POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

# Geospatial Mapping and Visualization for Reliability Analysis of the Italian Transmission Grid

Tesi di Laurea Magistrale in
Automation and Control Engineering - Ingegneria dell'Automazione

Author: **Yang Xuan**

Student ID: 10948738
Advisor: Prof. Loredana Cristaldi
Co-advisors: Prof. Christian Laurano, Andrea Moschetti, Luca Martiri

Academic Year: 2025-26

# Abstract

This thesis presents a data integration and geospatial visualization framework for reliability analysis of the Italian National Transmission Grid. It links Terna's facility registers with outage records by resolving heterogeneous identifiers and enriching line endpoints with geographic coordinates. The workflow includes structured text normalization, voltage- and area-based candidate filtering, two-pass fuzzy matching, a fallback dictionary for missing user codes, and geospatial validation against physical line lengths. The resulting standardized dataset enables spatially explicit RAMS metrics, including length-normalized Severity Factor, and supports interactive mapping of vulnerabilities. Empirical results show substantial gains in matched assets and reveal regional patterns that differ from raw failure counts, with high-frequency outages concentrated in southern regions and pronounced high-impact anomalies in specific northern areas. The study demonstrates that integrating logical outage data with physical topology provides more actionable insights for grid resilience assessment and maintenance planning.

**Keywords:** Geospatial Mapping, Reliability Analysis, Severity Factor, Italian Transmission Grid

# Abstract in lingua italiana

Questa tesi presenta un framework di integrazione dati e visualizzazione geospaziale per l'analisi dell'affidabilità della Rete di Trasmissione Nazionale italiana. Collega i registri degli impianti di Terna con i record di interruzione risolvendo identificatori eterogenei e arricchendo gli endpoint di linea con coordinate geografiche. Il flusso di lavoro include la normalizzazione del testo strutturato, il filtraggio dei candidati basato sulla tensione e sull'area, il fuzzy matching a due passaggi, un dizionario di fallback per i codici utente mancanti e la convalida geospaziale rispetto alle lunghezze fisiche delle linee. Il set di dati standardizzato risultante consente metriche RAMS spazialmente esplicite, incluso il fattore di severità normalizzato per lunghezza, e supporta la mappatura interattiva delle vulnerabilità. I risultati empirici mostrano guadagni sostanziali negli asset abbinati e rivelano modelli regionali che differiscono dai conteggi grezzi dei guasti, con interruzioni ad alta frequenza concentrate nelle regioni meridionali e anomalie pronunciate ad alto impatto in specifiche aree settentrionali. Lo studio dimostra che l'integrazione dei dati logici sulle interruzioni con la topologia fisica fornisce informazioni più fruibili per la valutazione della resilienza della rete e la pianificazione della manutenzione.

**Parole chiave:** Mappatura Geospaziale, Analisi di Affidabilità, Fattore di Severità, Rete di Trasmissione Nazionale

# Contents

# 1 | Introduction

## 1.1. Context and Motivation

Modern society depends heavily on a stable and continuous power supply to support everything from daily life and healthcare to transportation and communication. Large-scale power outages can quickly disrupt these essential services, leading to severe economic and social consequences.

The transmission grid acts as the backbone connecting power generation to end users. Its reliability is closely tied to national economic stability. As power systems become more complex and face changing environmental conditions, transmission line failures can cause significant damage. Therefore, it is crucial to analyze the causes of these failures, identify weak points, and develop preventive maintenance strategies to improve grid resilience [1, 2].

In Italy, the National Transmission Grid (Rete di Trasmissione Nazionale, RTN) is managed by the Transmission System Operator (TSO), Terna. Because Terna makes its data publicly available, researchers have access to detailed annual grid operations and outage records. Using this historical data, recent studies have performed Reliability, Availability, Maintainability, and Safety (RAMS) analyses on Italian Overhead Transmission Lines (OTL). These works introduced metrics like the Severity Factor (SF) to measure the impact of component failures and help optimize maintenance efforts [3].

However, existing research often overlooks a key aspect: the spatial dimension. Previous studies mainly focused on statistical and time-series analyses using tabular data. While this data contains valuable details about failures (such as duration and energy lost), it lacks a direct connection to the physical location of the grid assets. Transmission lines cover large and diverse geographical areas, and failures often cluster in specific locations or terrains. Looking only at tables makes it hard for grid operators to spot these regional risks.

This research aims to address this gap by linking the tabular outage data with physical locations. Specifically, it connects the line lengths and site names from Terna's facility list (Elenco Impianti RTN) to the historical database of grid failures. Mapping this data geographically helps identify vulnerable areas and failure hotspots that are not visible in simple tables. Ultimately, this approach provides a practical tool for grid managers to make better decisions and improve the overall reliability of the transmission network.

## 1.2. Problem Statement

Recent work by Moschetti et al. successfully integrated Terna's data into a unified relational database, greatly simplifying data access and creating a strong foundation for future research [4]. However, a major limitation remains: even though the database contains a vast amount of historical outage records, the failure data for each user site is treated independently, while the physical connections between sites are not represented. As a result, the spatial context and topological structure of the grid are missing from the current logical data.

To solve this problem, this research aims to connect the physical grid relationships contained in Terna's facility tables with the outage records in the database. By matching the names and codes of user sites, isolated failure events can be linked directly to the corresponding physical infrastructure.

Through this mapping process, it becomes possible to retrieve geographic details such as the coordinates of user sites and the lengths of transmission lines. Moreover, by visualizing the integrated dataset on an interactive map, the complex network can be explored more intuitively, supporting deeper spatial analyses of grid reliability.

## 1.3. Goals of the Thesis

Although the geographic coordinates of individual user sites are available, the logical database does not contain explicit information about the physical interconnections between them, namely the transmission lines, including their spatial trajectories and physical lengths. To enable a reliability analysis that accounts for both regional distribution and line length, it is therefore necessary to integrate the logical records with the corresponding physical network infrastructure. This integration allows the extraction of line-length attributes required for subsequent length-normalized assessments of the Italian National Transmission Grid. Based on this objective, the study is structured around the following three main tasks:

1. **Mapping Physical Infrastructure to Logical Outage Records.** The first

objective is to break down the existing data silos. By extracting site names and codes from Terna's facility tables, this research aims to develop a matching process that links the isolated outage events in the database to their corresponding physical sites. This mapping strategy is designed to overcome data inconsistencies and successfully retrieve crucial spatial attributes, specifically the exact geographic coordinates of the user sites and the actual physical lengths of the transmission lines.

2. **Development of an Interactive Web Map Tool.** Once the complete spatial and relational dataset is obtained, the second objective is to construct a user-oriented visual analytics platform. This goal focuses on developing an interactive web map that transforms massive and abstract database tables into intuitive geographic visualizations. This tool will not only clearly reconstruct the topological structure of the grid but also allow grid operators to visually observe the spatial distribution of failure events across different geographic regions and line corridors.

3. **Introduction of Length-Normalized Spatial Metrics.** The first two objectives lay the foundation for more advanced reliability analysis. The third objective is to utilize the extracted line length information to innovate traditional statistical evaluation methods. This study will introduce length-normalized metrics, discarding the potentially misleading comparisons based on absolute failure counts. By calculating the failure rate per unit of length and projecting the severity factor into real geographic space, this objective aims to more fairly and accurately identify localized vulnerabilities and high-risk failure hotspots hidden within the grid.

## 1.4. Thesis Outline

The remainder of this thesis is organized as follows:

Chapter 2 (Background) reviews the fundamental concepts and the existing literature relevant to this study. It introduces the structure of the Italian Transmission Grid, common reliability assessment metrics, and the current state of the art and previous works.

Chapter 3 (Data Integration and Spatial Matching Strategy) details the core methodology used to bridge the gap between logical records and physical assets. It explains the extraction of asset data, the entity resolution process, and the strategies applied to overcome data inconsistencies when mapping outage events to their geographic coordinates.

Chapter 4 (Geospatial Mapping and Visualization) presents the development of the interactive visual analytics platform. This chapter describes the system architecture, the georeferencing process of the transmission infrastructure, and the features of the web map designed to support operational decision-making.

Chapter 5 (Reliability Analysis and Results) discusses the findings derived from the geospatial mapping. It introduces the methodology for calculating length-normalized metrics and presents a detailed geographic evaluation of grid outages, regional vulnerabilities, and the spatial distribution of the severity factor.

Chapter 6 (Conclusions) summarizes the main contributions of this research, acknowledges the limitations of the current spatial model, and suggests potential directions for future developments.

# 2 | Background

## 2.1. The Italian Transmission Grid

RTN is the core backbone of the Italian power system. It is responsible for the efficient and safe transmission of electricity from generation facilities to load centers and distribution networks across the country. The grid is centrally managed and dispatched by Terna. Controlling over 97% of the high-voltage electricity infrastructure, Terna is the largest independent grid operator in Europe and manages nearly 75,000 km of transmission lines and approximately 1,000 electrical substations [5]. The RTN spans highly diverse geographical and climatic environments, and its physical architecture and operational strategies directly determine the energy security of the entire nation.

### 2.1.1. Voltage Levels and Infrastructure Architecture

The physical infrastructure of the transmission grid primarily consists of tens of thousands of kilometers of transmission lines. To meet different distance and capacity requirements, RTN assets are classified into voltage levels [6]:

- **Extra-High Voltage Backbone (AAT – Altissima Tensione).** This includes 380 kV and 220 kV lines. This section of the grid handles long-distance transmission of large electricity volumes at the national level. Its physical topology is highly meshed (magliata), meaning each node (substation or large power plant) is typically connected via two or more lines to ensure compliance with the N-1 security criterion and provide high system redundancy [7].
- **High Voltage Sub-transmission Network (AT – Alta Tensione).** This includes 150 kV, 132 kV, and a smaller portion of lines at or below 100 kV (e.g., 60 kV). The 150 kV and 132 kV lines usually operate in a meshed configuration within independent regional islands. Conversely, lines below 100 kV mostly adopt a radial (radiale) topology, which is mainly responsible for distributing electricity from the backbone to lower-level distributors or large energy-intensive industrial users, such as steelworks and chemical plants [7].

### 2.1.2. Operational and Geographical Areas

Due to Italy's elongated peninsular geography and complex mountainous terrain, the national grid is not physically homogeneous. For efficient localized management and dispatching, Terna divides the RTN into eight territorial areas (Area Territoriale) [6]. These areas define the physical boundaries for Terna's daily maintenance activities and serve as the core spatial scale for the spatio-temporal reliability analysis in this study:

- Torino: Northwestern regions of Piedmont, Liguria, and Aosta Valley.
- Milano: Lombardy region.
- Padova: Northeastern regions of Veneto, Trentino-Alto Adige, and Friuli-Venezia Giulia.
- Firenze: Central-northern regions of Tuscany and Emilia-Romagna.
- Roma: Central regions of Lazio, Umbria, Marche, Abruzzo, and Molise.
- Napoli: Southern regions of Campania, Apulia, Basilicata, and Calabria.
- Palermo: Sicily.
- Cagliari: Sardinia.

### 2.1.3. Interconnections and Cross-border Grid

The RTN is not an isolated, closed system. It connects the eight areas through numerous inter-area lines (collegamento interarea). In particular, when connecting the islands to the mainland, the RTN employs High-Voltage Direct Current (HVDC) technology, such as the SA.CO.I and SA.PE.I submarine cable projects linking Sardinia to the mainland [5]. Furthermore, the RTN is integrated with the power grids of France, Switzerland, Austria, and Slovenia in northern Italy, and connects to the Greek grid in the south via the Galatina converter station [6]. These cross-border interconnections not only guarantee national energy security but also serve as a crucial hub in the European electricity market.

## 2.2. Reliability Assessment Metrics

### 2.2.1. Traditional RAM Evaluation Metrics

Traditionally, RAM analysis has been mainly used to support system design requirements and to assess the operational performance of utility assets. It mainly covers the following three dimensions:

1. **Reliability (Failure Rate).** Reliability refers to the ability of a system to perform its required function under specified conditions for a specified period of time. It is commonly quantified by the failure rate. For transmission line networks, the failure rate can be defined as the number of failures per unit length within a given voltage level or area over a given time period (typically one year). The formulation is:

$$\lambda = \frac{N_i}{L_i \times T} \tag{2.1}$$

   where $N_i$ is the total number of failures, $L_i$ is the total length of transmission lines, and $T$ is the observation period.

2. **Maintainability (Repair Rate).** Maintainability describes the ability of the system to be restored to service within a specified time after a failure. It is commonly quantified by the repair rate $\mu$, which is the reciprocal of the mean time to repair for the $i$-th line ($\mathrm{MTTR}_i$):

$$\mu = \frac{1}{\mathrm{MTTR}_i} = \frac{N_i}{L_i \sum_{n=1}^{N_i} t_n} \tag{2.2}$$

3. **Availability.** Availability is the ability of an item to be in a state to perform a required function under given conditions at a given time instant or over a given time interval, assuming that the required external resources are provided. For a line, it can be defined as the percentage of time that the line is available to perform its function. It can be calculated as:

$$A = \frac{\mathrm{MTTF}_i}{\mathrm{MTBF}_i} \cdot 100\% \tag{2.3}$$

   where MTBF is the mean time between failures, given by the sum of the mean time to fail (MTTF) per km of line, which is calculated as:

$$\mathrm{MTTF}_i = \frac{T - \displaystyle\sum_{n=1}^{N_i} \frac{t_n}{L_i}}{\frac{N_i}{L_i}} \tag{2.4}$$

   and the MTTR per km of line (over a certain observation period T).

4. **Undelivered Energy.** A complementary reliability trend indicator is the total energy not delivered because of outages. For a given category (or line class), it can

be computed as the sum of interrupted power over outage durations:

$$E_{U,i} = \sum_{i=1}^{N_i} P_i t_i \tag{2.5}$$

where $P_i$ is the interrupted power and $t_i$ is the outage duration for the $i$-th failure event.

### 2.2.2. Severity Factor Methodology

Frequency-based indicators present an important limitation: they treat all outages as equivalent, even though the real impact strongly depends on the affected components and the magnitude of interrupted power. A high-frequency, low-impact outage does not have the same consequence as a low-frequency event that causes major industrial losses. To address this issue, Martiri et al. introduced the SF, shifting the evaluation focus from outage counts to energy not supplied (ENS) [3].

1. **Base Severity Factor.** The base SF for a given outage category (or area) is defined as the share of undelivered energy associated with that category over the total undelivered energy in the system:

$$\mathrm{SF}_j = \frac{\sum_{i=1}^{N_i} (P_i \cdot t_i)}{\sum_{N_i} \left( \sum_{i=1}^{N_i} (P_i \cdot t_i) \right)} = \frac{E_{U,i}}{E_{U,\mathrm{tot}}} \tag{2.6}$$

   where $\mathrm{SF}_j$ is the severity factor for outage category $j$, $N_i$ is the total number of outages in that category, $P_i$ is the mean interrupted power at the relevant user location, and $t_i$ is the outage duration. The numerator $E_{U,i}$ represents the cumulative undelivered energy for category $j$, while the denominator $E_{U,\mathrm{tot}}$ is the total undelivered energy across all categories.

2. **Double Normalized Severity Factor.** Combining frequency and length normalization yields the final double-normalized severity factor. A frequency weight is introduced to mitigate bias from heterogeneous outage counts:

$$\gamma_j = \frac{\sum_{i=1}^{C} K_i}{K_j} \tag{2.7}$$

   where $K_j$ is the number of outages in category $j$ and $\sum_{i=1}^{C} K_i$ is the total number of outages across all $C$ categories. A length weight is introduced to remove the base-rate effect of different line lengths:

$$\lambda_i = \frac{L_{\text{tot}}}{L_i} \tag{2.8}$$

where $L_{\text{tot}}$ is the total network length and $L_i$ is the length of the line class (or area) under analysis. The final form is:

$$\text{SF}_j = \frac{\lambda_i \gamma_i \sum_{i=1}^{N_i} (\text{ENS}_i)}{\sum_{N_i} \left( \sum_{i=1}^{N_i} \gamma_i \, (\text{ENS}_i) \right)} \tag{2.9}$$

This formulation shifts the assessment from absolute ENS to a normalized, comparable severity measure across heterogeneous categories [3].

## 2.3. Related Works

Power grid reliability assessment and spatial vulnerability analysis are core interdisciplinary topics in modern power system engineering and complex network science. This section reviews the relevant literature from three perspectives: the macroscopic assessment of power grid structural vulnerability, the application of geospatial analysis, and specific empirical studies on the RTN.

### 2.3.1. General Assessment of Power Grid Reliability and Structural Vulnerability

Globally, scholars have transitioned from pure probabilistic statistics to complex network theory when quantifying transmission network performance. Traditional reliability assessments widely adopt the RAMS framework, focusing on the calculation of failure rates and MTTR at the system level [1]. However, modern research further indicates that grid vulnerability is not uniformly distributed across the entire network. In their classic study on the North American power grid, Albert et al. abstracted the power system into a topological network, demonstrating that while the grid exhibits strong robustness against random failures, it displays extreme structural vulnerability when facing the targeted loss of specific critical nodes [8]. This finding has led to an academic consensus: evaluating grid reliability cannot solely rely on global absolute failure statistics; it is essential to deeply analyze the heterogeneous roles played by different components and lines within the system topology.

### 2.3.2. Geospatial Analysis in Power Systems

Because transmission infrastructure is widely distributed across highly complex geographical environments, purely logical topological analysis has gradually revealed its limitations. Integrating topology with GIS has become a research frontier in recent years. Bompard et al. pointed out that the physical span and geographical constraints of transmission lines directly determine the propagation paths of disasters and the risk of cascading failures [9]. Projecting the abstract grid graph onto a real geographic coordinate system not only reveals the localized clustering of vulnerabilities but also significantly enhances the dispatching efficiency of disaster recovery and preventive maintenance. However, in practical engineering, achieving high-precision data alignment between multi-source heterogeneous grid operation records and GIS spatial architectures remains a highly challenging technical bottleneck.

### 2.3.3. Empirical Studies on the RTN

Focusing on the RTN, the research team at Politecnico di Milano utilized the massive historical database published by Terna to establish a systematic RAMS assessment framework based on real operational data [3, 10]. Unlike the aforementioned complex network simulations based on theoretical physics, previous works are entirely driven by real data and pioneered the introduction of the SF to comprehensively quantify the energy loss caused by failures.

The research achieved significant results across multiple logical slices: at the component level, OHTL were confirmed as the most vulnerable elements; in the temporal dimension, a severe trend of declining system availability was revealed; in terms of physical structure, by introducing a normalized SF incorporating a line length parameter, the high-voltage subtransmission network (132 kV and 150 kV) was accurately pinpointed as a substantially high-risk group; in the failure cause analysis, external environmental interference was identified as the dominant factor leading to severe outages.

### 2.3.4. Limitations of Current Studies and Proposed Contribution

In summary, incorporating physical topology and spatial dimensions into power grid reliability research is highly advantageous for vulnerability assessment. However, owing to the confidentiality of critical infrastructure, database fragmentation, and unstructured proprietary formats, acquiring complete geospatial data and linking it seamlessly with historical outage logs remains a pervasive data barrier globally. Consequently, a substantial body of recent international literature still relies predominantly on tabular, logical data, evaluating grid reliability through non-spatial statistical models [11, 12].

Previous empirical research on the Italian National Transmission Grid [3, 10]aligns with this widespread methodological constraint. While these studies conducted highly detailed tabular analyses, even introducing "line length" into the mathematical formulation of the severity framework, their analytical scope was inherently confined by these data silos to abstract voltage categories within the logical database. Without an explicit spatial mapping between logical entities and actual geographical corridors or the eight operational areas, the absence of physical topological attributes prevents grid managers from intuitively discerning which specific geographic regions bear the highest localized risks.

This research is dedicated to overcoming this pervasive data barrier. By constructing a mapping dictionary between logical outage logs and real spatial infrastructure, this study bridges the data silos and reconstructs the missing spatial dimension in the regional empirical analysis of the RTN. This integration shifts the analytical paradigm from non-spatial tabular statistics to localized topological evaluation, providing a novel and actionable decision-support perspective for the spatial vulnerability assessment of the power grid.

# 3 | Data Integration and Spatial Matching Strategy

## 3.1. The Foundation: Database of Grid Locations and Outages

Following Moschetti et al., to address inconsistent formats (mostly PDFs) and frequent table-structure changes in Terna's historical data (2008–2023), a data-cleaning pipeline was designed and a standardized, high-quality relational database was built [4]. At the conceptual level, the database includes three core entities: Outages (Guasti), User Sites (Siti Utente), and Locations (Località). The logical schema and the meaning of its fields are summarized below.

**Outages.** Outages is the core repository for recording grid outage events, effectively consolidating multiple associated records triggered by the same fault source through a shared annual fault number. This table not only details the specific date and time of the outage, the type of network equipment causing the anomaly and its voltage level, alongside the officially classified fault cause codes, but also links to the specific affected sites via user codes. Crucially, it records the decisive physical metrics used to quantify the severity of the fault, such as the power interrupted at the moment of failure (PI), the outage duration, and the highly critical total Energy Not Supplied or Energy Not Received (ENS/ENR).

**User Sites.** User Sites integrates the metadata and connection characteristics of all specific physical sites connected to the high-voltage grid. To properly address data loss or overwriting issues caused by changes in site attributes (such as ownership transfers or voltage level modifications) over the years, this table innovatively adopts a time-tracking mechanism. It utilizes the unique site identifier code along with the data's "start year" and "end year" as a composite primary key to define the valid time window of the information. Furthermore, the core fields in the table clearly describe the actual name of the site, its owner, functional classification (load or generation), operating voltage parameters, and the topological connection configuration for the site's direct interaction with the grid.

**Locations.** Locations is specifically designed to store the geospatial information of the user sites, serving as a critical bridge for realizing the spatial visualization of grid data. Given that the original foundational data published by the operator is limited to administrative divisions at the municipality and province levels, this table builds upon this base by utilizing external geocoding tools (such as Nominatim and Google Elevation API) to perform an in-depth expansion of the spatial dimension. This operation introduces the crucial precise fields of latitude, longitude, and elevation for each location, thereby making it possible to accurately map the originally isolated logical outage records onto the real physical spatial network.

A simplified version of the Entity-Relationship (E-R) schema of the database is showed in Figure 3.1.

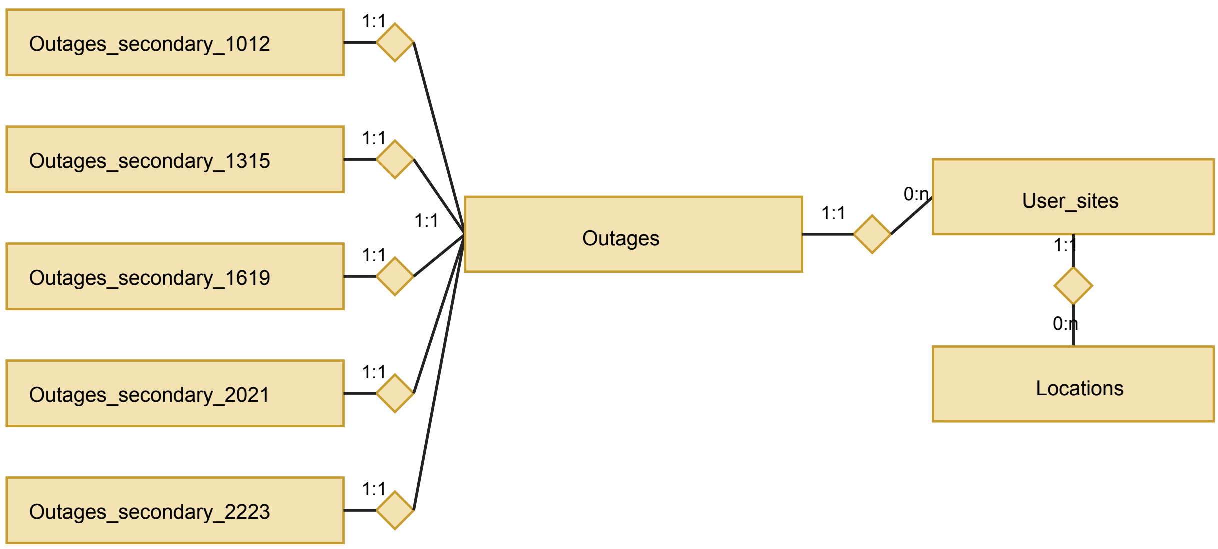


Figure 3.1: Entity-Relationship (E-R) schema of the database.

## 3.2. Extraction of Terna Transmission Lines Data

In this section, the structure of the transmission line data table published by Terna is introduced, dividing it into a line section and a substation section. Furthermore, it presents the integrated table obtained after merging Terna's line information with the code information from the database.

### 3.2.1. Structure of Transmission Lines Table

When processing the first half of the asset register (i.e., transmission lines, Linee), this study extracted and cleaned the data into 8 independent CSV tables, corresponding precisely to Terna's 8 operational areas (Area Territoriale). During this process, the focus was on capturing the following key physical fields:

**Core Coding Rule: The First Digit X (Asset Type).** It explicitly distinguishes topological "edges" from "nodes": if the digit is 2, it represents a "Line" (Linea) asset; if the digit is 3, it represents a "Substation" (Stazione) asset.

**Core Coding Rule: The Second Digit Y (Voltage Level and Function).** For lines (when X=2): 0 denotes a direct current (DC) line regardless of voltage; 1 denotes a 380kV alternating current (AC) main trunk line; 2 denotes a 220kV AC line; 3 denotes a 120÷150kV AC secondary line; 4 denotes a 30÷70kV line; and 5 denotes a line below 30kV. For substations (when X=3): 0 denotes an AC/DC converter station; 1 denotes a substation with a maximum operating voltage of 380kV; 2 denotes a 220kV substation; 3 denotes a substation operating at ≤150kV; and 5 denotes a switching station (Smistamento).

**Line Section.** This section is strictly categorized according to Terna's eight operational areas and voltage levels. It specifies the unique CODICE for each transmission line, the line name (DENOMINAZIONE, typically formatted as origin-destination) which reveals the topological connection relationship between nodes, and the crucial physical length of the line circuit. The structure of transmission lines table is showed in Table 3.1.

| CODICE | DENOMINAZIONE | km |
|---|---|---|
| 21.301 | SAN ROCCO – TURANO | 16,924 |
| 21.302 | BRUGHERIO – TAVAZZANO | 31,500 |
| 21.303 | TAVAZZANO – TURANO | 26,811 |
| 21.304 | CASSANO – VERDERIO | 17,627 |
| 21.305 | LONATO – NAVE | 29,332 |

Table 3.1: Structure of Transmission Lines Table

### 3.2.2. Independent Substation Table

In addition to the transmission lines divided by areas, the second half of the Elenco Impianti RTN details the asset information of electrical substations (Stazioni Elettriche).

During data processing, this study did not mix this data with the lines, but rather extracted it separately and saved it as a single, new table.

This separated processing significantly enhances the fault tolerance and completeness of the data in practical engineering. During the subsequent foreign key mapping between the outage logs and the user sites table, it is common to encounter situations where some sites cannot be found in the database using the conventional alphabetical user code (codice_utente) due to irregular recording or system code changes. To address this data missing issue, this independently established substation table acts as a critical "Fallback Dictionary". When the conventional alphabetical code fails, the system invokes the numeric code recorded in the substation table as an alternative identifier for a secondary search. This straightforward engineering design effectively overcomes the problem of identifier inconsistency and maximally salvages outage records that would otherwise potentially be discarded [13, 14].

| CODICE | DENOMINAZIONE | Provincia | tensione | AREA |
|---|---|---|---|---|
| 31.201 | RONDISSONE | Torino | 380 | Torino |
| 31.202 | CASANOVA | Torino | 380 | Torino |
| 31.203 | PIOSSASCO | Torino | 380 | Torino |
| 31.204 | LEINI | Torino | 380 | Torino |
| 31.205 | VENAUS | Torino | 380 | Torino |

Table 3.2: Electrical Stations Table.

## 3.3. Matching Strategy

### 3.3.1. Site Matching Logic

After extracting the line data, the crucial step is matching the origin and destination of the lines from the PDF (referred to as sito_1 and sito_2 in the code) with the sites in the database. The system extracts the name and voltage level of each transmission line and uses these two parameters to search within the consolidated database of user sites (siti_utente) and locations (località).

### 3.3.2. Acquiring Foreign Keys and Coordinates

If the match is successful, the algorithm extracts the user code (codice_utente) of the site. This code acts as a Foreign Key, linking the physical transmission line in reality to the outage table (guasti) in the database. Simultaneously, the algorithm fetches the latitude and longitude coordinates (lat, lon) of these two sites, giving the transmission line a real geographic position on the map.

When matching official texts with the database, issues like typos, inconsistent abbreviations, or missing codes frequently occur during the entity resolution process [15]. To ensure high matching accuracy, the following cleaning and validation mechanisms were designed in the code.

### 3.3.3. Text Normalization

The naming of sites in Terna's historical documents is highly irregular. To facilitate programmatic searching, the code first uses Regular Expressions (Regex) to standardize common Italian abbreviations (for example, replacing "S." with "SAN" and "C.LE" with "CENTRALE"). This data cleaning step eliminates interference caused by different writing habits and significantly improves the initial match rate [16].

### 3.3.4. Two-Pass Fuzzy Matching

To handle spelling errors or different operational prefixes (such as CP, CS), the code uses the thefuzz library to calculate string similarity based on the Levenshtein distance [17].

Step 1: The system retains all operational prefixes and attempts a complete fuzzy match.

Step 2: If no results are found, a fallback mechanism is triggered. The system strips the operational prefixes and performs a secondary search using only the core name. Finally, by calculating the similarity score, it selects the best-matching site in the database.

### 3.3.5. Fallback Mechanism

It is common to find sites in the database missing the standard alphabetical user codes. To prevent these line records from being discarded, the previously extracted substation table (Stazioni Elettriche.csv) is utilized as a "fallback dictionary." When the standard search fails, the system checks this substation table. If a match is found, it replaces the missing alphabetical code with the substation's numeric code. This straightforward design salvaged a massive amount of data that would have otherwise been lost.

### 3.3.6. Output of the Enhanced Result Table

Following this pipeline, the output is no longer merely cleaned text, but an enhanced intermediate result table. Compared to the raw data, the new table successfully supplements the standard site codes (featuring alphabetical codes) and precise geographical coordinates for both endpoints of every line. This highly standardized table thoroughly eliminates the irregularities and spatial blind spots inherent in the raw data, enabling each physical transmission line to seamlessly interface with the macroscopic system. Consequently, it lays a directly actionable foundational data basis for subsequent network visualization and spatial topological analysis.

### 3.3.7. Matching Results

After this data processing pipeline, we obtained a more refined data table, as illustrated in the figure.

| codice | sito_1 | codice_sito_1 | sito_2 | codice_sito_2 | tensione | lunghezza | stato_estero | lat_1 | lon_1 | lat_2 | lon_2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 23.635 | ZOGNO | MMILZOGM | SORISOLE | MMILSORM | 132 | 8.325 | | 45.7937 | 9.6662 | 45.7446 | 9.66119 |
| 23.636 | BERTINI | MMILBERM | VERDERIO | MMILVRDM | 132 | 3.572 | | 45.6467 | 9.46594 | 45.6617 | 9.43229 |
| 23.637 | VERDERIO | MMILVRDM | REA DALMINE | MMILREAM | 132 | 15.763 | | 45.6617 | 9.43229 | 45.6472 | 9.60192 |
| 23.639 | SORISOLE | MMILSORM | CURNO | MMILRNOM | 132 | 5.573 | | 45.7446 | 9.66119 | 45.6915 | 9.61097 |
| 23.646 | SAN PIETRO D'ORZIO | MMILSPOM | ZOGNO | MMILZOGM | 132 | 16.458 | | 45.8741 | 9.65302 | 45.7937 | 9.6662 |

Table 3.3: Table after data matching.

To quantify the efficacy of the proposed data integration pipeline, a statistical evaluation of the spatial matching results was conducted, as summarized in Table 3.1. Out of a total of 3,973 topological nodes (stations) extracted from the official PDF registers, the primary entity resolution phase successfully matched 2,131 stations using standard alphabetical user codes (codice_utente).

| Matching Category | Number of Stations | Percentage (%) |
|---|---|---|
| Total Extracted Stations | 3,973 | 100.00% |
| Matched via Alpha-code (Primary) | 2,131 | 53.64% |
| Matched via Numeric-code (Fallback) | 354 | 8.91% |
| Unmapped / No-code | 1,047 | 26.35% |
| Unclassified | 441 | 11.10% |

Table 3.4: Statistical summary of the spatial matching results.

Crucially, the implementation of the fallback dictionary mechanism salvaged an additional 354 stations (8.91%) by mapping them to numeric substation codes. Consequently, the combined automated pipeline achieved a successful resolution rate of over 62.5%.

# 4 | Geospatial Mapping and Visualization

## 4.1. System Architecture and Technology Stack

In modern power system reliability research, traditional purely probabilistic statistical models have gradually revealed their limitations. Because extreme weather events (such as lightning strikes and blizzards) and natural disasters often exhibit highly localized spatial characteristics, integrating Geographic Information Systems (GIS) to spatially map grid infrastructure has become a critical approach for assessing the vulnerability of modern transmission networks [18]. After completing the logical data cleaning and foreign key matching, this chapter focuses on resolving the core spatial dimension issues: how to acquire missing physical coordinates and transform the vast grid topological relationships into an intuitive and analytically capable interactive spatial map.

### 4.1.1. System Architecture and Technology Stack

The spatial mapping and visualization pipeline of this system is built on the Python ecosystem, aiming to achieve a seamless connection from low-level data storage to high-level interactive decision-making. The architecture is primarily divided into three functional modules:

**Data Interface Layer:** Utilizes `mysql.connector` to establish a real-time connection with the relational database, dynamically extracting the matched line topologies and node data.

**Spatial Computing Layer:** Employs the `geopy` library to call external map APIs for geocoding; utilizes the `geopandas` and `shapely` libraries to load and process official high-precision administrative boundary polygon data, executing complex spatial geometric operations and topological validation.

**Map Rendering Front-end:** Abandoning traditional static plotting tools, this research adopts the `folium` library and its underlying `Leaflet.js` engine. This technology stack can directly compile spatial topological objects in Python's memory into Web pages containing dynamic interactive scripts, providing lightweight visualization support for subsequent complex network spatial analyses [19].

## 4.2. Georeferencing the Transmission Infrastructure

Although the preliminary database integrated a massive amount of site information, some historical substations still faced issues of insufficient precision or missing geographic coordinates. To ensure that all topological nodes could be accurately located on the map, this study developed an automated georeferencing engine.

### 4.2.1. Dual-Engine Geocoding Mechanism

For sites with only toponymic descriptions but no specific latitude and longitude coordinates, the system introduces an automated geocoder. Since open-source crowdsourced maps (such as OpenStreetMap) may have coverage gaps regarding remote electrical facilities, this research designed a dual-engine fallback mechanism to balance efficiency and parsing success rate [20]. First, the system initiates a primary query request to the open-source map engine (Nominatim). If this engine fails to resolve a specific location name or triggers a network timeout exception, the system automatically downgrades and triggers a secondary, precise search using a backup commercial map engine (Google Maps V3 API). This complementary strategy significantly reduces the coordinate omission rate.

### 4.2.2. Spatial Boundary Validation via Polygons

After obtaining coordinates through external APIs, to prevent external engines from returning false geographic locations due to toponymic ambiguity (e.g., transnational towns with the same name), the system introduces a rigorous spatial boundary validation step. The algorithm loads official administrative boundary data and utilizes the `shapely` library to convert the latitude and longitude returned by the APIs into geometric point objects. Subsequently, the system executes a strict "Point-in-Polygon" topological check. Only when the coordinate is mathematically verified to fall strictly within the geometric boundaries of its respective operational area (e.g., Piedmont, Lombardy) is it officially written into the database, thereby ensuring the absolute reliability of the spatial data.

## 4.3. Development of the Interactive Web Map

After ensuring all nodes possess reliable coordinates, the visualization module takes over the construction of the map. Breaking through the limitations of traditional static maps, this module generates a spatial analytics dashboard that supports multi-layer data analysis and rich-text interactions.


Figure 4.1: A complete map of Italy's power grid

### 4.3.1. Topology Coloring and Layer Isolation

To clearly display the grid hierarchy, the algorithm reads the operating voltage parameters of the lines and applies strict color coding when generating spatial polyline vectors. For example, the 380 kV extra-high voltage backbone network and the 150 kV sub-transmission network are visually assigned high-contrast distinguishing colors.

Furthermore, given that displaying the national grid data simultaneously would cause severe visual clutter, the system utilizes layer control components to strictly divide the national infrastructure into 8 independent operational areas. Users can toggle specific feature groups to isolate and focus on analyzing the spatial vulnerability distribution of a particular region, greatly improving the efficiency of macroscopic dispatch analysis.


Figure 4.2: Transmission Lines of Roma Area

## 4.3.2. Spatial Overlap Handling and Data Aggregation

At the macro-map level, multiple substation entities located within the same municipality often share identical rough coordinates, causing markers on the map to overlap and obscure one another. To resolve this issue, the system developed a location grouping logic. Before rendering, the algorithm iterates through the coordinate dictionary and packages facilities with the same latitude and longitude at the data structure level. On the map surface, these overlapping entities are aggregated into a single visual anchor, effectively reducing the redundancy of map elements.

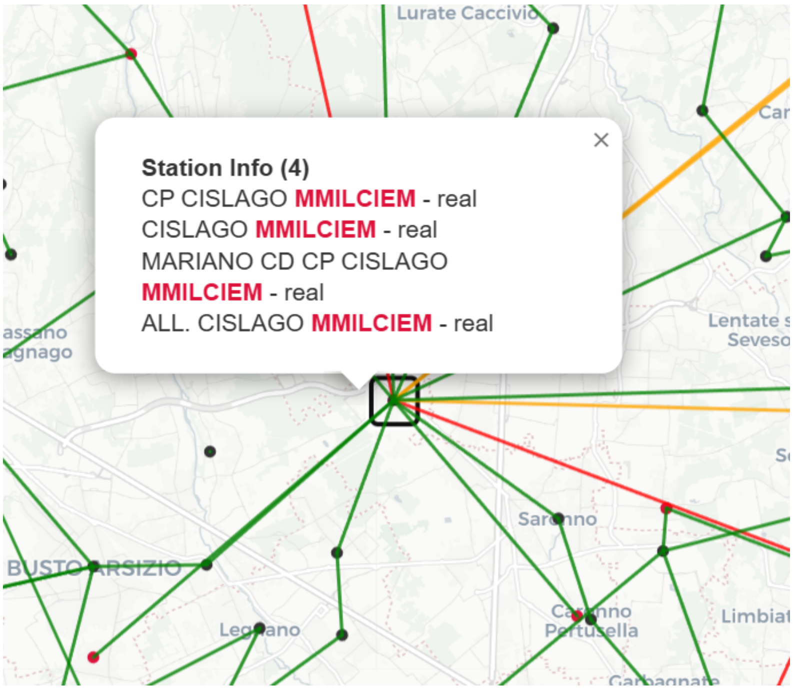


Figure 4.3: Multiple sites share the same coordinates

### 4.3.3. Data Binding and Dynamic Retrieval

The core value of the interactive map lies in its function as a spatial query interface for the underlying database. When rendering transmission line and substation primitives, the algorithm directly binds core physical parameters (such as line length, precise operating voltage, origin and destination node codes) to the spatial entities by generating customized HTML popups. When a user clicks on a specific grid corridor or site on the map, the system dynamically retrieves and displays these encapsulated low-level data in real time. This "what you see is what you get" data retrieval mechanism transforms the map from a mere display tool into a spatial decision support system, assisting grid managers in rapidly locating failure nodes and querying line topological parameters.

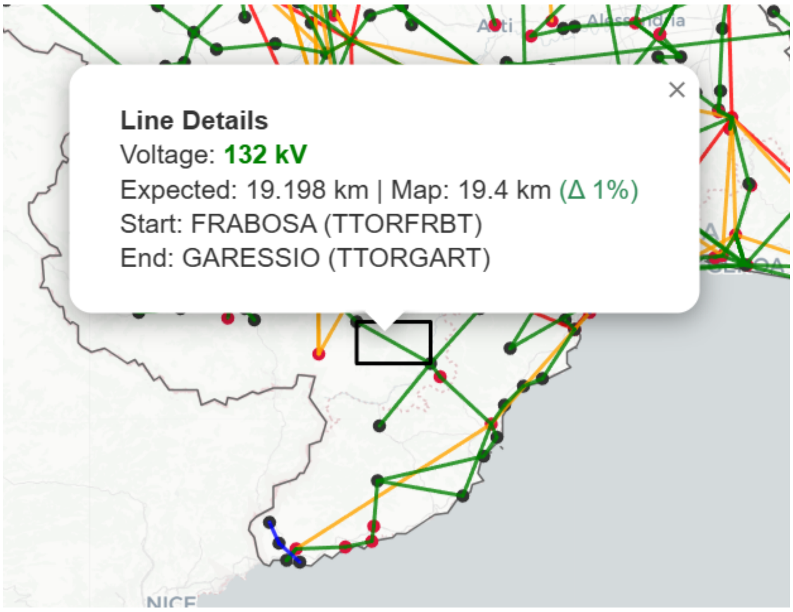


Figure 4.4: Line Details

## 4.4. Mapping Results

To comprehensively evaluate the final efficacy of the dual-engine geocoding mechanism and the Web map rendering pipeline, this study conducted a quantitative statistical analysis of the spatial mapping success rate (detailed in Table 4.1).

Through external APIs and topological inference, the system successfully inferred and supplemented the spatial coordinates of 1,405 substations, accounting for 35.36% of the total valid sites. At the transmission corridor level, the official register contains a total of 4,264 physical lines. After rigorous foreign key matching and coordinate mapping, 3,757 lines were successfully instantiated and rendered onto the interactive map, achieving an overall spatial utilization rate of 88.11%. This exceptionally high coverage rate proves that the constructed visualization system can highly accurately reconstruct the macroscopic overview of the Italian national transmission grid.

| **Mapping & Validation Metrics** | **Count** | **Percentage (%)** |
|---|---|---|
| Total Official Transmission Lines | 4,264 | 100.00% |
| Lines Successfully Mapped | 3,757 | 88.11% |
| Inferred Stations | 1,405 | 35.36% |
| Lines with Error ≤ 100% | 3,213 | 75.35% |

Table 4.1: Table 4.1: Statistical summary of transmission line spatial mapping and validation results.

Furthermore, this study executed a rigorous physical validation on all mapped lines (as previously described, comparing the geodesic straight-line distance between the two topological endpoints with the official cable length). Statistics indicate that a total of 3,213 lines (accounting for 75.35% of the total lines) have an error rate of less than or equal to 100% (i.e., the straight-line distance falls within the reasonable physical limits of the actual cable length). These high-fidelity line data, having passed strict spatial logic tests, provide highly reliable data support for conducting the single-line-based reliability assessment in the next chapter.

# 5 | Reliability Analysis and Results

## 5.1. Evaluation Methodology Based on Operational Areas

The reliability analysis shifts the research focus from the macroscopic national grid to the eight operational areas defined by Terna. These areas serve as independent research units to evaluate the spatiotemporal evolution of key reliability metrics, including the Severity Factor, failure rate, repair rate, and total undelivered energy.

The geographic size and infrastructure scale of these eight areas vary significantly. Comparing absolute values, such as the total number of failures or total undelivered energy, inherently introduces statistical bias favoring larger or heavily industrialized regions. To eliminate this scale-induced discrepancy, a length-normalized approach is applied across all evaluated metrics.

The length parameter $L$ utilized in the normalization calculations represents the actual physical cable length extracted directly from the official facility lists. The lengths of the mapped transmission lines are aggregated by their respective operational areas to establish a precise infrastructure baseline. Utilizing this physical length baseline ensures that the calculated metrics objectively reflect the actual vulnerability per kilometer of transmission line within each distinct region.

## 5.2. Severity Factor

### 5.2.1. Number of Failures

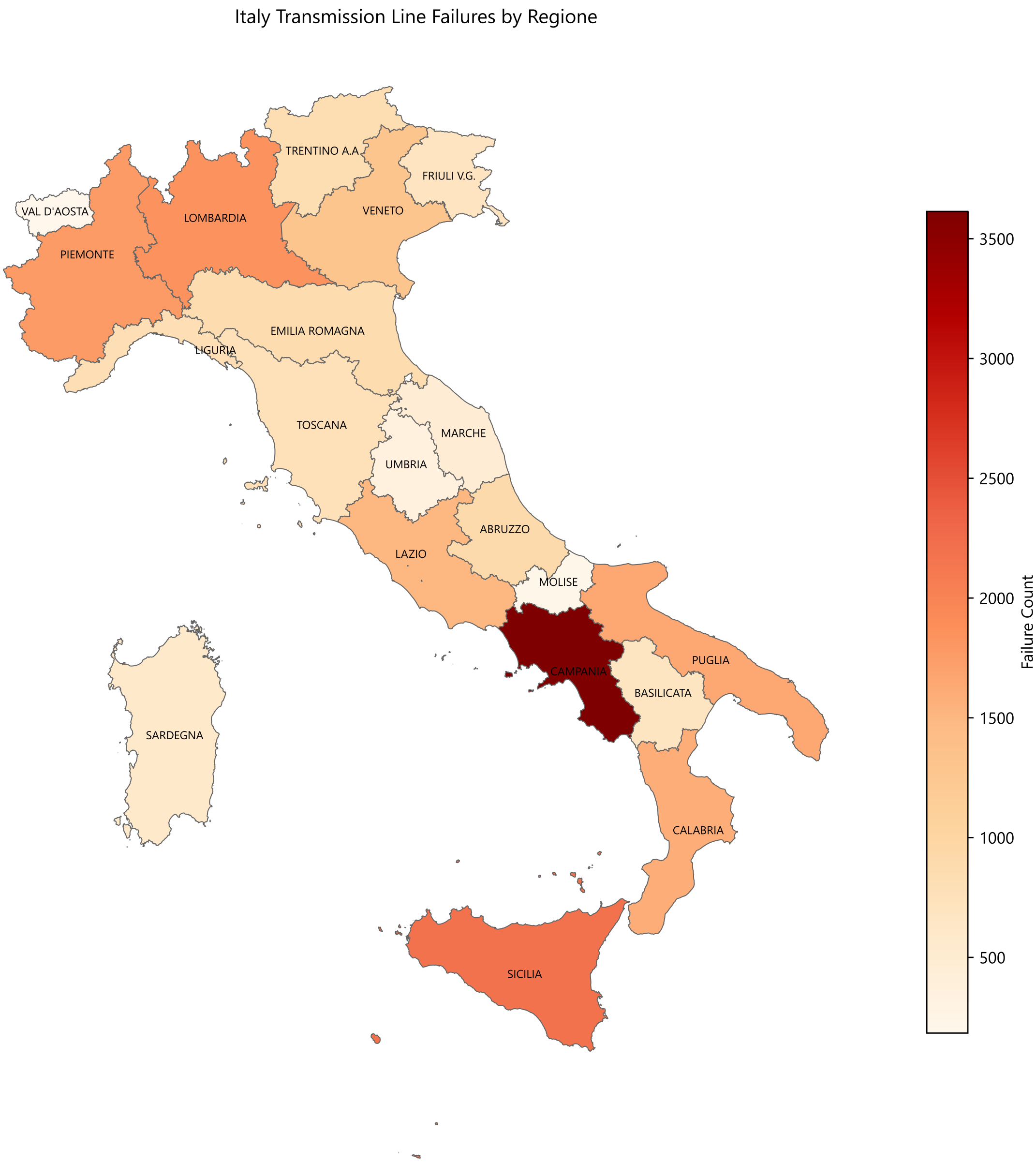


Figure 5.1: Heatmap of failures by region

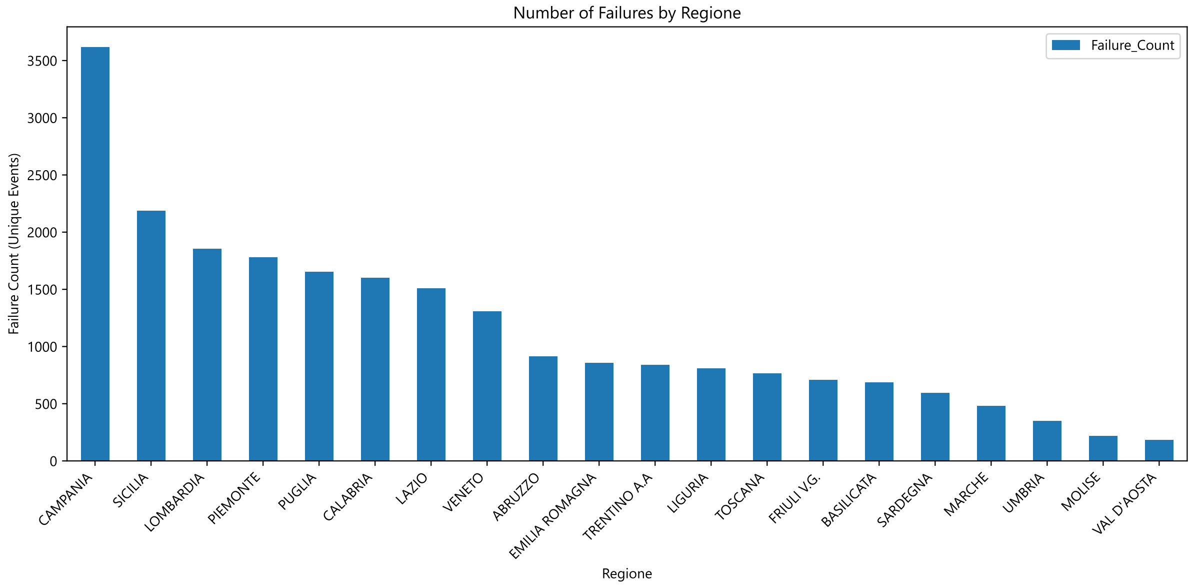


Figure 5.2: Number of failures by region

From Figure 5.1 and 5.2, it can be seen that Campania's absolute number of failure events is far ahead, exceeding 3,500, followed closely by Sicilia and Lombardia. These figures show that the southern regions indeed exhibit typical characteristics of high-frequency failures regarding the exposure of physical lines.

### 5.2.2. Unnormalized SF

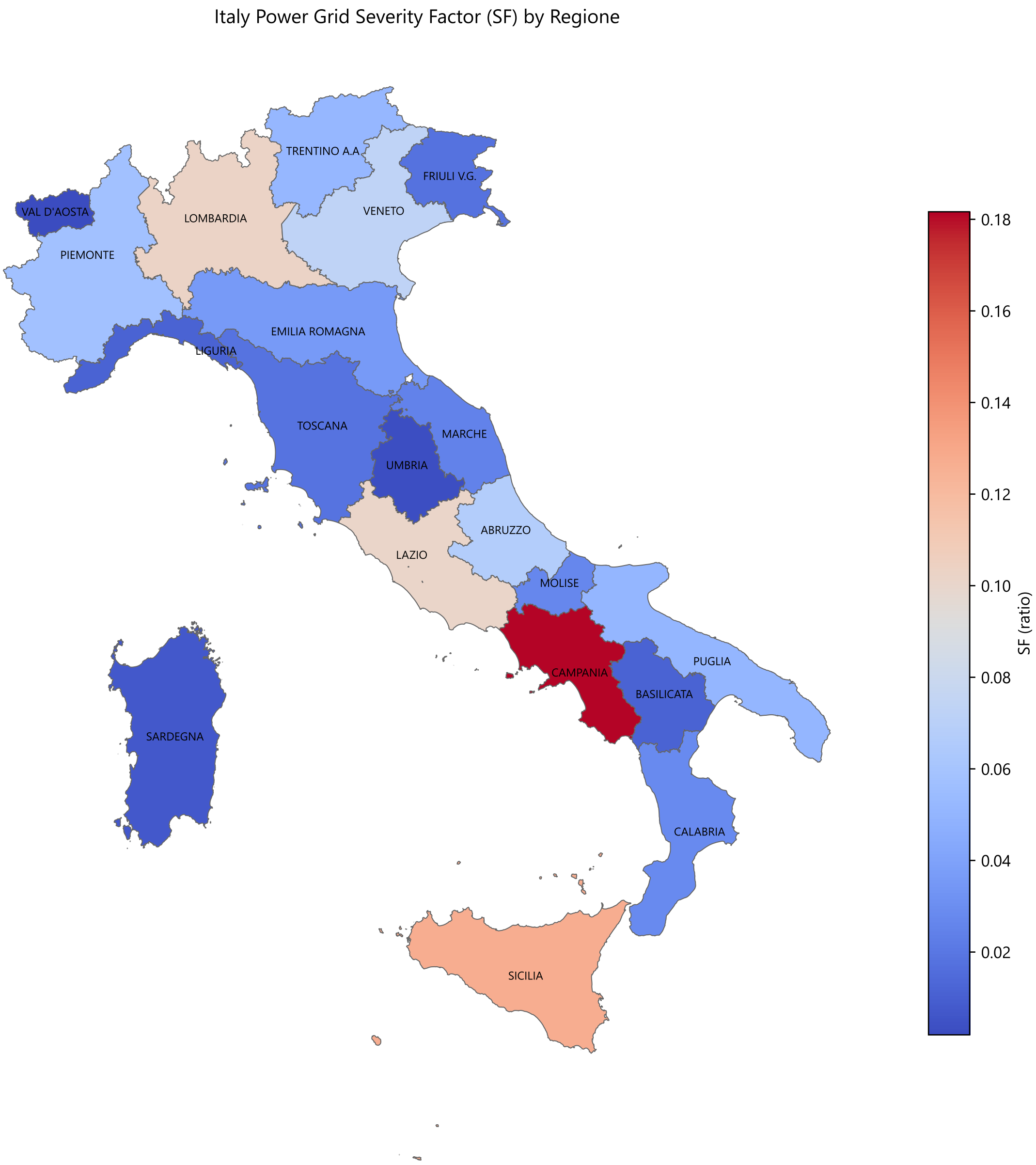


Figure 5.3: Heatmap of severity factor by region

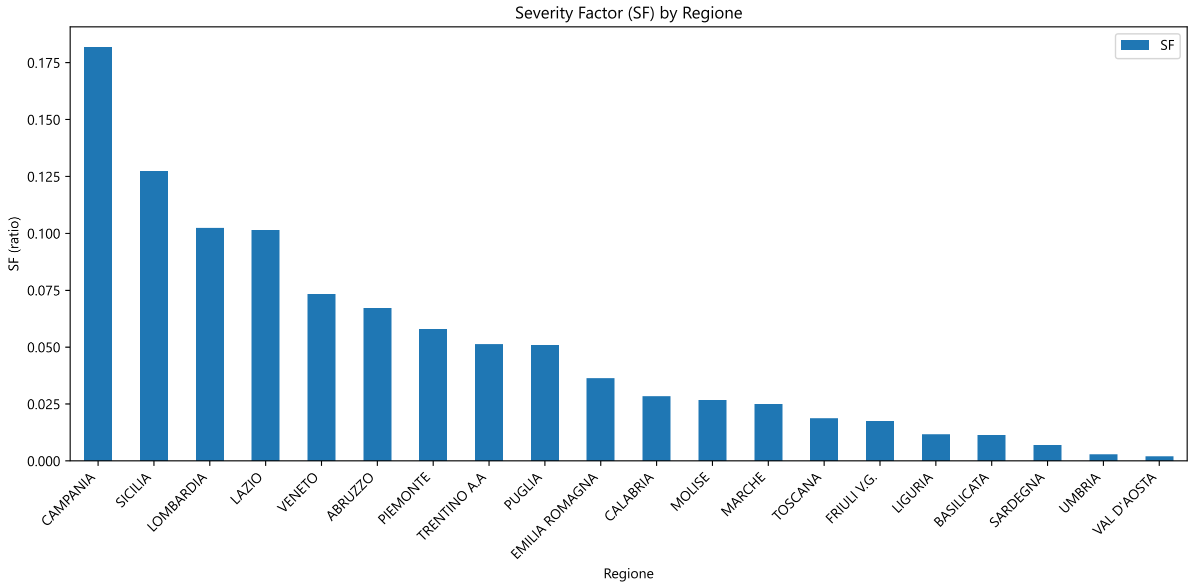


Figure 5.4: Severity factor by region

From Figure 5.3 and 5.4, if the physical network size is not considered, Campania, Sicilia, and Lombardia rank in the top three in terms of numerical value, with Campania approaching 0.175. This result is consistent with the results obtained from the Number of Failures analysis above, demonstrating that Unnormalized SF is numerically highly susceptible to the influence of the absolute number of failures.

### 5.2.3. Normalized SF

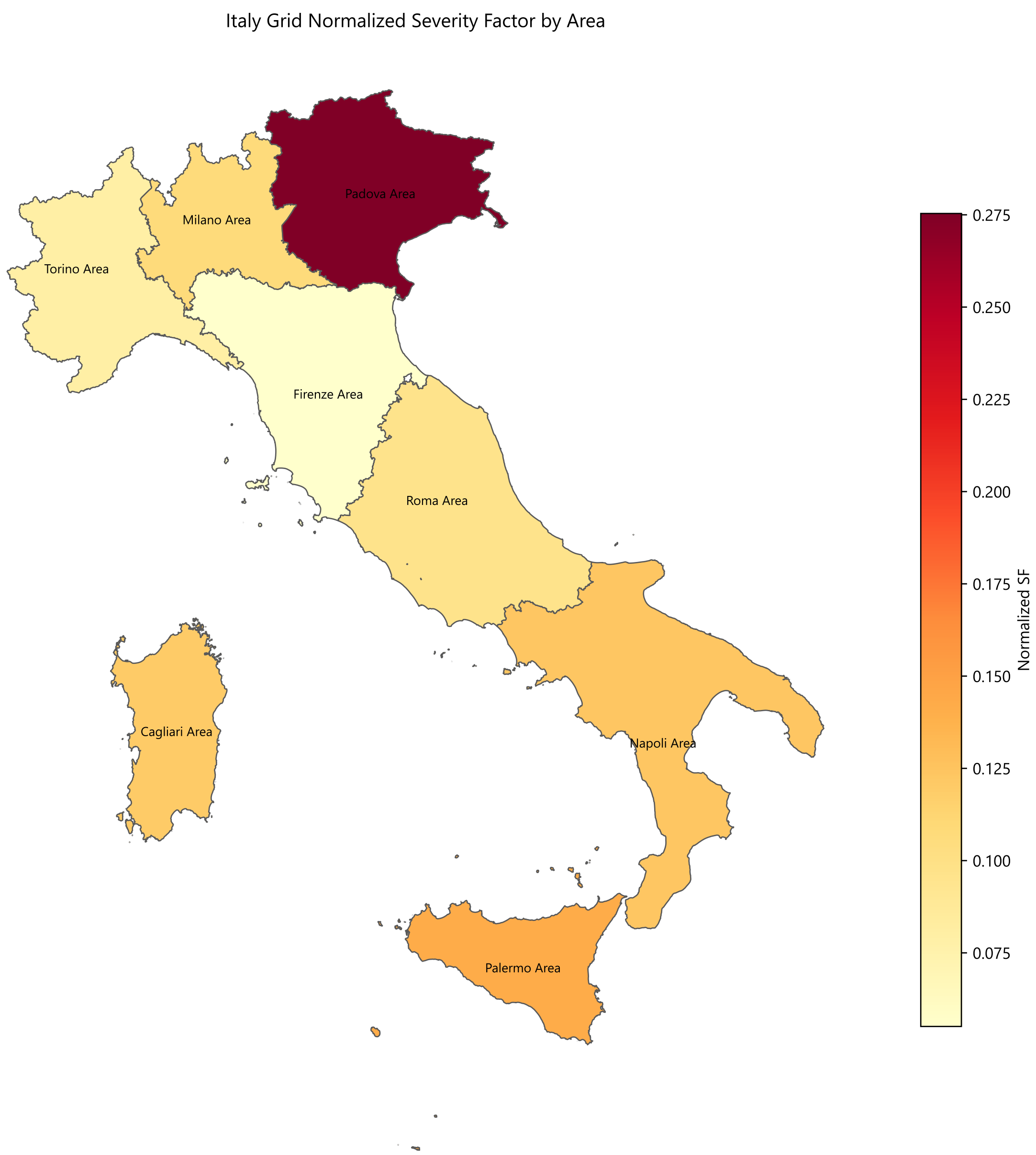


Figure 5.5: Heatmap of normalized severity factor by operational area

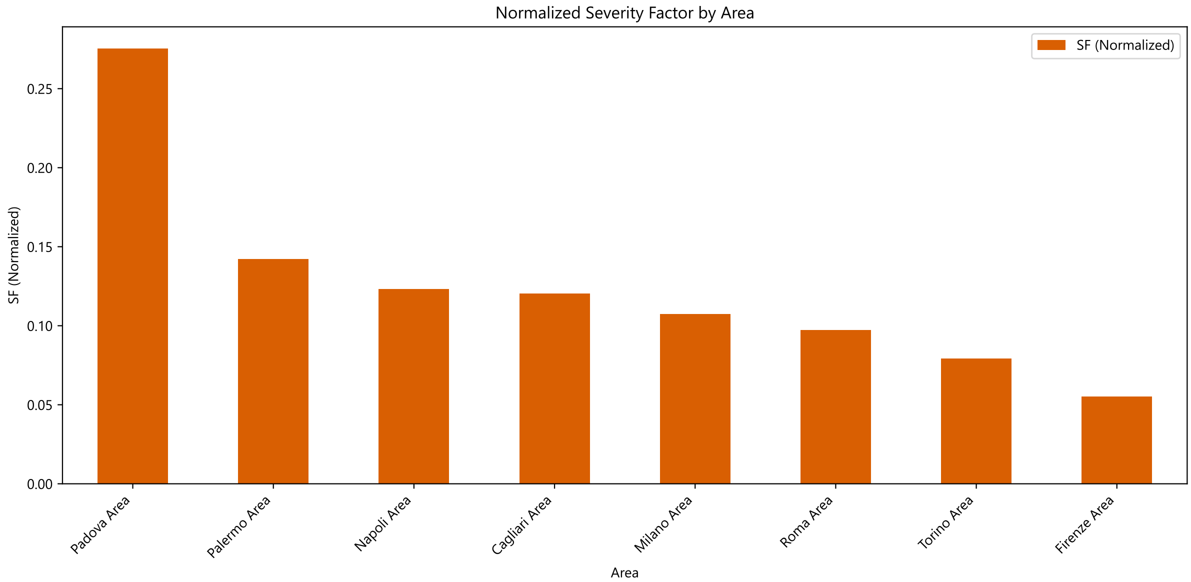


Figure 5.6: Normalized severity factor by operational area

### 5.2.4. Interpretation of Normalized SF Rankings

Neglecting the scale of the grid infrastructure and relying solely on absolute failure counts or unnormalized Severity Factor misleadingly identifies Campania, Sicilia, and Lombardia as the most vulnerable regions. This phenomenon is typically attributed to the frequency of extreme weather events in the southern regions and the massive consumer base in the Lombardia area. However, the introduction of the double-normalized Severity Factor, which eliminates biases introduced by transmission line length and failure frequency, triggers a radical inversion in the regional rankings.

From Figure 5.5 and 5.6, it shows the normalized SF of the Padova area surges to an anomalous peak exceeding 0.275, significantly outpacing the second-ranked Palermo area. Despite the Padova area not ranking among the top three nationwide in absolute failure counts, its grid topology is densely coupled with high-energy-consuming heavy industries, such as steel metallurgy. Consequently, a fault per unit kilometer of transmission line inflicts disproportionately devastating consequences in terms of ENS. This topological reality proves that the Padova area constitutes a quintessential High-Impact, Low-Frequency (HILP) vulnerable core within the Italian national grid.

## 5.3. RAMS Analysis

### 5.3.1. Failure Rate Trends

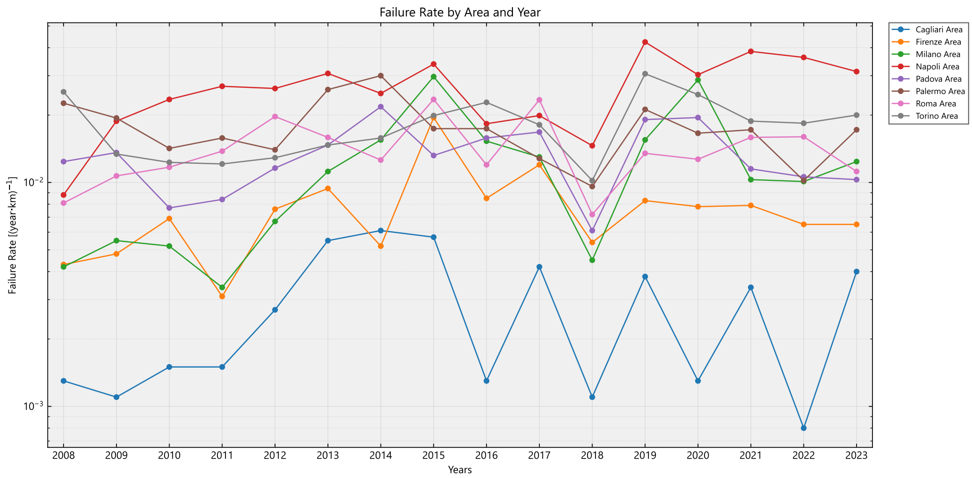


Figure 5.7: Failure rate trends by operational area

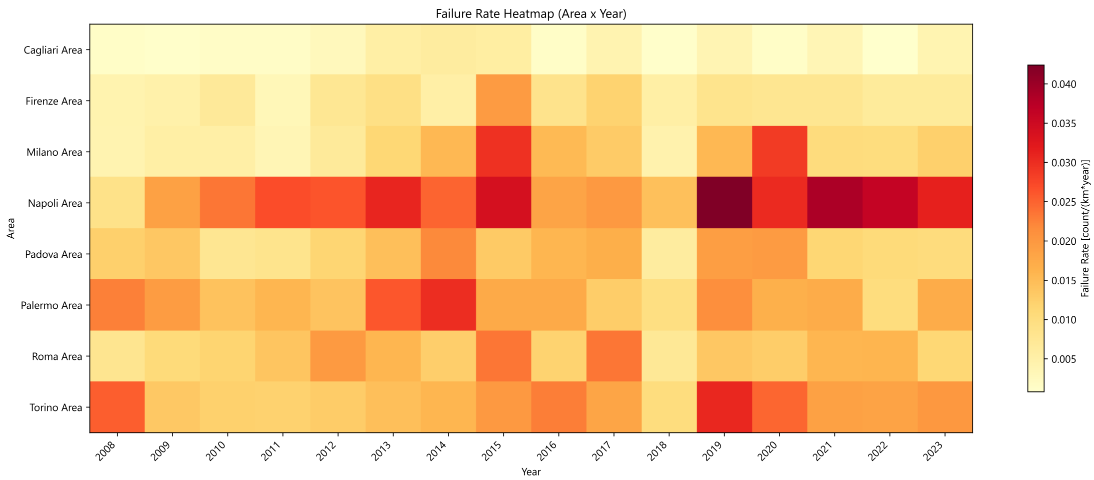


Figure 5.8: Failure rate heatmap by operational area and year

Figure 5.7 and 5.8 clearly show that the Napoli and Palermo areas remain in a persistent high-failure-rate band, with a marked deepening in recent years (e.g., 2019–2020). This pattern is consistent with the southern grid's increased exposure to climate change and extreme weather, which drives more frequent faults.

### 5.3.2. Repair Rate Trends

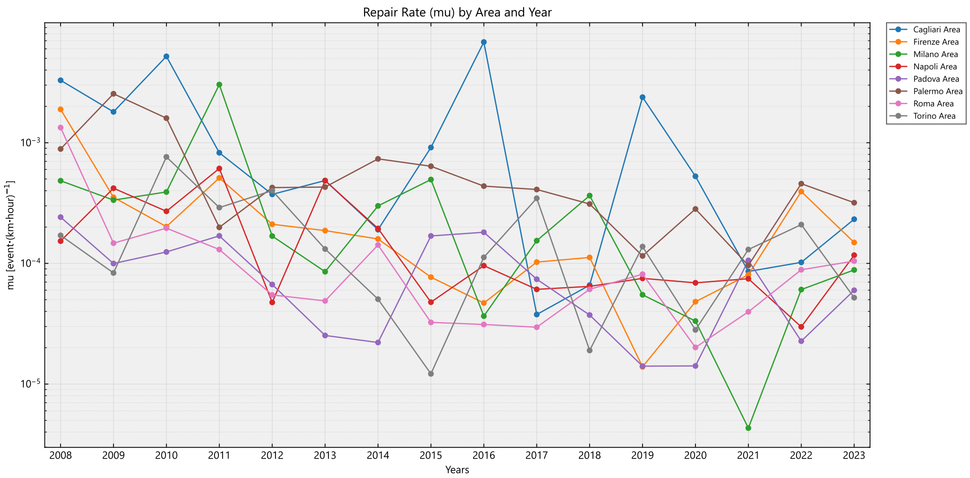


Figure 5.9: Repair rate trends by operational area

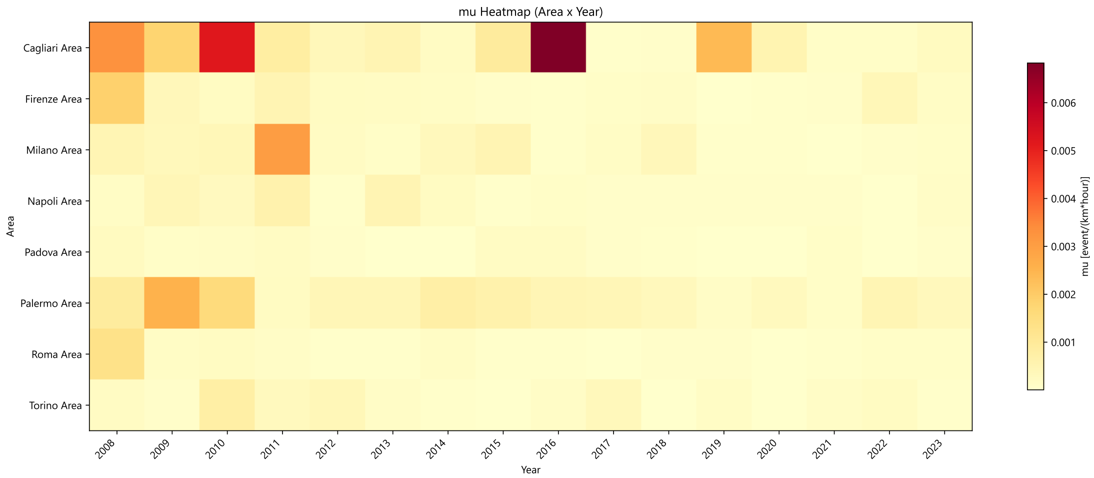


Figure 5.10: Repair rate heatmap by operational area and year

The repair rate, a key indicator of maintainability, exhibits dramatic temporal fluctuations.From Figure 5.9 and 5.10, we see thst some regions experienced precipitous drops in repair speed in specific years, such as the Milan area in 2021 and the Torino area in 2015. This abnormal decrease in maintainability indicates that the power grid has encountered complex cascading faults or severe physical environmental obstacles, directly disrupting the grid's rapid repair and restoration mechanisms.

### 5.3.3. Availability and Undelivered Energy

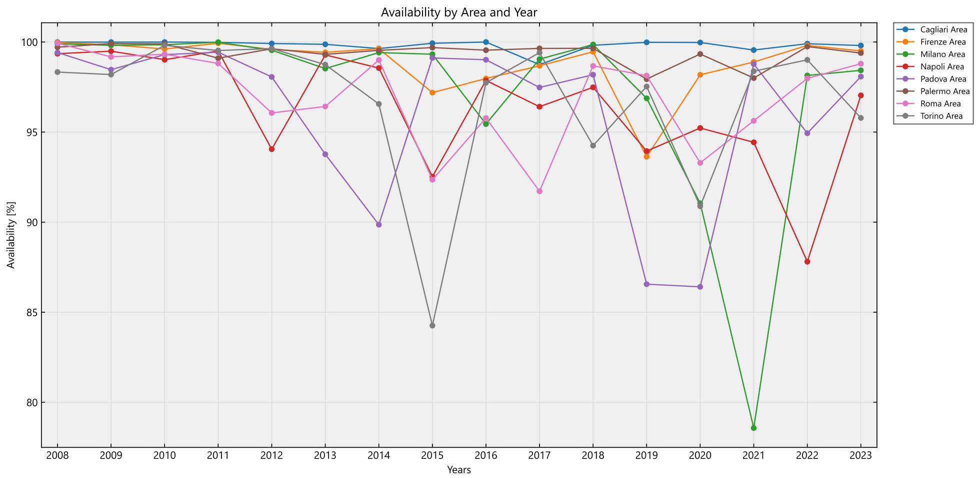


Figure 5.11: Availability trends by operational area

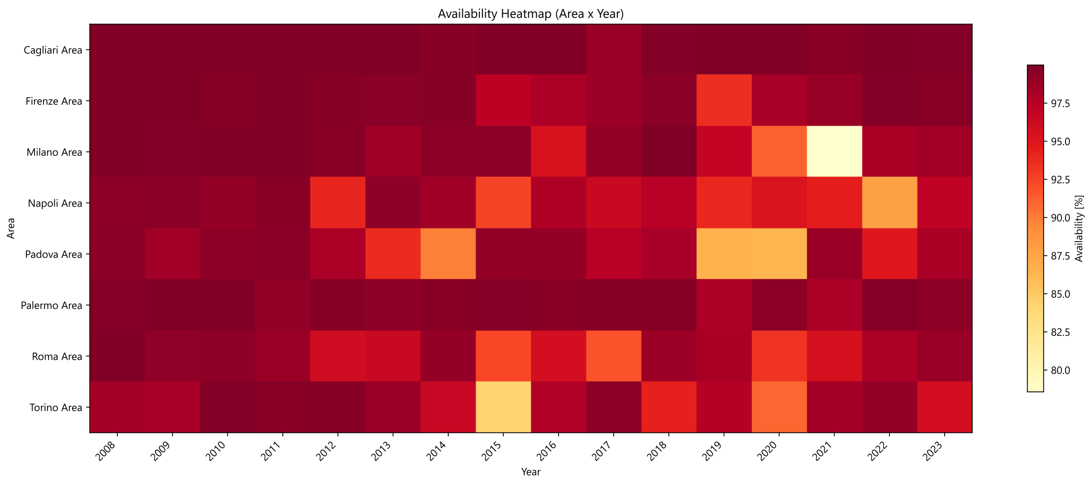


Figure 5.12: Availability heatmap by operational area and year

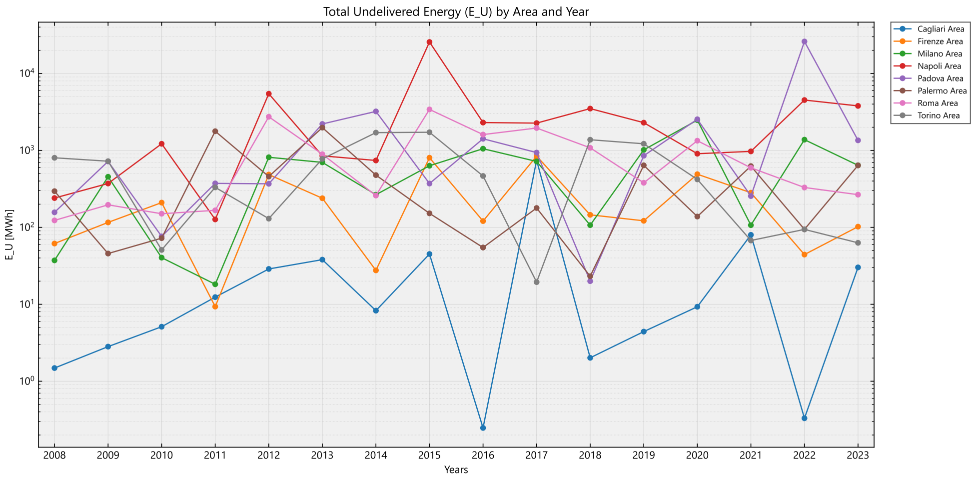


Figure 5.13: Undelivered energy $E_U$ trends by operational area

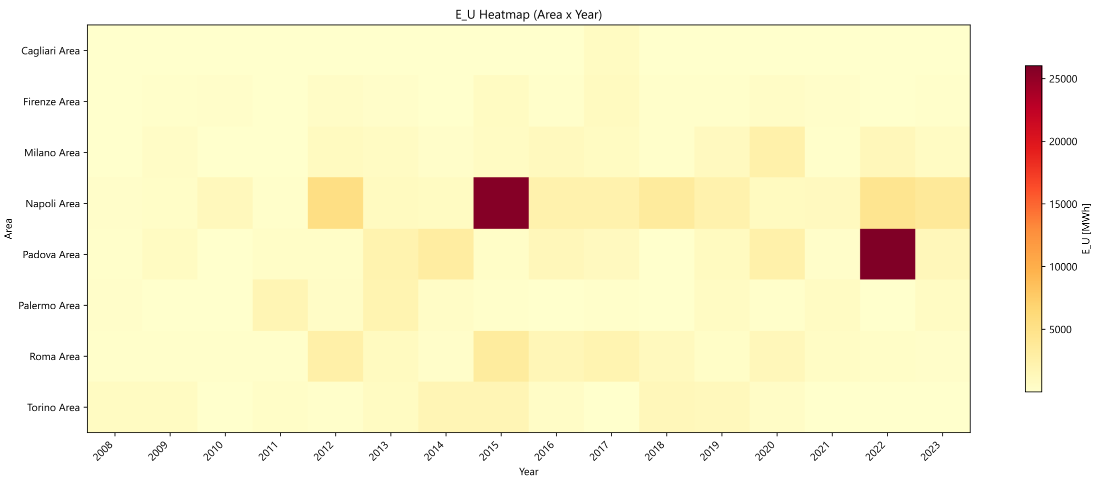


Figure 5.14: Undelivered energy $E_U$ heatmap by operational area and year

The spatiotemporal distribution of system availability and Total Undelivered Energy ($E_U$) reveals extreme "black swan" events during grid operation, directly substantiating the highly anomalous Severity Factor observed in the Padova area.From Figure 5.11 and 5.12, we noticed that under standard operational conditions, the baseline availability of the national transmission network is consistently sustained above 97.5%. However, the Milano area experienced a precipitous collapse in availability to approximately 78% in 2021, forming a distinct low-value depression on the heatmap that perfectly correlates with its drastically reduced maintainability metrics in the same year. The availability in the Padova area also demonstrated significant degradation in 2015, 2019, and 2020, plunging to the 85% threshold multiple times, which reflects the structural vulnerability of the regional grid resilience. The Total Undelivered Energy heatmap (Figure5.13) exposes two catastrophic energy loss anomalies against a 15-year predominantly stable baseline. The Napoli area exhibited a severe surge in unsupplied energy in 2015, a phenomenon primarily driven by its massive base of high-frequency fault occurrences. More exceptionally, the Padova area registered an absolute peak block in 2022, with the single-year energy loss exceeding $2.5 \times 10^4$ MWh (Figure5.14). This isolated and highly destructive spatial anomaly unequivocally validates the topological reality of the Padova area functioning as a HILP vulnerable core. The extreme unsupplied energy shock not only dictates the regional Severity Factor but also exposes the systemic crisis of heavy industrial load nodes under cascading failures.

# 6 | Conclusions

## 6.1. Summary of Contributions

This chapter summarizes the research contributions of the entire thesis, frankly analyzes the limitations present in the data integration and spatial mapping processes, and outlines clear development directions for the in-depth stability analysis of the RTN based on the constructed geographic topology database.

This thesis successfully bridged the gap between logical reliability records and physical spatial infrastructure, providing a novel geographic perspective for the vulnerability assessment of the RTN. The main contributions of this research are summarized as follows:

1. **Data Integration and Spatial Mapping Strategy:** To overcome the data silos between unstructured asset registers and relational outage databases, this study designed an automated extraction and matching pipeline. By employing the Camelot library for PDF parsing, combined with text normalization, two-pass fuzzy matching, and a substation fallback dictionary mechanism, the system effectively resolved entity inconsistencies. Ultimately, 88.11% of the official transmission lines were successfully mapped and instantiated from the historical records.

2. **Development of a Geospatial Visualization Framework:** This research constructed an interactive visual analytics platform using a dual-engine geocoding mechanism (Nominatim and Google Maps API) along with strict polygon-based spatial boundary validation. The system successfully transformed abstract database tables into intuitive geographic visualizations, implementing logic to handle spatial overlaps of user sites within the same municipality and enabling dynamic data binding for topological parameters.

3. **Identification of Vulnerabilities via Length-Normalized Metrics:** By extracting actual physical lengths from the mapped infrastructure, this study innovated traditional statistical evaluation methods by computing a length-normalized Severity Factor. This spatial method removed the bias caused by different scales

and clearly showed the “Padova Anomaly”, demonstrating that the Padova area, due to its compact topology and heavy industrial load, bears the most critical economic risk, contrary to the illusion created by absolute failure numbers in southern regions. Furthermore, the spatiotemporal RAMS analysis accurately pinpointed extreme operational anomalies, such as the persistent high failure rates in the Napoli and Palermo areas, and the precipitous collapse in grid availability in the Milano area in 2021.

## 6.2. Limitations

Despite the successful implementation of the spatial mapping pipeline and visualization framework, the current methodology presents several limitations primarily stemming from raw data quality and geocoding constraints:

1. **Entity Resolution Constraints and Unmapped Data:** The historical data published by Terna frequently exhibits irregular naming conventions and missing standard alphabetical user codes (codice_utente). Although the implemented fallback dictionary mechanism significantly improved the matching rate by utilizing numeric substation codes, 26.35% of the extracted stations remained unmapped. For these unmapped or unclassified sites, it is currently not possible to directly exploit the guasti table for physical reliability analysis.

2. **Geocoding Granularity and Spatial Overlaps:** The spatial enrichment process relied heavily on the municipality and province data provided in Terna’s official documents, as precise latitude and longitude coordinates for each facility were inherently unavailable. Consequently, the geocoding engines assigned identical spatial coordinates to distinct user sites located within the same municipality. While the visualization module mitigated visual clutter through data aggregation, this coarse spatial granularity prevents micro-level topological analysis within dense urban centers.

3. **Geometric Validation Discrepancies:** The physical validation comparing actual cable lengths to geodesic straight-line distances revealed that while 75.35% of the mapped lines fell within a reasonable error margin, the remaining lines exhibited significant deviations. Straight-line approximations calculated via APIs cannot fully account for actual topographic routing, mountain detours, or geometric constraints in the physical transmission corridors, inevitably introducing spatial distortion into the visualized network topology.

## 6.3. Future Developments

Having successfully established the underlying data pipeline and achieved macro-regional severity assessments, the fully integrated and visualized topological relationships between transmission lines and substations pave the way for advanced reliability analyses. Future research should focus on the following directions:

1. **Cascading Failures and Topological Propagation:** Future studies will transcend static severity statistics to observe the propagation dynamics of blackout events within the grid network. By integrating graph theory and complex network analysis, research can investigate how the failure of a single critical node triggers the overloading of adjacent lines, thereby simulating and predicting the propagation paths of cascading failures that lead to large-scale blackouts.

2. **Identifying the Most Dangerous Line and Make It More General:** The current assessment primarily operates at the macro operational area (Area Territoriale) level. After further resolving API coordinate precision and manual calibration issues, future research should refine the spatial granularity, utilizing a single transmission line as the smallest unit of study. By correlating historical weather data with the physical attributes of specific lines, the most vulnerable and high-risk individual lines within the RTN can be precisely identified, providing grid operators with forward-looking recommendations for micro-topological upgrades and preventive maintenance.

# Bibliography

[1] R. N. Allan and Billinton, *Reliability Evaluation of Power Systems*. Springer Science & Business Media, Nov. 2013, google-Books-ID: 4QryBwAAQBAJ.

[2] L. Wu, G. Kaiser, C. Rudin, and R. Anderson, "Data quality assurance and performance measurement of data mining for preventive maintenance of power grid," in *Proceedings of the First International Workshop on Data Mining for Service and Maintenance*, ser. KDD4Service '11. New York, NY, USA: Association for Computing Machinery, 2011, pp. 28–32. [Online]. Available: https://dl.acm.org/doi/10.1145/2018673.2018679

[3] L. Martiri, A. Moschetti, C. Laurano, L. Cristaldi, S. Toscani, and M. Faifer, "Analyzing Failure Severity and Trends in Italian Power Grid Transmission Lines," in *2025 International Conference on Clean Electrical Power (ICCEP)*, Jun. 2025, pp. 996–1001, iSSN: 2474-9664. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11143742

[4] A. Moschetti, L. Martiri, C. Laurano, L. Cristaldi, D. Martinenghi, and M. Faifer, "A Data Preparation Pipeline for the Analysis of the Italian Power Grid Reliability," in *2025 International Conference on Clean Electrical Power (ICCEP)*, Jun. 2025, pp. 1002–1007, iSSN: 2474-9664. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11143620

[5] Terna S.p.A., "2024 Annual Report," Terna S.p.A., Rome, Italy, Tech. Rep., Mar. 2025. [Online]. Available: https://www.terna.it

[6] ——, "Elenco Impianti RTN al 2013," Terna S.p.A., Rome, Italy, Tech. Rep., Jan. 2014.

[7] ——, "Guida Tecnica: Classificazione e Registrazione delle Interruzioni degli Utenti Direttamente e Indirettamente Connessi alla RTN," Terna S.p.A., Rome, Italy, Tech. Rep. Allegato A54, Rev. 09, Sep. 2017.

[8] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the North

American power grid," *Physical Review E*, vol. 69, no. 2, p. 025103, Feb. 2004. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.69.025103

[9] E. Bompard, R. Napoli, and F. Xue, "Analysis of structural vulnerabilities in power transmission grids," *International Journal of Critical Infrastructure Protection*, vol. 2, no. 1, pp. 5–12, May 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874548209000031

[10] L. Cristaldi, M. Faifer, C. Laurano, R. Ottoboni, S. Toscani, and M. Zanoni, "Severity Factor and Reliability Trend Evaluation of the Italian Transmission Lines," in *2023 International Conference on Clean Electrical Power (ICCEP)*, Jun. 2023, pp. 671–677, iSSN: 2474-9664. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10247465

[11] S. Garip, Özdemir, and N. Altın, "Power System Reliability Assessment— A Review on Analysis and Evaluation Methods," *Journal of Energy Systems*, vol. 6, no. 3, pp. 401–419, Sep. 2022. [Online]. Available: https://dergipark.org.tr/en/pub/jes/article/1099618

[12] Y. Dechgummarn, P. Fuangfoo, and W. Kampeerawat, "Reliability Assessment and Improvement of Electrical Distribution Systems by Using Multinomial Monte Carlo Simulations and a Component Risk Priority Index," *IEEE Access*, vol. 10, pp. 111 923–111 935, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9925206

[13] V. Pérez and C. Aybar, "Challenges in Geocoding: An Analysis of R Packages and Web Scraping Approaches," *ISPRS International Journal of Geo-Information*, vol. 13, no. 6, p. 170, Jun. 2024. [Online]. Available: https://www.mdpi.com/2220-9964/13/6/170

[14] G. F. B. d. Medeiros, L. C. Degrossi, and M. Holanda, "QualiOSM: An Architecture to Improve Data Completeness on OpenStreetMap," *Journal of Information and Data Management*, vol. 12, no. 2, Sep. 2021. [Online]. Available: https://journals-sol.sbc.org.br/index.php/jidm/article/view/2092

[15] P. Christen, "The Data Matching Process," in *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*, P. Christen, Ed. Berlin, Heidelberg: Springer, 2012, pp. 23–35. [Online]. Available: https://doi.org/10.1007/978-3-642-31164-2_2

[16] W. McKinney, "Data Structures for Statistical Computing in Python," *SciPy*

*2010*, May 2010. [Online]. Available: https://proceedings.scipy.org/articles/Majora-92bf1922-00a

[17] L. Boytsov, "Indexing methods for approximate dictionary searching: Comparative analysis," *ACM J. Exp. Algorithmics*, vol. 16, pp. 1.1:1.1–1.1:1.91, 2011. [Online]. Available: https://dl.acm.org/doi/10.1145/1963190.1963191

[18] M. Panteli and P. Mancarella, "Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies," *Electric Power Systems Research*, vol. 127, pp. 259–270, Oct. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037877961500187X

[19] C. Tominski and H. Schumann, *Interactive Visual Data Analysis.* CRC Press, Apr. 2020, google-Books-ID: 2qXaDwAAQBAJ.

[20] P. Mooney, P. Corcoran, and A. C. Winstanley, "Towards quality metrics for OpenStreetMap," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 514–517. [Online]. Available: https://dl.acm.org/doi/10.1145/1869790.1869875

# List of Figures

# List of Tables

# List of Symbols

| Variable | Description | SI unit |
|---|---|---|
| $\lambda$ | failure rate | 1/(km·year) |
| $N_i$ | number of failures in category $i$ | – |
| $L_i$ | line length in category $i$ | km |
| $T$ | observation period | year |
| $\mu$ | repair rate | 1/h |
| $\mathrm{MTTR}_i$ | mean time to repair for line $i$ | h |
| $t_n$ | outage duration of event $n$ | h |
| $A$ | availability | % |
| $\mathrm{MTTF}_i$ | mean time to failure for line $i$ | h |
| $\mathrm{MTBF}_i$ | mean time between failures for line $i$ | h |
| $E_{U,i}$ | undelivered energy in category $i$ | MWh |
| $P_i$ | interrupted power of event $i$ | MW |
| $t_i$ | outage duration of event $i$ | h |
| $\mathrm{SF}_j$ | severity factor for category $j$ | – |
| $E_{U,\mathrm{tot}}$ | total undelivered energy | MWh |
| $K_j$ | number of outages in category $j$ | – |
| $C$ | total number of outage categories | – |
| $\gamma_j$ | frequency weight for category $j$ | – |
| $L_{\mathrm{tot}}$ | total network length | km |
| $\lambda_i$ | length weight for category $i$ | – |
| $\mathrm{ENS}_i$ | energy not supplied in event $i$ | MWh |

# Acknowledgements

I would like to express my sincere gratitude to Prof. Loredana Cristaldi and Prof. Christian Laurano for their guidance on the topic selection and research direction of this thesis. I am also thankful to Luca Martiri for his valuable suggestions on refining the details of this paper. Special thanks go to Andrea Moschetti, who provided meticulous guidance from the very beginning of the thesis work and supported me throughout the entire research and writing process.