.1\%$, which is slightly higher than the capacity derate of substations with extensive damage (70%), as shown in FEMA (2013). Thus, a line is connected if its capacity derate is less or equal than the one caused by extensive damage in one of its connecting substations and the other with no damage. It is disconnected if its derate is greater than the threshold. Furthermore, we chose three risk values for the test cases, with $\gamma \in \{95.0\%, 97,5\%, 99.0\%\}$, as explained in Step 8 of Section 4.2.2.

The models and algorithms developed in this work were programmed in the Julia v1.5 language, using the embedded package JuMP for mathematical optimization with Gurobi solver v9.0. We set an optimality gap of 1% for all cases and a convergence tolerance of 3% for the alternating direction algorithm, with an initial solution given by $\boldsymbol{\xi} = \mathbf{1}$ [2]. We executed all experiments in a Dell PowerEdge R360 server with an Intel Xeon CPU E5-2630 v4 processor running at 2.20GHz, with 64 GB of RAM.

In what follows, Section 6.1 analyses the impact of earthquakes from different seismic sources using the developed attacker-defender model. Then, Section 6.2 proposes various investment plans based on the DDSRO model's solutions, depending on the disposable

[2] In this thesis, we chose the optimality gap and the convergence tolerance for the sake of computational tractability.

budget and the degree of conservativeness of the decision-maker, for enhancing power network seismic resilience.

6.1. Identification and analysis of critical earthquakes

In this experimental setup, we considered three recovery stages for the one-month operation horizon. The first recovery stage lasts for 3 days and considers the available capacities of components immediately after an earthquake. The other two stages consider 3 days and 7 days of recovery, which last fir 4 days and 23 days, respectively. In addition, we used 4 representative days of 24 hours to model the behavior of each stage.

The objective values of the earthquake attacker-defender model solutions for each seismic source under different values of γ are presented in Figure 6.1, considering no prior investments in the network. As expected, the objective values obtained are non-decreasing in γ due to the enlargement that this parameter causes in the uncertainty sets. Further, we note that the objective values vary strongly depending on the seismic source. For instance, Zone 4 and Zone 6 produce almost no load shedding after an earthquake generated by these sources. Moreover, the source that produces the largest load shedding is not always the same. For $\gamma = 95.0\%$, Zone 5 causes the worst load shedding value, while for $\gamma = 97.5\%$ and $\gamma = 99.0\%$ the most damaging seismic source is Zone 7. Furthermore, the last three bars in the figure show the expected value of the worst-case costs from the earthquake attacker-defender model under the random variable that determines the seismic source. Hence, it helps to understand the overall risk to which the network is subjected.

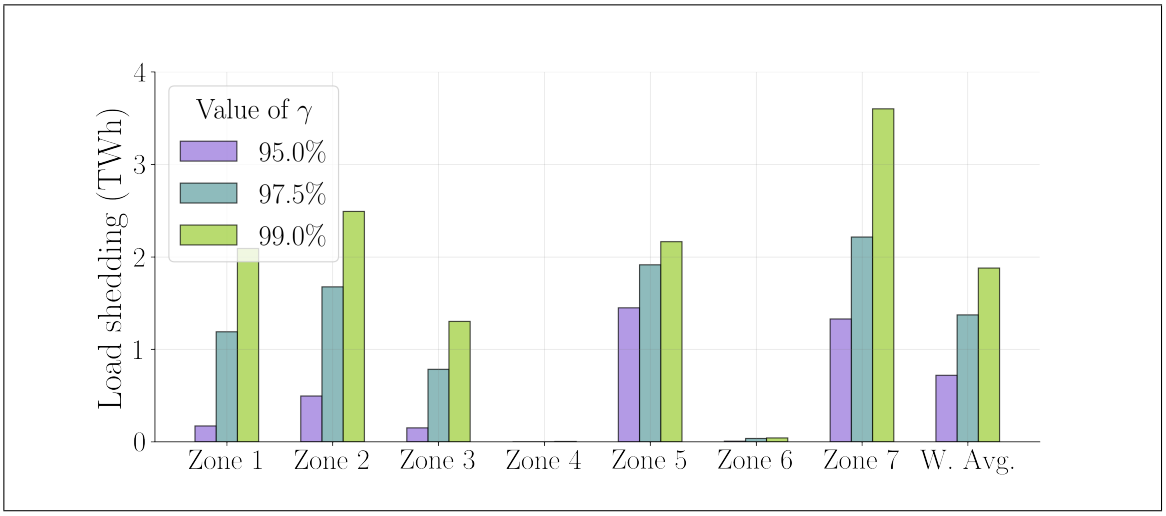

Figure 6.1. Objective values for seismic sources and their weighted average

To explore what the attacker-level solutions look like, we show in Figure 6.2 examples of the damage of substations and transmission lines three days after the occurrence of an earthquake in seismic source Zone 7 [3], for two values of γ. This figure provides valuable insights into the vulnerabilities of the given network configuration. The attacker level builds optimized seismic attacks based on the earthquake data employed by the model, accounting for the temporal and spatial correlation of the substations' capacity derates, so we can describe the worst possible earthquakes generated in each seismic zone for different levels of probability mass included (defined by γ) through this data-driven optimization scheme. Therefore, the earthquake attacker-defender model delivers helpful comprehension for decision-making of seismic-resilient planning by leveraging the data at hand.

[3]Another example can be found in Appendix E.

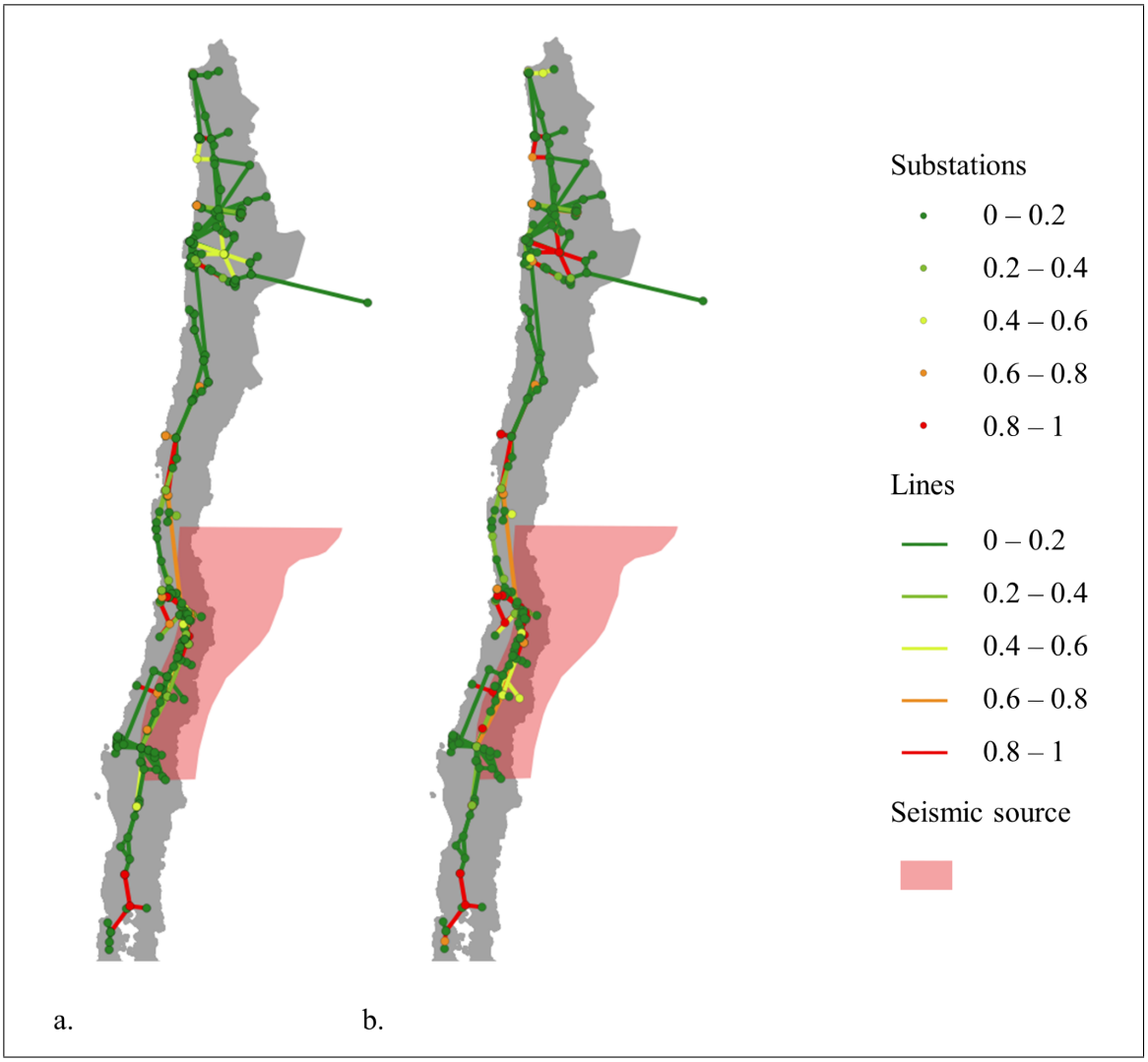

Figure 6.2. Maps of damage three days after the occurrence of an earthquake originated from seismic source Zone 7: a) $\gamma = 95\%$, and b) $\gamma = 99\%$

Regarding the computational tractability of the earthquake attacker-defender model, Table 6.1 presents the minimum, maximum and average running time over the seven seismic sources until convergence of Algorithm 2, under the different γ values. We can see

that the running times do not vary significantly, and the problem is solved within minutes, which is very efficient for its practical application by decision makers.

Table 6.1. Running time in minutes.

γ	95.0%	97.5%	99.0%
Min	3.3	4.1	3.3
Max	7.9	11.7	6.9
Avg	5.3	7.4	5.0

6.2. Planning against critical earthquakes

For the experiments of this section, we used two recovery stages during the month of operation. The first recovery stage lasts for 7 days and considers the damage states immediately after the earthquake, and the second recovery stage considers 7 days of recovery and ends 23 days after. We modeled the load and renewable generation profiles through a single representative day composed of 6 ordered blocks of 4 hours each.

This part analyzes the DDSRO model's solutions and computational performance for three values of γ and two different resilience budgets of 50 and 100 million USD.

Table 6.2. Investment results for the DDSRO model.

γ	95.0%		97.5%		99.0%	
I	50	100	50	100	50	100
Load shedding (GWh)	494.4	494.2	707.1	698.9	887.4	873.5
L.S. reduction (%)	2.3	2.4	9.4	10.5	8.4	9.8
Expanded lines	15	19	14	19	16	23
Batteries installed	0	1	0	0	0	0
Invested in lines (%)	100	88.9	100	100	100	100
ENC	6.0	7.4	6.0	9.4	9.8	12.9

In Table 6.2 we present a summary of various relevant features of the investment solutions. First, we confirm that the expected value of load shedding increases with γ and decreases with a larger resilience budget. Then, we note that the reduction in the expected load shedding for different investment configurations compared to the power network with no investment is significant, especially for $\gamma = 97.5\%$ and $\gamma = 99.0\%$. Moreover, The first 50 million USD proves to be proportionally more beneficial than investing 100 million USD. Therefore, the optimized seismic-resilient investment proves to be of benefit regarding the improvement of energy supply. In addition, the model decides to invest almost the totality of the resilience budget in transmission lines, as reported in the number of expanded lines, batteries installed, and the percentage of the budget invested in lines. Thus, under this analysis, current battery energy storage systems are not cost-effective concerning seismic resilience.

Moreover, we used a diversification index to analyze the diversification degree of the investments: the index termed *effective number of constituents* or simply ENC (Carli et al., 2014). In our context, the ENC index is defined in eq. (15). The ENC reaches a minimum of 1 if the investment decision considers investments in only one network component, and a maximum equal to $|\mathcal{L}^C| + |\mathcal{B}|$ when the investment decision assigns an equal percentage of the budget to every component in which the the model can invest.

$$ENC = \frac{I^2}{\sum_{j \in \mathcal{L}^C} \left(c_j^l x_j^l\right)^2 + \sum_{b \in \mathcal{B}} \left(c_b^p \overline{P_b}\right)^2} \tag{15}$$

The ENC reported in Table 6.2 show that the diversification degree of the investment is consistently non-decreasing in the budget and γ. Hence, more diversified investments, i.e., investments distributed equitably and in many assets, are desirable when the power network faces higher levels of seismic-induced component damage and when the decision-maker has a larger resilience budget.

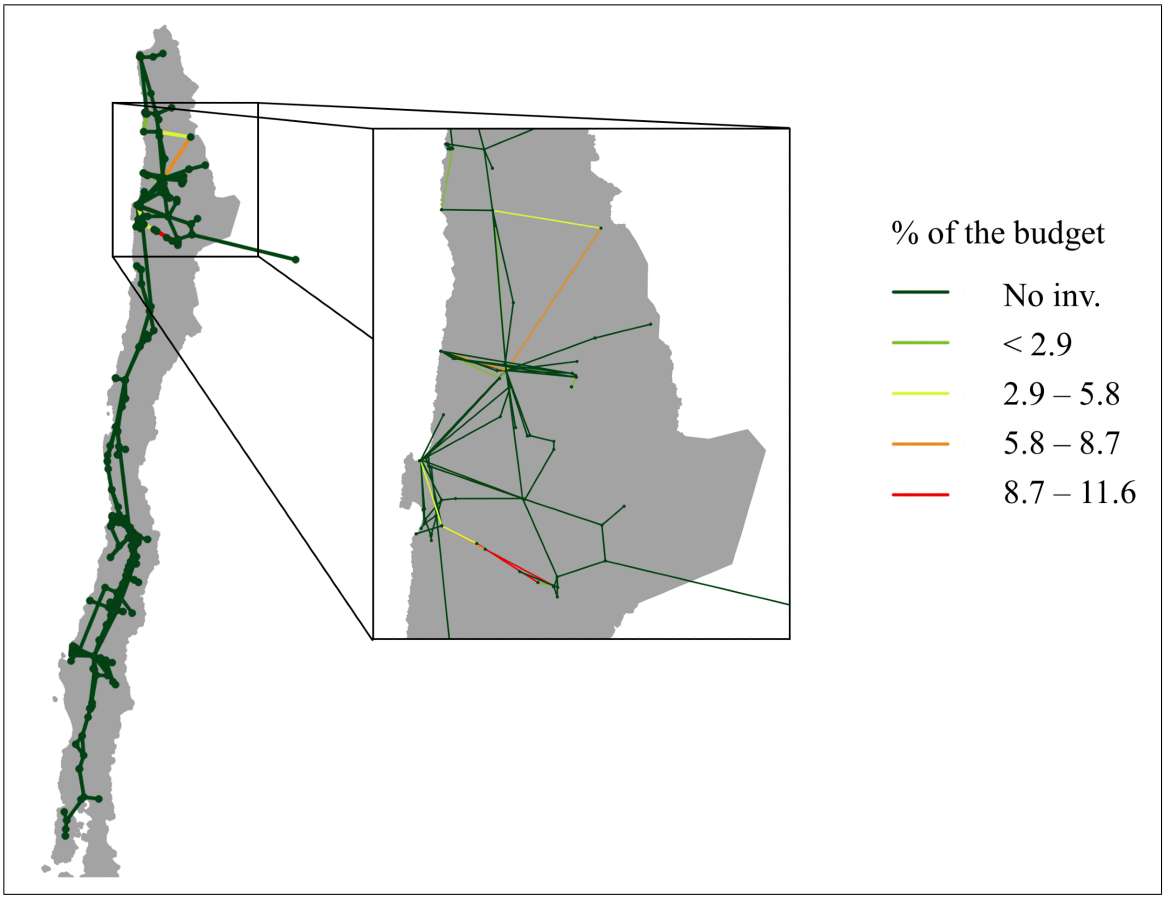

Figure 6.3. Investment solution for a 100 MMUSD budget and $\gamma = 99.0\%$.

To further explore the DDSRO solutions, we summarized the first five most significant investments made under each investment scheme in Table 6.3, in descending order of magnitude for the two resilience budgets given a fixed value of γ. For the complete names and locations of these components, see Table B.1 in the Appendix. First, we note that the optimizer invests in L_5 when the resilience budget is 50 MMUSD, despite the value of γ, while there is no common investment among the five most extensive investments for the 100 MMUSD case. Further, we see that L_2, L_3, and L_5 appear three or more times between the five largest investments, so we can expect that these transmission lines significantly impact the resilience of the Chilean power system. An interesting fact to notice

is that, as presented in Table B.1, 13 of the 17 components presented here are located in the far north macro-zone of Chile, under the definition of macro-zones in (MinEnergia, 2020), so based on our analysis the best investment opportunities for enhancing the seismic resilience are found there. Furthermore, a closer look at an investment solution, as shown in Figure 6.3, allows to observe that the model invests most of the budget in a specific zone [4]. These results present evidence that there are no one-fits-all solutions for enhancing the power system's seismic resilience and ratifying the value of using a flexible optimization framework since the optimal investments that minimize the load shedding change according to the level of risk to which the system is exposed.

Table 6.3. First five largest investments

γ	I	Investments
95.0%	50	L_1, L_2, L_3, L_4, L_5
	100	L_6, L_7, L_8, B_1, L_3
97.5%	50	L_9, L_2, L_5, L_{10}, L_{11}
	100	L_9, L_{10}, L_2, L_3, L_4
99.0%	50	L_{12}, L_5, L_{13}, L_{14}, L_{15}
	100	L_6, L_{13}, L_5, L_{16}, L_{12}

Regarding computational tractability, the running time and the number of master iterations until convergence of Algorithm 1 for the DDSRO model are shown in Table 6.4. We can observe that the running time increases with the allowed budget and the value of γ. It is important to remark here that with each master iteration, the algorithm adds one seismic-induced damage scenario for each seismic source considered, which in our study case means a total of seven scenarios. These addition of scenarios are computationally expensive and explains why running time varies widely depending on the number of master

[4]This solution shows that model prefers investments in the most economical macro-zone to make transmission lines expansions, as noted in the Table D.1

iterations, as seen in the cases when four such iterations are needed. However, the running time presented, which varies between 15 an 92 minutes for a large real-world power system, shows an exceptional efficiency of the proposed approach, given that this model is to be employed in planning studies.

Table 6.4. Computational performance of the DDSRO model.

γ	95.0%		97.5%		99.0%	
I	50	100	50	100	50	100
Running time (min)	15.4	19.3	19.0	21.2	56.2	91.8
Master iterations	3	3	3	3	4	4

7. CONCLUSIONS

We have developed a data-driven optimization framework for seismic-resilient power network planning. First, the proposed earthquake attacker-defender model harnesses a novel uncertainty set for high-dimensional uncertain parameters to unveil the worst-case network contingencies under different risk levels, using a sequential representation of the evolution of the system response and restoration after the hazard interdiction. Further, the DDSRO investment planning model includes multiple earthquake attacker-defender problems in its inner levels while accounting for distributional information for seismic sources and choosing optimal transmission lines capacity expansion and siting and sizing energy storage systems to enhance the power network seismic resilience. These problems are efficiently solved using the proposed solution method. Computational experiments on a 281-node representation of the Chilean power system show the effectiveness of the proposed framework.

Computational experiments on a 281-node representation of the Chilean power system showed the effectiveness of the proposed framework. On the one hand, The earthquake attacker-defender model identified worst-case contingencies for each seismic source considered and different risk levels. Further, we showed that mapping these solutions with geographical information systems helps visualize critical parts of the power network easily. On the other hand, we solved the DDSRO model for several combinations of budgets and conservativeness levels, which provided meaningful insights: 1) unfulfilled power demand can be significantly reduced with seismic-resilient planning, 2) investments in transmission lines are preferred over energy storage systems under the case study, 3) different investment configurations share no obvious preferable assets, and 4) optimum investment decisions tend to be more diversified while the risk level increases.

There are valuable future research opportunities in at least four areas. The first corresponds to refining the data on structural fragility, local site conditions, and recovery times

for each vulnerable component of the network. A second opportunity is seen in incorporating the component repair and recovery problem into the system operator's decision process. The third line of research should respond to the impacts of investments in new technologies, such as storage and power generation with hydrogen or the construction of concentrated solar power plants. Finally, it is desirable to improve the solution methodology to make it more efficient, especially in solving the data-driven stochastic-robust problem through the column and constraint generation algorithm, due to the significant sensibility of the execution time to the number of master iterations.

REFERENCES

Araneda, J., Rudnick, H., Mocarquer, S., & Miquel, P. (2010). Lessons from the 2010 chilean earthquake and its impact on electricity supply. In *2010 international conference on power system technology* (p. 1-7).

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual acm-siam symposium on discrete algorithms* (p. 1027–1035). USA: Society for Industrial and Applied Mathematics.

Babaei, S., Jiang, R., & Zhao, C. (2020). Distributionally robust distribution network configuration under random contingency. *IEEE Transactions on Power Systems*, *35*(5), 3332–3341.

Bagheri, A., Zhao, C., Qiu, F., & Wang, J. (2019). Resilient transmission hardening planning in a high renewable penetration era. *IEEE Transactions on Power Systems*, *34*(2), 873-882.

Baker, J. W. (2015). Introduction to probabilistic seismic hazard analysis. *White paper, version 2.1*, 77.

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research*, *52*(1), 35–53.

Bishop, C. M., et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

Brubaker, S. C., & Vempala, S. S. (2008). *Isotropic pca and affine-invariant clustering.*

Carli, T., Deguest, R., & Martellini, L. (2014). Improved risk reporting with factor-based diversification measures. *EDHEC-Risk Institute Publications*.

Castillo, A. (2014). Risk analysis and management in power outage and restoration: A literature survey. *Electric Power Systems Research*, *107*, 9-15.

Celebi, M. E., Kingravi, H. A., & Vela, P. A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*, *40*(1), 200-210.

Dilley, M. (2005). *Natural disaster hotspots: a global risk analysis* (Vol. 5). World Bank Publications.

Duan, C., Fang, W., Jiang, L., Yao, L., & Liu, J. (2018). Distributionally robust chance-constrained approximate ac-opf with wasserstein metric. *IEEE Transactions on Power Systems*, *33*(5), 4924–4936.

Dvorkin, Y., Fernandez-Blanco, R., Wang, Y., Xu, B., Kirschen, D. S., Pandžić, H., ... Silva-Monroy, C. A. (2017). Co-planning of investments in transmission and merchant energy storage. *IEEE Transactions on Power Systems*, *33*(1), 245–256.

Eidinger, J. (2009, 06). Wenchuan earthquake impact to power systems.

Espinoza, S., Poulos, A., Rudnick, H., de la Llera, J. C., Panteli, M., & Mancarella, P. (2020). Risk and resilience assessment with component criticality ranking of electric power systems subject to earthquakes. *IEEE Systems Journal*, *14*(2), 2837-2848.

FEMA. (2013). *Hazus-mh 2.1 earthquake model technical manual*.

Frank, S., & Rebennack, S. (2016). An introduction to optimal power flow: Theory, formulation, and examples. *IIE Transactions*, *48*(12), 1172–1197.

Gacitua, L., Gallegos, P., Henriquez-Auba, R., Lorca, A., Negrete-Pincetic, M., Olivares, D., ... Wenzel, G. (2018). A comprehensive review on expansion planning: Models and tools for energy policy analysis. *Renewable and Sustainable Energy Reviews*, *98*, 346–360.

Gan, W., Ai, X., Fang, J., Yan, M., Yao, W., Zuo, W., & Wen, J. (2019). Security constrained co-planning of transmission expansion and energy storage. *Applied Energy*, *239*, 383-394.

Goda, K., & Atkinson, G. (2010, 12). Intraevent spatial correlation of ground-motion parameters using sk-net data. *Bulletin of the Seismological Society of America*, *100*, 3055-3067.

GSL. (2007). *GNU General Public License - Statistics.* Retrieved from `https://www.gnu.org/software/gsl/doc/html/statistics.html#covariance` ([Online; accessed 25-May-2021])

Jayaram, N., & Baker, J. W. (2010). Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. *Earthquake Engineering & Structural Dynamics*, *39*(10), 1109–1131.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Konno, H. (1976). A cutting plane algorithm for solving bilinear programs. *Mathematical Programming*, *11*(1), 14–27.

Koç, Y., Warnier, M., Kooij, R., & Brazier, F. (2014). Structural vulnerability assessment of electric power grids. In *Proceedings of the 11th ieee international conference on networking, sensing and control* (p. 386-391).

Lagos, T., Moreno, R., Espinosa, A. N., Panteli, M., Sacaan, R., Ordonez, F., ... Mancarella, P. (2019). Identifying optimal portfolios of resilient network investments against natural hazards, with applications to earthquakes. *IEEE Transactions on Power Systems*, *35*(2), 1411–1421.

Leyton, F., Ruiz, S., & Sepúlveda, S. (2009). Preliminary re-evaluation of probabilistic seismic hazard assessment in Chile: from Arica to Taitao Peninsula. *Advances in Geosciences*, *22*, 147–153.

Lorca, A., & Sun, X. A. (2016). Multistage robust unit commitment with dynamic uncertainty sets and energy storage. *IEEE Transactions on Power Systems*, *32*(3), 1678–1688.

Lorca, A., & Sun, X. A. (2017). The adaptive robust multi-period alternating current optimal power flow problem. *IEEE Transactions on Power Systems*, *33*(2), 1993–2003.

Maluenda, B., Negrete-Pincetic, M., Olivares, D. E., & Lorca, Á. (2018). Expansion planning under uncertainty for hydrothermal systems with variable resources. *International Journal of Electrical Power & Energy Systems*, *103*, 644–651.

Martin., A. (1990). Hacia una nueva regionalización y cálculo del peligro sísmico en Chile. *University of Chile*. (Master's Thesis)

Mccormick, G. P. (1976, December). Computability of global solutions to factorable nonconvex programs: Part i – convex underestimating problems. *Math. Program.*, *10*(1), 147–175.

MinEnergia. (2020). *Long-Term Energy Planning.* Retrieved from `https://energia.gob.cl/planificacion-energetica-de-largo-plazo-proceso` ([Online; accessed 02-June-2021])

Nan, C., & Sansavini, G. (2017). A quantitative method for assessing resilience of interdependent infrastructures. *Reliability Engineering & System Safety*, *157*, 35–53.

Nazemi, M., Moeini-Aghtaie, M., Fotuhi-Firuzabad, M., & Dehghanian, P. (2019). Energy storage planning for enhanced resilience of power distribution networks against earthquakes. *IEEE Transactions on Sustainable Energy*, *11*(2), 795–806.

Ning, C., & You, F. (2018a). Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. *Computers Chemical Engineering*, *112*, 190-210.

Ning, C., & You, F. (2018b). Data-driven stochastic robust optimization: General computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era. *Computers Chemical Engineering*, *111*, 115-133.

Ning, C., & You, F. (2019). Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*, *125*, 434–448.

Owen, A., & Zhou, Y. (2000). Safe and effective importance sampling. *Journal of the American Statistical Association*, *95*(449), 135–143.

Parker, G., Stewart, J., Boore, D., Atkinson, G., & Hassani, B. (2020). Nga-subduction global ground-motion models with regional adjustment factors. *Report No. August*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Poulos, A., Espinoza, S., de la Llera, J., & Rudnick, H. (2017). Seismic risk assessment of spatially distributed electric power systems. In *16th world conf. on earthquake eng., santiago* (pp. 1949–3029).

Poulos, A., Monsalve, M., Zamora, N., & de la Llera, J. C. (2019). An updated recurrence model for Chilean subduction seismicity and statistical validation of its poisson nature. *Bulletin of the Seismological Society of America*, *109*(1), 66–74.

Romero, N. R., Nozick, L. K., Dobson, I. D., Xu, N., & Jones, D. A. (2013). Transmission and generation expansion to mitigate seismic risk. *IEEE Transactions on Power Systems*, *28*(4), 3692-3701.

Rudnick, H., Mocarquer, S., Andrade, E., Vuchetich, E., & Miquel, P. (2011). Disaster management. *IEEE Power and Energy Magazine*, *9*(2), 37–45.

Salmeron, J., Wood, K., & Baldick, R. (2004). Analysis of electric grid security under terrorist threat. *IEEE Transactions on Power Systems*, *19*(2), 905-912.

Shang, C., & You, F. (2018). Robust optimization in high-dimensional data space with support vector clustering. *IFAC-PapersOnLine*, *51*(18), 19-24. (10th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2018)

Shumuta, Y. (2011). *2011 tohoku chiho-taiheiyo-oki earthquake—damage of electric power facilities in tohoku electric power co., inc.* Central Research Institute of Electric Power Industry, Civil Engineering.

Sun, X., & Lorca, (2015, 02). Adaptive robust optimization for daily power system operation. *Proceedings - 2014 Power Systems Computation Conference, PSCC 2014*.

Tapia, T., Álvaro Lorca, Olivares, D., Negrete-Pincetic, M., & Lamadrid, A. (2021). A robust decision-support method based on optimization and simulation for wildfire resilience in highly renewable power systems. *European Journal of Operational Research*, *294*.

Velloso, A., Pozo, D., & Street, A. (2020). Distributionally robust transmission expansion planning: a multi-scale uncertainty approach. *IEEE Transactions on Power Systems*, *35*(5), 3353–3365.

Verástegui, F., Lorca, Á., Olivares, D. E., Negrete-Pincetic, M., & Gazmuri, P. (2019). An adaptive robust optimization model for power systems planning with operational uncertainty. *IEEE Transactions on Power Systems*, *34*(6), 4606–4616.

Wang, Q., Watson, J.-P., & Guan, Y. (2013). Two-stage robust optimization for nk contingency-constrained unit commitment. *IEEE Transactions on Power Systems*, *28*(3), 2366–2375.

Wang, Y., Chen, C., Wang, J., & Baldick, R. (2016). Research on resilience of power systems under natural disasters—a review. *IEEE Transactions on Power Systems*, *31*(2), 1604-1613.

Yan, J., Hu, B., Xie, K., Tang, J., & Tai, H.-M. (2020). Data-driven transmission defense planning against extreme weather events. *IEEE Transactions on Smart Grid*, *11*(3), 2257-2270.

Yuan, W., Wang, J., Qiu, F., Chen, C., Kang, C., & Zeng, B. (2016). Robust optimization-based resilient distribution network planning against natural disasters. *IEEE Transactions on Smart Grid*, *7*(6), 2817–2826.

Yuan, W., Zhao, L., & Zeng, B. (2014). Optimal power grid protection through a defender–attacker–defender model. *Reliability Engineering & System Safety*, *121*, 83–89.

Yue, D., & You, F. (2016). Optimal supply chain design and operations under multi-scale uncertainties: Nested stochastic robust optimization modeling framework and solution algorithm. *AIChE Journal*, *62*(9), 3041-3055.

Zeng, B., & Zhao, L. (2013). Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, *41*(5), 457–461.

Zhang, Y., Shen, S., & Mathieu, J. L. (2016). Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads. *IEEE Transactions on Power Systems*, *32*(2), 1378–1388.

Zhao, L., & Zeng, B. (2012). An exact algorithm for two-stage robust optimization with mixed integer recourse problems. *submitted, available on Optimization-Online. org*.

Zhao, P., Gu, C., Cao, Z., Shen, Y., Teng, F., Chen, X., ... Li, S. (2020). Data-driven multi-energy investment and management under earthquakes. *IEEE Transactions on Industrial Informatics*, 1-1.

Zheng, X., & Chen, H. (2020). Data-driven distributionally robust unit commitment with wasserstein metric: Tractable formulation and efficient solution method. *IEEE Transactions on Power Systems*, *35*(6), 4940–4943.

APPENDICES

A. DUAL OPERATIONAL PROBLEM

When the loss function in (3a) is defined as $L(\boldsymbol{y_{dht}}) = \sum_{i\in\mathcal{G}^F} C_i^g P_{dhit}^g + \sum_{b\in\mathcal{B}} C_b^{LS} P_{bdht}^{LS}$, the dual problem of model (3a)-(3u) is as follows:

$$
\begin{aligned}
\max_{\pi} \quad & \Big\{ \sum_{t\in\mathcal{T}} \sum_{d\in\mathcal{D}} \sum_{h\in\mathcal{H}} \Big\{ \sum_{i\in\mathcal{G}^F} -\overline{P_i^g}(1-\varphi_{b(i),t})\pi_{dhit}^1 \\
& + \sum_{i\in\mathcal{G}^V} -\overline{P_{dhi}^g}(1-\varphi_{b(i),t})\pi_{dhit}^2 \\
& + \sum_{j\in\mathcal{L}^C} -(\overline{f_j}+x_j^l)\big((1-\zeta_{jt})(\pi_{dhjt}^4+\pi_{dhjt}^5)+(1-z_{jt})(\pi_{dhjt}^6+\pi_{dhjt}^7)\big) \\
& + \sum_{j\in\mathcal{L}\setminus\mathcal{L}^C} -\overline{f_j}\big((1-\zeta_{jt})(\pi_{dhjt}^4+\pi_{dhjt}^5)+(1-z_{jt})(\pi_{dhjt}^6+\pi_{dhjt}^7)\big) \\
& + \sum_{j\in\mathcal{L}} -M z_{jt}(\pi_{dhjt}^8+\pi_{dhjt}^9) \\
& + \sum_{b\in\mathcal{B}} \big(- D_{bdh}(\pi_{bdht}^3+\pi_{bdht}^{12}) + \underline{\theta}\pi_{bdht}^{10} - \overline{\theta}\pi_{bdht}^{11}\big) \\
& + \sum_{b\in\mathcal{B}} (1-\varphi_{bt})\big(- \overline{E_b}\pi_{bdht}^{13} - \overline{P_b}(\pi_{bdht}^{14}+\pi_{bdht}^{15})\big)\Big\}\Big\} \\
\text{s.t.} \quad & \Big\{ \pi_{dhit}^1 + \pi_{b(i),d,h,t}^{12} + C_i^g \tau_t \frac{p_d^D}{H} \geq 0 \quad \forall i\in\mathcal{G}^F \\
& \pi_{dhit}^2 + \pi_{b(i),d,h,t}^{12} \geq 0 \quad \forall i\in\mathcal{G}^V \\
& \pi_{bdht}^3 + \pi_{bdht}^{12} + C_b^{LS}\tau_t\frac{p_d^D}{H} \geq 0 \quad \forall b\in\mathcal{B} \\
& - \pi_{dhjt}^4 + \pi_{dhjt}^5 - \pi_{dhjt}^6 + \pi_{dhjt}^7 - \pi_{dhjt}^8 + \pi_{dhjt}^9 - \pi_{b^+(j),d,h,t}^{12} \\
& \qquad + \pi_{b^-(j),d,h,t}^{12} = 0 \quad \forall j\in\mathcal{L}\Big\}, \quad \forall d\in\mathcal{D}, h\in\mathcal{H}, t\in\mathcal{T} \\
& \Big\{ \sum_{j|b^+(j)=b} B_j(\pi_{dhjt}^8 - \pi_{dhjt}^9) + \sum_{j|b^-(j)=b} B_j(\pi_{dhjt}^9 - \pi_{dhjt}^8) \\
& \qquad - \pi_{bdht}^{10} + \pi_{bdht}^{11} = 0 \quad \forall h\in\mathcal{H}
\end{aligned}
$$

$$
\begin{aligned}
&- \pi^{12}_{b,d,h-1,t} + \pi^{14}_{b,d,h-1,t} - p^H \eta^{cha} \pi^{16}_{bdht} \geq 0 \quad \forall h \in \mathcal{H} \setminus \{1\} \\
&- \pi^{12}_{b,d,h=H,t} + \pi^{14}_{b,d,h=H,t} - p^H \eta^{cha} \pi^{17}_{bdt} \geq 0 \\
&\pi^{12}_{b,d,h-1,t} + \pi^{15}_{b,d,h-1,t} + \frac{p^H}{\eta^{dis}} \pi^{16}_{bdht} \geq 0 \quad \forall h \in \mathcal{H} \setminus \{1\} \\
&\pi^{12}_{b,d,h=H,t} + \pi^{15}_{b,d,h=H,t} + \frac{p^H}{\eta^{dis}} \pi^{17}_{bdt} \geq 0 \\
&\pi^{13}_{b,d,h-1,t} + \pi^{16}_{b,d,h-1,t} - \pi^{16}_{bdht} \geq 0 \quad \forall h \in \mathcal{H} \setminus \{1,2\} \\
&\pi^{13}_{b,d,h=1,t} - \pi^{16}_{b,d,h=2,t} + \pi^{17}_{bdt} \geq 0 \\
&\pi^{13}_{b,d,h=H,t} + \pi^{16}_{b,d,h=H,t} - \pi^{17}_{bdt} \geq 0 \Big\}, \quad \forall b \in \mathcal{B}, d \in \mathcal{D}, t \in \mathcal{T} \\
&\boldsymbol{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7, \pi_8, \pi_9, \pi_{10}, \pi_{11}, \pi_{13}, \pi_{14}, \pi_{15} \geq 0}
\end{aligned}
$$

B. NETWORK COMPONENTS

The nomenclature and the macro-zone of each component mentioned in this work are presented in Table B.1. Macro-zones are defined as in MinEnergia (2020), namely: Far north, Near north, Center, South.

Table B.1. Network components nomenclature and location

Lines		
Compact name	Line	Macro-zone
L_1	Charrua220 - Concepcion154	Center
L_2	Ohiggins220_BP2 - Farellon220	Far north
L_3	LVilos220 - DonaCarmen220	Near north
L_4	Parinacota220 - Condores220	Far north
L_5	Farellon220 - Chimborazo220	Far north
L_6	Ohiggins220_BP2 - Puri220	Far north
L_7	Miraje220 - Atacama220_BP2	Far north
L_8	LVilos220 - Nogales220	Near north
L_9	Laberinto220 - NvaZaldivar220	Far north
L_{10}	SanLuis220 - ASanta220	Center
L_{11}	NvaZaldivar220 - Sulfuros220	Far north
L_{12}	Crucero220 - Tocopilla220_BP2	Far north
L_{13}	Palestina220 - Domeyko220	Far north
L_{14}	Ohiggins220_BP1 - Palestina220	Far north
L_{15}	Encuentro220 - Collahuasi220	Far north
L_{16}	Kapatur220_BP1 - Laberinto220	Far north
Storage systems		
Compact name	Attachment substation	Macro-zone
B_1	Tamarugal066	Far north

C. WEIGHTS OF SEISMIC SOURCES

The weights of seismic sources used in Section 6 are provided in the following table.

Table C.1. Seismic sources' weights.

Zone	Weight
Zone 1	0.131607
Zone 2	0.224456
Zone 3	0.058417
Zone 4	0.094556
Zone 5	0.337124
Zone 6	0.088668
Zone 7	0.065171

D. EXPANSION COSTS OF TRANSMISSION LINES

The costs of transmission lines expansion used in Section 6, based on the information provided by MinEnergia (2020), are as shown in the following table.

Macro-zone	220 kV (250 MW)	500 kV (1500 MW)
Far north	173	452
Near north	226	590
Center	366	956
South	375	981

Table D.1. Expansion costs of transmission lines.

The reference cost for 220 kV lines for the near north, the center, and the south, was estimated based on the expansion cost relationship between the 500 kV and 220 kV lines in the large north.

E. MORE EXAMPLES OF ATTACKER-DEFENDER SOLUTIONS

Here we present another example of a solution identified by the attacker-defender model.

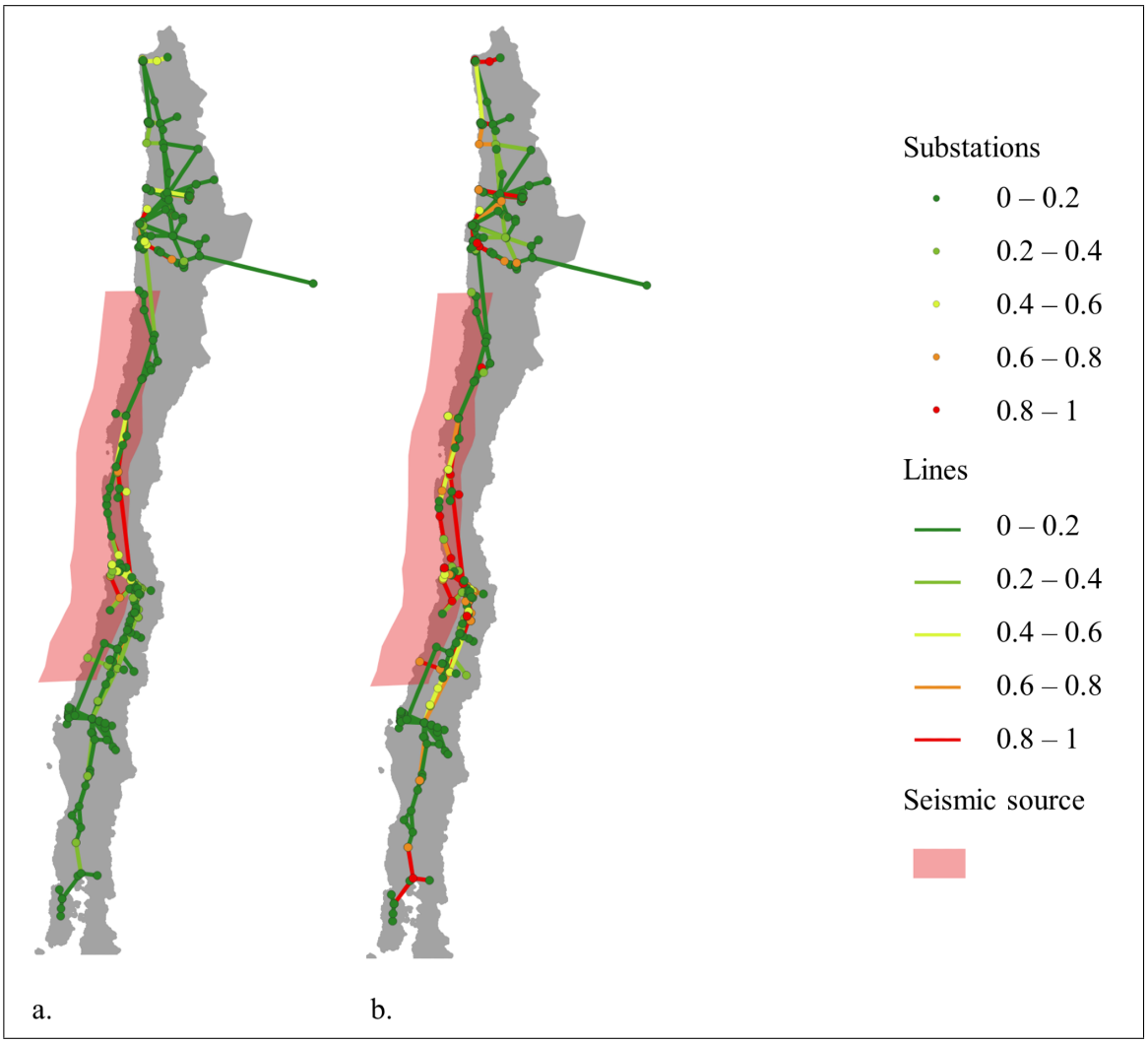

Figure E.1. Maps of damage three days after the occurrence of an earthquake originated from seismic source Zone 2: a) $\gamma = 95\%$, and b) $\gamma = 99\%$