Assessing the Impact of Extreme Weather Events on Power Grid Failures: A Data-Driven Analysis in Sicily

Chiara Galvagno

Dissertation written under the supervision of Professor Pedro Afonso Nunes Correia de Costa Fernandes

Dissertation submitted in partial fulfilment of requirements for the MSc in Business Analytics at the Universidade Católica Portuguesa

May 29, 2025

Abstract

This thesis investigates the impact of extreme weather events on electric transmission networks in Sicily. As climate change increases the frequency and intensity of such events, and as decarbonization and integration of renewable energy introduce new operational complexities, the need for resilient grid infrastructure has become increasingly urgent. Sicily presents a compelling case due to its central Mediterranean location, geological and island characteristics, aging infrastructure, and growing renewable capacity, all of which increase vulnerability to weather-induced disruptions.

A three-stage modeling framework is developed to predict grid failure occurrence, severity, and economic impact using daily panel data from 2014 to 2023, combining high-resolution weather data with transmission outage reports. The first stage employs binary classification to forecast outage occurrence; the second, multiclass classification to assess severity; and the third, regression analysis to estimate unserved energy and associated economic loss. While acknowledging that outages may have multiple causes, the analysis identifies weather, especially wind, as a key driver. Results also show that interpretable models, such as logistic regression, can provide operational value for prioritization and planning.

The study contributes to climate-resilient grid planning by demonstrating the utility of supervised learning for predictive diagnostics and highlighting the importance of improved data integration in critical infrastructure monitoring.

Keywords: Extreme weather events, power grid resilience, Sicily, outage prediction, outage severity, economic impact, machine learning, supervised learning, Energy Not Supplied (ENS), panel data

Title: Assessing the Impact of Extreme Weather Events on Power Grid Failures: A Data-Driven Analysis in Sicily

Author: Chiara Galvagno

Resumo

A presente dissertação centra-se no impacto de eventos meteorológicos extremos nas redes de transmissão elétrica da Sicília, Itália. Com o aumento da frequência e intensidade desses fenómenos e com a integração de energias renováveis, é urgente a necessidade de infraestruturas resilientes. A Sicília é um caso relevante devido à localização no Mediterrâneo, características climáticas e insulares, infraestrutura envelhecida e crescente incorporação de energias renováveis na produção de energia, fatores que aumentam a vulnerabilidade a perturbações induzidas pelos fenómenos atmosféricos.

Neste contexto, foi desenvolvido um esquema de modelação em três etapas para prever a ocorrência de falhas na rede elétrica, sua gravidade e impacto económico, utilizando dados meteorológicos e registos de interrupção de energia, com periodicidade diária de 2014 a 2023. A primeira etapa recorreu à classificação binária para prever a ocorrência de falhas; a segunda etapa, à classificação em múltiplas classes para avaliar a gravidade dessas falhas; e a terceira, à análise de regressão para estimar a energia não fornecida e custos decorrentes. Embora se reconheça que as falhas podem ter múltiplas causas, a análise identifica os fenómenos atmosféricos, especialmente o vento, como fatores determinantes. Os resultados demonstram, ainda, que modelos interpretáveis, como a regressão logística, podem ter valor operacional em termos de identificação de prioridades e planeamento.

Assim, este estudo contribui para a promoção de redes resilientes aos fenómenos atmosféricos, ao demonstrar a utilidade da aprendizagem automática supervisionada e a importância de uma melhor integração de dados na monitorização de infraestruturas críticas.

Palavras-chave: eventos meteorológicos extremos, resiliência da rede elétrica, Sicília, previsão de falhas na rede elétrica e sua gravidade, impacto económico, aprendizagem automática supervisionada, Energia Não Fornecida (ENS), dados em painel.

Título: Avaliação do Impacto de Eventos Meteorológicos Extremos nas Falhas da Rede Elétrica: Uma Análise Baseada em Dados na Sicília

Autora: Chiara Galvagno

Contents

List of Tables

List of Figures

List of Acronyms

AI	Artificial Intelligence
ARERA	Autorità di Regolazione per Energia Reti e Ambiente
AUC	Area Under the Curve
EHV, HV	Extra-high Voltage, High Voltage
ENS	Energy Not Supplied
IIA	Independence of Irrelevant Alternative
IPCC	Intergovernmental Panel on Climate Change
ISPRA	Istituto Superiore per la Protezione e la Ricerca Ambientale
KNN	K-Nearest Neighbors
MED	Major Event Day
MERIDA	MEteorological Reanalysis Italian DAtaset
OHTL	Overheard Transmission Lines
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
RF	Random Forest
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic
SAIDI	System Average Interruption Duration Index
SDG	Sustainable Development Goal
TSO	Transmission System Operator
VIF	Variance Inflaction Factor
XGBoost	Extreme Gradient Boosting

1 Introduction

Climate change is exerting an increasingly significant influence on ecosystems, societies, and critical infrastructures worldwide. One of the most visible consequences is the increase in the frequency and severity of extreme weather events, which cause widespread damage to land, urban areas, and populations, placing particular stress on infrastructures such as power grids. These grids are vital to the daily functioning of societies, supporting everything from essential services to economic fabric.

Several factors contribute to the vulnerability of power grids to extreme weather, including aging infrastructure, exposed above-ground lines, and high interconnectivity between components. Given the severe consequences of power outages, strengthening the resilience of the grid has become an urgent priority. This necessity is underscored by the recent surge in research and energy and climate policy initiatives focusing on this area. Grid resilience refers to a system's ability to withstand, recover from, and adapt to disruptive events, including those induced by climate extremes.

This thesis focuses on the region of Sicily, Italy, a Mediterranean territory characterized by a high exposure to extreme weather events and several critical grid constraints, including aging thermal capacity, spatial mismatches in renewable production, and limited topological redundancy. Despite these challenges, data-driven approaches to anticipating and mitigating outage risks in the region remain scarce.

A core limitation for both researchers and system operators is the lack of predictive tools that merge fine-grained meteorological data with high-frequency outage information. Traditional reliability indicators are insufficient for capturing the dynamics of low-probability, high-impact failures. Furthermore, region-specific insights into the drivers and consequences of weather-induced outages are essential for effective risk assessment, yet largely absent in the existing literature.

In response to these gaps, this study adopts a supervised learning approach to model and forecast power outages in Sicily based on historical weather and failure data. It introduces a three-stage modeling pipeline: first, the binary prediction of outage occurrence; second, the classification of failure severity conditional on occurrence; and third, the estimation of economic impact using regression-based forecasts of unserved energy translated into monetary terms. This framework is designed not only to improve predictive accuracy but also to provide operationally interpretable insights for grid resilience planning.

The central research questions that guide this study are the following:

1. *To what extent can high-resolution meteorological and spatial features predict the occurrence and severity of weather-related outages in the Sicilian transmission grid?*

2. *How do these outage risks vary across locations, and which weather variables most significantly affect system vulnerability?*

3. *What is the estimated economic impact of extreme-weather-induced outages, and how can predictive models inform targeted mitigation strategies and resilience investments?*

To address these questions, the analysis combines supervised classification and regression models, drawing on daily panel data at the locality level. It incorporates domain-specific metrics such as the Energy Not Supplied (ENS) and evaluates models with a focus on interpretability and out-of-sample performance.

The findings are intended to support public and private decision makers by identifying vulnerable regions and conditions, quantifying the costs of outages, and suggesting data-driven interventions to improve resilience under increasing climate stress.

The structure of the thesis is as follows. Section 2 presents a review of the literature. Section 3 describes the dataset and preprocessing steps. Section 4 combines the methodology and predictive models with the presentation of the results. Section 5 discusses their implications. Section 6 concludes the study.

2 Literature Review

In recent years, the increasing frequency and severity of extreme weather events, driven by climate change, have posed significant challenges to the resilience of power grids worldwide (Campbell, 2012; Slater et al., 2021). Resilience, in the context of complex infrastructure systems, refers to a system's capacity to anticipate, absorb, and recover from hazardous events in a timely and efficient manner (Calvin et al., 2023). Specifically, power grid resilience is defined as the ability of electrical infrastructure to withstand and recover from external disturbances, such as extreme weather events, ensuring rapid restoration of energy transmission and distribution services (Bonanno et al., 2019).

As extreme weather events become more frequent and severe, it is necessary to understand their impact on power system operations to develop effective mitigation strategies. This has led to a surge in research at the intersection of meteorology and energy infrastructure, with a focus on enhancing system resilience. However, research on the specific vulnerabilities of national and regional transmission networks in Italy remains underdeveloped.

In the Italian context, and particularly for islanded systems like Sicily, these resilience challenges are intensified by geographical constraints, aging infrastructure, and climate-driven stressors, as highlighted in recent operational analyses and planning reports by Terna S.p.A, the Italian Transmission System Operator (TSO).

This review synthesizes key research on extreme weather events, their classification and measurement, and their impact on power systems. Examines state-of-the-art methodologies for assessing these impacts, with a particular focus on power outages, defined as loss of electrical power supply to an end user (Yang et al., 2023). These events can have far-reaching consequences, including disruption of economic activity, daily life, and potential safety risks for consumers. For energy suppliers, they can lead to significant economic losses and reputational damage. Moreover, for communities, these outages may escalate into significant public safety hazards.

The review also evaluates the primary data sources available for studying these phenomena. It justifies the selection of specific datasets for weather events and grid failures, and analyzes the methodological challenges of integrating heterogeneous data. Additionally, it explores advanced analytical techniques proposed in the literature, particularly machine learning and statistical modeling, for assessing the resilience of power grids to extreme weather.

The focus is primarily on Sicily, selected due to its geographical position in the center of the Mediterranean Sea, making it particularly vulnerable to climate extremes, and the recent occurrence of unprecedented weather events (ANSA, 2025; Copernicus Data Space Ecosystem, 2023; Scicchitano et al., 2021).

The review is based on peer-reviewed articles, technical reports, and industry publications, including updated documentation from Terna.

2.1 Extreme weather events and their trends

In the last few years, the world has witnessed a significant increase in the frequency, intensity, and severity of natural disasters, largely due to climate change and its associated consequences. These events have the potential to disrupt ecosystems, infrastructure, and societal stability (Avolio et al., 2025; Calvin et al., 2023).

Traditional risk assessment models rely on the assumption of stationarity, the idea that the statistical properties of climate variables, such as mean and variance, remain constant over time (Slater et al., 2021). However, as hydroclimatic extremes become increasingly non-stationary due to atmospheric circulation shifts, land cover changes, and human interventions, these models are proving inadequate (Fowler and Hennessy, 1995; Slater et al., 2021). The increasing unpredictability of extreme events demands new forecasting approaches that account for non-stationary climate conditions.

According to Jufri et al. (2019), an *extreme weather event* is defined as a meteorological occurrence characterized by magnitudes situated at the upper or lower ends of the variability scale, representing rare conditions with low frequency but substantial potential for severe and prolonged impacts on infrastructure systems.

While these global climate shifts are affecting ecosystems worldwide, certain regions are particularly vulnerable. The Mediterranean region has been identified as a climate ”hotspot” due to its sensitivity to climate change and susceptibility to severe weather events resulting from variations in sea temperature and complex orography (Giorgi, 2006). The Intergovernmental Panel on Climate Change (IPCC) (2023) confirms the Mediterranean’s vulnerability to extreme weather events, including increased frequency and intensity of heat waves, prolonged and severe droughts, more intense storms and cyclones affecting wind and flood patterns. Sicily occupies a central position within this sensitive region and is therefore experiencing the aforementioned effects first-hand, accompanied by a decrease in average precipitation and an escalation of variability during the dry season (Giorgi, 2006). Historical records also show a statistically significant increase in extreme precipitation events, especially for short-duration storms, which are projected to intensify further between 2050 and 2100 (Forestieri et al., 2018).

2.2 Definitions and Metrics for Extreme Weather Events

Clear definitions of extreme weather events are essential for evaluating their impact on energy systems. Standardized metrics support risk quantification, forecasting, and the design of resilience strategies for transmission infrastructure.

Extreme weather events are typically classified based on magnitude, duration, and spatial extent (Jufri et al., 2019). A common approach uses return periods (T), defined as the inverse of the probability of an event’s occurrence (Ma et al., 2019). In Italy, the Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) provides national guidelines for identifying temperature and precipitation anomalies (Fioravanti et al., 2013), which remain widely used in regulatory and scientific assessments.

In power system reliability analysis, however, event classification often shifts from meteorological parameters to their operational consequences. A prominent example is the Major Event Day (MED) framework, defined in IEEE Std 1366 (2012). MEDs are identified when the System Average Interruption Duration Index (SAIDI) exceeds a statistical threshold computed using the 2.5 Beta Method. This log-normal approach ensures that extreme weather-induced disruptions do not distort long-term reliability indices.

Climate stressors like heatwaves, cold spells, intense precipitation, and drought are commonly recognized as threats to energy systems (Bonanno et al., 2019; Dwivedi et al., 2023). Yet, operational insights gathered from Terna's internal documentation and a field visit to the Misterbianco substation, including an interview with Ing. Chiarenza (Head of Plant Unit, Terna S.p.A.), emphasize that the most critical hazards in Sicily include lightning strikes, convective storms (e.g., vegetation contact and windborne debris), and wildfires. Due to data limitation, these risks are proxied using extreme temperatures and wind speeds, which are better captured in the observational dataset.

Metric	**Definition**	**Relevance to Power Grid**
Warm Spell Duration Index (WSDI)	Number of consecutive days with max temperature above the 90th percentile	Identifies prolonged heatwaves that stress transformers and increase peak energy demand
TXx / TNx	Absolute max/min temperature recorded in a given period	Indicates heat stress potential on transmission lines
Cold Spell Duration Index (CSDI)	Number of consecutive days with minimum temperature below the 10th percentile	Tracks extreme cold events that may impact grid components
RX1day / RX5day	Maximum 1-day and 5-day precipitation totals	Critical for flood risk assessment at substations and underground networks
R95p / R99p	Precipitation above 95th and 99th percentiles	Assesses extreme rainfall's role in storm-induced outages
Consecutive Dry Days (CDD)	Number of consecutive days with precipitation below 1mm	Evaluates prolonged droughts affecting hydroelectric generation

Table 1: Key meteorological indices used for assessing extreme weather events and their impact on power grids

The standardized indices described in Table 1, originally developed for comprehensive climate assessments (Fioravanti et al., 2013), provide a useful conceptual foundation for evaluating meteorological stressors on power grids.

2.3 Vulnerability of Power Grids to Extreme Weather

2.3.1 Overview

Power grid vulnerability to extreme weather arises from two interrelated dimensions: (1) direct physical damage to infrastructure, such as transmission lines, substations, and power plants, and (2) operational disruptions, including fluctuations in electricity demand, renewable energy output, and system stability (Dwivedi et al., 2023).

Yang et al. (2023) demonstrated that extreme weather events have resulted in an increasing trend of high-impact outages in the United States, notably shortening the return periods of major hurricanes. Such events have substantial economic impacts, costing billions annually due to prolonged downtimes and extensive infrastructure damage. In addition, extreme weather alters electricity demand and renewable energy output, further destabilizing the grid. For example, extreme heatwaves substantially increase cooling demands, exacerbating transformer stress and leading to voltage instability (Protezione Civile Sicilia, 2023).

Many studies have identified extreme heat, storms, heavy precipitation, droughts, and strong winds as threats to power infrastructure on a global scale. In the United States, for instance, these events were responsible for 80% of primary outages between 2003 and 2012. Furthermore, the frequency and severity of these impacts have increased in direct proportion to climate change (Kenward and Raja, 2014). A similar vulnerability has been observed in Europe, underscoring the need for advanced predictive modeling and enhanced resilience planning (Stankovski et al., 2023; Panteli and Mancarella, 2015).

2.3.2 Italian and Sicilian Context

Italy's power grid increasingly faces challenges from extreme weather events, driven primarily by rising temperatures, intense precipitation, prolonged droughts, and strong winds. Reports from Terna S.p.A. emphasize that these climate disturbances have led to frequent voltage imbalances, transformer overloads, and transmission line failures, necessitating targeted resilience interventions (Terna S.p.A., 2023).

In Sicily, specific regional vulnerabilities are particularly acute due to its Mediterranean climate and insular geography. Studies and historical data highlight several distinct weather-induced threats:

- Lightning strikes are a primary cause of infrastructure damage in Sicily, often leading to faults of overhead lines.

- Heat waves, which significantly increase electricity demand for cooling, lead to overloaded transformers and reduce the efficiency of transmission lines due to thermal stress. The frequency and intensity of these heat waves are expected to continue rising, exacerbating infrastructure stress during peak load periods (Protezione Civile Sicilia, 2023; Avolio et al., 2025).

- Heavy precipitation events and flooding pose significant threats to substations and underground power distribution networks. Storm surges and flash flooding from intense storms and Mediterranean hurricanes (Medicanes) represent additional risks for coastal transmission infrastructure (Scicchitano et al., 2021).
- Strong wind events and severe convective storms may damage overhead transmission lines and towers, causing localized blackouts and infrastructure failures (Scicchitano et al., 2021).
- Drought conditions threaten hydroelectric power production, particularly critical in regions like Sicily, and exacerbate wildfire risks along power transmission corridors. These events lead to prolonged power disruptions and reduced overall grid reliability (Terna S.p.A., 2023).
- Variability in renewable energy generation, primarily solar and wind power, introduces instability during extreme weather events, creating challenges for grid operators in balancing power supply and demand effectively (Di Gloria et al., 2024).

Other extreme weather-related risks identified in national-level studies, such as snow and ice storms causing mechanical failures (e.g., snow sleeves), are considered to be irrelevant for Sicily due to the island's mild winter climate and negligible snowfall.

Given these unique regional challenges, improving the resilience of Sicily's power grid requires customized interventions, advanced forecasting models, and integrated climate adaptation strategies. However, despite national efforts outlined by Terna, research on Sicily's grid vulnerabilities remains underdeveloped. Specifically, studies lack (1) high-resolution outage data correlated with extreme weather events, (2) predictive models tailored to the island's unique climate and grid infrastructure, and (3) resilience strategies integrating local renewable energy sources into adaptive grid management. Addressing these gaps will be crucial to improve the reliability of the network in the face of climate change (Terna S.p.A., 2023; Di Gloria et al., 2024).

2.4 Description of the Italian and Sicilian Power Grid

2.4.1 Overview

A comprehensive understanding of the Italian electric power system is essential for assessing its vulnerability to extreme weather events.

The system consists of three interconnected macro-components: generation, transmission, and distribution. Electricity generation occurs through various energy sources, including fossil fuels (natural gas, coal), renewable energy (solar, wind, hydroelectric) and energy imports through interconnected networks with neighboring countries (Terna S.p.A., 2023).

The Italian transmission grid operates under a regulated natural monopoly model, with Terna S.p.A. serving as the sole operator of the national transmission network. Terna is responsible for the planning, development, management, and maintenance of the Extra High Voltage (EHV) and High Voltage (HV) grid infrastructures across Italy. Its role is strictly governed by the national regulatory framework established by ARERA (Autorità di Regolazione per Energia Reti e Ambiente), the Italian Regulatory Authority for Energy, Networks and Environment. Close collaboration between Terna and government agencies ensures that transmission system operations align with national energy policies, grid security standards, and decarbonization targets (Terna S.p.A., 2023).

The grid has an extend over approximately 75,000 kilometers throughout Italy, and it is hierarchically structured into three voltage levels:

- Extra High Voltage (EHV): 380 kV lines form the backbone of the transmission system, interconnecting the main power generation plants, strategic substations, and cross-border interconnections. These lines are designed for large-scale bulk power transfer over long distances.

- High Voltage (HV): 220 kV and 150 kV lines distribute power regionally, linking EHV substations to local distribution networks and large industrial consumers.

- Medium Voltage (MV) and Low Voltage (LV): These are not directly managed by Terna but by local Distribution System Operators (DSOs). They distribute electricity to residential, commercial, and small industrial users.

Damage to the EHV network can lead to severe, large-scale outages (blackouts) due to the interconnected nature of the system and the sheer volume of power flows involved. However, EHV infrastructure is also engineered with greater safety margins, involving robust physical construction (wider right-of-ways, stronger towers, lightning protection systems) and sophisticated operational defenses such as automatic load shedding, islanding protocols, and fast fault isolation.

By contrast, HV infrastructure, generally experiences higher failure rates due to more extensive network coverage, closer proximity to vegetation and urban structures, and comparatively less redundancy. Failures at the HV level are more frequent but typically localized in their impacts.

The overhead transmission lines (OHTL) that connect power generators and load centers are recognized as a critical component of the power grid, due to their high exposure to external factors (Cristaldi et al., 2023). They are responsible for the stability of the network and the failure of these lines can result in significant economic and service performance losses. The severity of these outages is typically defined by the reliability, availability, and maintainability of the grid.

The Italian transmission system incorporates multi-tiered defense mechanisms designed to detect, isolate, and mitigate faults, thereby preserving system security and service continuity. These systems, formalized in Terna's operational protocols (Terna S.p.A., 2022, 2025b, 2024a), include:

- Automatic protection systems: Distance protection, differential protection, and automatic reclosing devices are deployed across the grid to detect faults and automatically isolate damaged sections within milliseconds, minimizing system disruption.
- Automatic load shedding schemes: Under-frequency and under-voltage automatic disconnection protocols (including selective load shedding and disconnection of non-programmable renewable sources) are activated to maintain system frequency and voltage stability during severe disturbances, especially in insular grids like Sicily.
- Black start capabilities: Specific power plants are equipped to restore grid service independently in the event of widespread outages, ensuring rapid system recovery.
- Dynamic real-time security assessment: The National Control Center continuously monitors electrical parameters (voltage, current, frequency) and forecasts system states to proactively manage risks and implement corrective actions before thresholds are breached.

2.4.2 Focus on the Sicilian Grid

Sicily's power grid is electrically connected to mainland Italy primarily through a set of high-voltage submarine cables, including key 380 kV interconnections linking the island to Calabria. These links form the only synchronous coupling between the island and the continental grid, making the system particularly sensitive to faults occurring along these corridors. Internally, Sicily's grid consists of a network of 380 kV and 220 kV transmission lines concentrated along the northern and eastern coasts, connecting major urban centers, industrial loads, and scattered generation plants (Terna S.p.A., 2023; Di Gloria et al., 2024).

The vulnerabilities of the Sicilian grid stem from a combination of structural, geographic, and operational factors. The limited redundancy of its high-voltage backbone means that failures, especially along interconnection points, can quickly spiral into regional blackouts. In addition, the aged infrastructure in mountainous and rural areas further increases the likelihood of physical degradation and unplanned outages. According to recent reports and field interviews, the grid's exposure to wildfires, storms, and vegetation-related faults adds further operational risk. Moreover, the island's increasing share of non-programmable renewable generation, which exceeded 38% in 2023, introduces balancing and stability challenges. During periods of high renewable output net load can fluctuate rapidly, reducing system inertia and making the grid more sensitive to frequency deviations. These effects are amplified during extreme weather events when protection schemes and load-shedding systems (such as the Western Sicily ”Telescatto” framework) must be activated. This growing reliance on intermittent renewables, while vital for decarbonization, requires improved forecasting and robust grid flexibility to avoid outages. (Terna S.p.A., 2025b, 2023; Di Gloria et al., 2024).

This evolving scenario requires targeted investments and adjustments to power grid infrastructure to maintain reliability, reduce failure risks, and align with the climate and energy transition goals established at both European and national levels. These efforts directly contribute to fulfilling Sustainable Development Goal (SDG) 7 (Affordable and Clean Energy) and SDG 13 (Climate Action), as well as to meeting the European Union's targets for achieving climate neutrality by 2050 under the European Green Deal framework.

In light of Sicily's specific vulnerabilities and the increasing pressure on its grid infrastructure, significant structural improvements and grid expansions have been planned for the coming years (Terna S.p.A., 2023). Terna's 2023 Development Plan outlines major investments, including new high-voltage direct current (HVDC) submarine connections and an expansion of the 380 kV network (Terna S.p.A., 2023). However, these projects face notable challenges, such as prolonged implementation timelines, environmental constraints, and the complexity of integrating high shares of renewable generation into a traditionally centralized grid.

Figure 1 illustrates the existing transmission network (panel a) compared to the enhancements planned for 2030 (panel b).

2.5 Methodologies for Analyzing Weather Impacts on Power Grids

To understand the repercussions of extreme weather events on power grids, robust analytical methodologies that integrate meteorological data with power system performance metrics are needed. Various approaches have been developed for this purpose, ranging from statistical analyses of historical outages to advanced machine-learning models that predict grid failures under different weather conditions. High-resolution reanalysis datasets, numerical weather prediction (NWP) models, and probabilistic risk assessments play a crucial role in quantifying weather-related disruptions and enhancing grid resilience. By leveraging these methodologies, researchers can identify vulnerabilities, improve forecasting accuracy, and develop adaptive strategies to mitigate climate-induced risks to power infrastructure.

2.5.1 Power Grid Resilience

Jufri et al. (2019) provide a comprehensive literature review that serves as a state-of-the-art examination of research conducted over the years on power grid resilience. This field is now gaining increasing attention due to the rising frequency and intensity of extreme weather events driven by climate change. One of the key aspects analyzed in their work is the distinction between grid resilience and grid reliability.

Reliability refers to a grid's ability to perform its function under normal conditions, with a primary focus on minimizing momentary and sustained interruptions within a predefined time

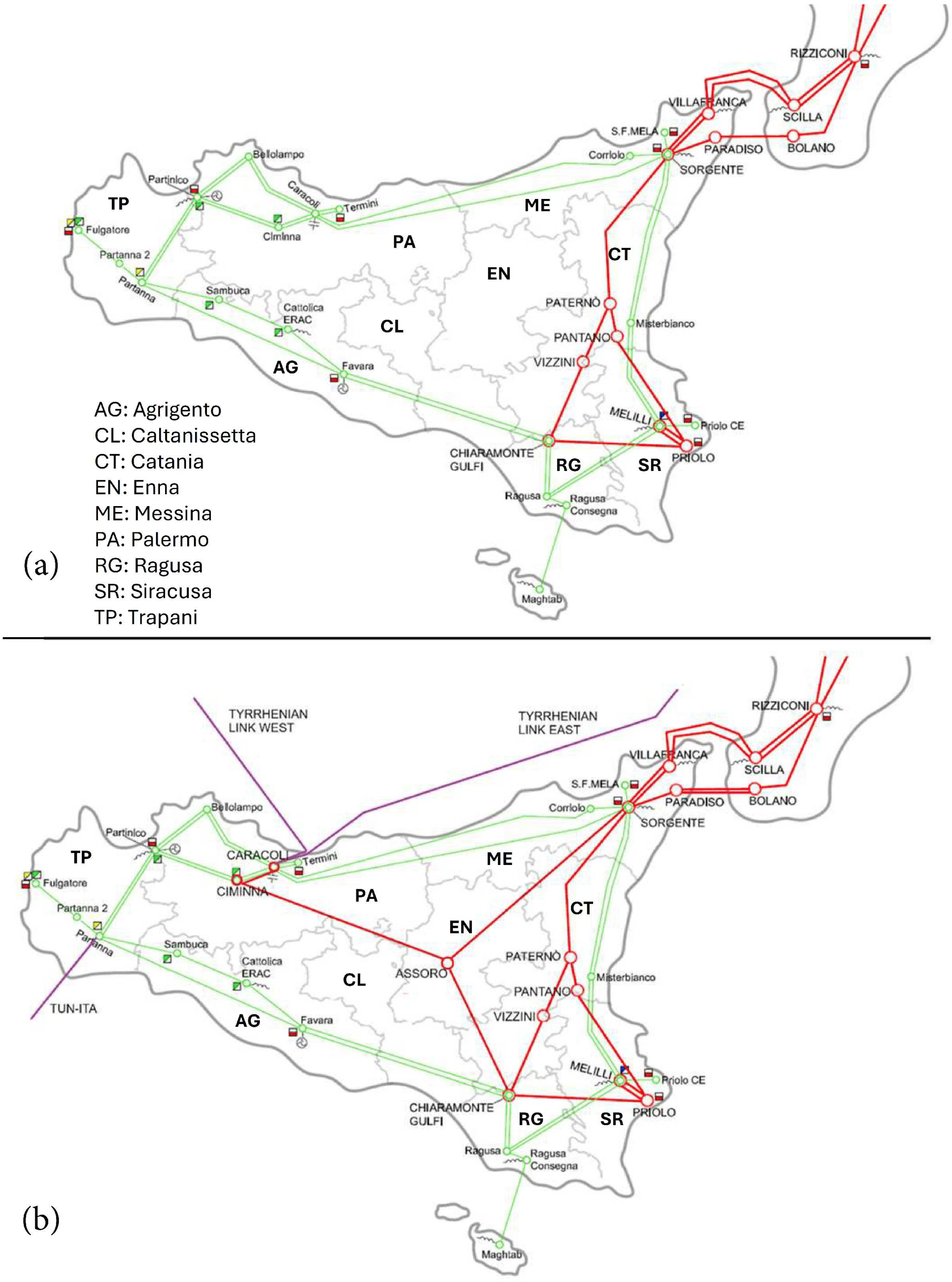

Figure 1: Sicilian power grid infrastructure. Current transmission grid as of 2024 (a) and planned transmission grid enhancements by 2030 according to Terna's 2023 Development Plan (b). Red lines represent the 380 kV grid, green lines indicate the 220 kV grid, and purple lines represent HVDC links. Source: Adapted from Di Gloria et al. (2024).

range (IEEE Std 1366, 2012).

By contrast, grid resilience is defined as the ability to anticipate, absorb, recover from, and adapt to extreme weather events. Unlike reliability, resilience focuses on low-probability, high-impact events such as hurricanes, floods, and wildfires. It requires active operational strategies rather than passive reliability measures and evaluates both system state transitions and infrastructure recovery time. Moreover, resilience assessments consider how quickly and effectively the grid restores functionality after disruptions. Since resilience methodologies must account for non-stationary climate risks, a grid that is considered reliable under normal conditions is not necessarily resilient to extreme weather conditions.

The resilience of the grid is influenced by three critical factors: weather intensity, exposure to the grid, and vulnerability to the grid. Weather intensity represents the severity of meteorological disturbances, ranging from normal fluctuations to extreme events such as storms, floods, and heatwaves. Grid exposure measures the extent to which power infrastructure is subjected to such disturbances, with higher exposure in regions prone to severe weather. Grid vulnerability, in turn, assesses how susceptible the system is to failures, with older or poorly maintained networks exhibiting greater fragility. As illustrated in Figure 2, the combined effect of these three parameters determines the overall impact on grid operations. Even in the presence of extreme weather events, a grid with low exposure and low vulnerability can maintain its functionality, whereas highly exposed and fragile grids are more likely to suffer severe disruptions.

Building on these resilience principles, the present study adopts a predictive modeling approach to operationalize the vulnerability analysis of the grid, described in detail in Section 4.

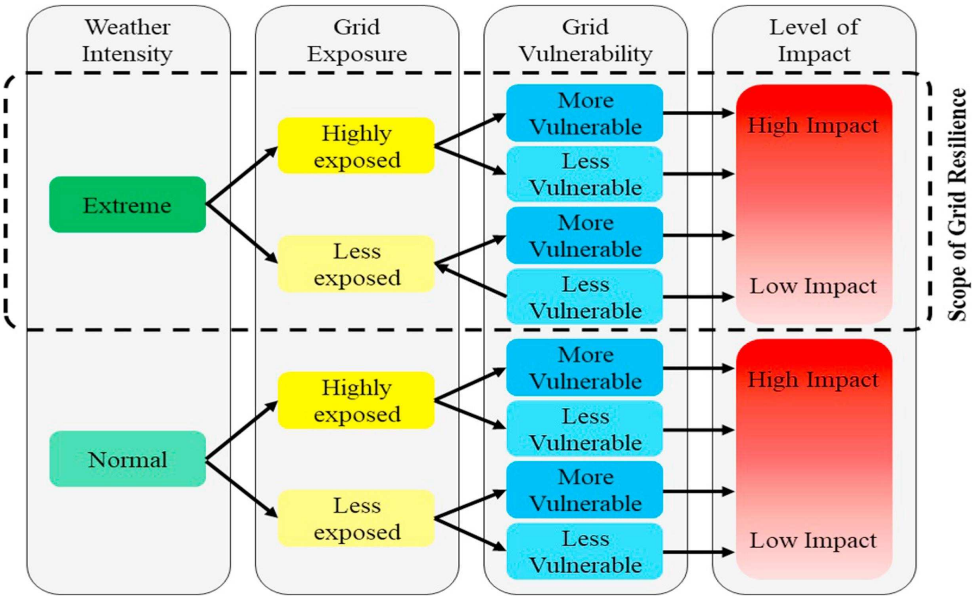

Figure 2: Factors influencing grid resilience, showing the interplay between weather intensity, grid exposure, and grid vulnerability. Adapted from Jufri et al. (2019).

2.5.2 Data Sources for Weather and Grid Analysis

Conducting reliable analyses to enhance the resilience of power grids requires robust meteorological and grid data. For Sicily, meteorological information is available from several institutional sources. The Servizio Informativo Agrometeorologico Siciliano (SIAS) provides real-time measurements every 10 minutes, including temperature, humidity, precipitation, and wind speed. ARPA Sicilia and the Dipartimento Regionale della Protezione Civile offer additional validated datasets.

While real-time observations provide high-frequency data, they often suffer from limited spatial coverage and historical depth. For long-term climate analysis, the MERIDA (Meteorological Reanalysis Italian DAtaset) project offers a high-resolution (7 km) reanalysis specifically developed to support energy sector resilience planning, integrating ERA5 global data with refined local modeling (Bonanno et al., 2019).

Regarding grid data, Terna S.p.A. provides mandatory reports on energy consumption, production, and outages. However, the available formats (often PDFs) and aggregation levels pose challenges for integration with meteorological datasets. Advanced preprocessing steps, including spatial interpolation and anomaly detection, are necessary to harmonize weather and grid failure data for analysis.

This research adopts the structured data pipeline developed by Moschetti and Barozzi (2024), which systematically integrates these heterogeneous sources into a relational database suitable for predictive modeling.

2.5.3 Techniques for Measuring Weather Impacts on Grids

A wide range of techniques has been developed to quantify and predict the impact of extreme weather on power grid performance. These include traditional statistical methods, engineering-based reliability frameworks, and increasingly, predictive models using machine learning.

Cristaldi et al. (2023) apply the Reliability, Availability, Maintainability (RAM) framework to evaluate historical failure trends in the Italian transmission network. Their work highlights rising failure rates in OHTLs due to aging infrastructure and increasing weather-related stress. By integrating the duration of the failure and the energy undeliverable into a severity factor (SF), the study quantifies the growing challenge of maintaining service continuity under adverse conditions.

Complementing these retrospective analyses, predictive modeling techniques aim to anticipate outage risks based on weather conditions. Outage Prediction Models (OPMs) use supervised learning algorithms, such as Random Forests and Gradient Boosting Machines, to estimate the probability and severity of weather-induced failures (Yang et al., 2023). These models can guide preventive reinforcement strategies by identifying high-risk regions. Outage Prediction Trees (OPTs), while similar, offer interpretable decision frameworks for operational decision-making.

For long-term risk assessment, statistical approaches like the Generalized Extreme Value

(GEV) distribution and Probability-Weighted Moments (PWM) are used to estimate the return periods of rare but high-impact weather events. These models help utilities evaluate infrastructure vulnerability to unprecedented storms and support resilience planning.

Table 2 summarizes these methodologies and their relevance to improving grid resilience in the face of climate extremes.

The present study builds on these approaches by combining supervised learning with economic severity estimation, offering an integrated predictive framework tailored to the Sicilian grid context.

Methodology	**Description**	**Relevance**
Reliability, Availability, Maintainability (RAM) Analysis	Evaluates Failure Rate, Uptime Ratio, Mean Time To Repair, and Undelivered Energy	Identifies grid weaknesses and failure trends under extreme weather
Outage Prediction Models (OPM)	ML models (RF, GBM) predicting outage likelihood and severity	Enhances weather-induced failure forecasting and grid reinforcement
Outage Prediction Trees (OPT)	Decision trees classifying outage risks based on weather conditions	Provides an interpretable risk framework for grid operators and policymakers.
Generalized Extreme Value (GEV) Distribution	Estimates return periods of extreme weather events	Assesses long-term outage risks from severe storms
Probability-Weighted Moments (PWM)	Refines extreme event modeling for rare but severe weather	Improves worst-case outage risk assessment

Table 2: Summary of methodologies for assessing power grid vulnerability to extreme weather events

2.6 Conclusion of the Literature Review

This review underscores the growing vulnerability of power grids to extreme weather events and the need for robust analytical methodologies to assess and mitigate their impacts. Despite substantial research on outage prediction and grid resilience, significant gaps remain in the understanding of regional vulnerabilities, particularly in Sicily. Addressing these challenges requires the integration of high-resolution meteorological data with grid performance metrics to develop predictive models that reflect evolving climate risks. This study builds on prior approaches by enhancing outage prediction and resilience assessment through structured data integration and the application of interpretable machine learning techniques.

3 Data Collection, Integration and Preprocessing

3.1 Data Sources

This study integrates two primary data sources to analyze the relationship between extreme weather events and power grid outages in Sicily during the period 2014–2023. The first dataset comprises high-resolution meteorological data, while the second contains detailed records of electricity outages and related grid information.

The meteorological data were obtained from the publicly available MERIDA dataset, which contains hourly records of weather conditions across Italy. The processed dataset includes temperature, precipitation, and wind speed measurements in Sicily, aggregated on a daily basis. In addition, binary indicators of extreme weather events were added to the dataset, based on percentile thresholds.

Power grid data were made available through collaboration with Carlo Barozzi, co-author of the MSc thesis by Moschetti and Barozzi (2024), and are derived from official outage records maintained by Terna S.p.A. The dataset is structured according to the classification system described in the technical guide (Terna S.p.A., 2017). The dataset includes, for each event, information on duration, cause, grid component affected, interruption type, asset ownership, and net energy not supplied (ENS). Outages marked as planned interventions (e.g., preventive maintenance) were excluded from the analysis to focus exclusively on unplanned events of operational concern.

3.2 Data Integration

The two datasets were merged using a left join on the daily-locality level, with meteorological observations serving as the base structure. This resulted in a panel dataset in which each row represents the weather and outage conditions for a given locality on a specific day. Localities correspond to operationally defined zones derived from Terna's substation coverage and meteorological station proximity. A binary variable, `outage_occurred`, was constructed to indicate whether at least one outage occurred on that day in the respective locality.

Notably, the original outage data were not recorded at the daily level but instead consisted of a separate row for each individual event. To align this structure with the panel format, outage events were aggregated at the daily-locality level. In cases where multiple outages occurred within the same day and locality, numerical fields such as Energy Not Supplied (`ENS_ENR_netta`) and disconnection duration were summed, while categorical variables, including cause (`codice_causa_AEEG`), interruption type (`tipo_interruzione`), and grid component affected (`elemento_rete_origine`), were aggregated into lists to retain multi-event granularity within a single record.

After merging and filtering for unplanned events, the resulting dataset includes 447,406 daily records across 124 localities in Sicily, spanning the period from 2014 to 2023. Outage events were infrequent, occurring in approximately 0.25% of all observations.[1]

Weather conditions displayed high spatial and temporal heterogeneity, with observed extremes in temperature, wind speed, and precipitation varying significantly across both seasons and geographical areas.

3.3 Variable Summary, Operational Meaning, and Validation

3.3.1 Meteorological Variables

The key meteorological variables include daily mean and maximum temperature, mean and maximum precipitation, and mean and maximum wind speed.

The selection of meteorological variables was based on their demonstrated capacity to influence the stability and performance of the power grid within the Sicilian context.

Increases in high air temperatures (`T2_C_max`) have the potential to increase line sag and reduce the cooling efficiency of transformers, especially in aging infrastructure. A correlation has been identified between these phenomena and an escalation in the incidence of fires, which represent potential risks to the integrity of power grid components. Furthermore, prolonged heatwaves often coincide with elevated electricity demand due to air conditioning use, further stressing the grid.

Precipitation variables capture both the intensity (`PREC_max`) and persistence (`PREC_max_roll3`) of rainfall. These conditions can precipitate flooding, insulation breakdowns, and landslides, which have been demonstrated to affect both overhead and underground lines.

Furthermore, wind gusts, defined as the maximum velocity of the wind, are of particular concern, as they can lead to various adverse effects, including contact between vegetation and conductors, conductor swinging, and mechanical damage to poles and towers. Cumulative wind indicators, such as the three-day maximum, were included to account for stress accumulation on components.

Each meteorological variable was transformed into a binary indicator flag denoting the occurrence of an extreme event, based on variable-specific criteria. For temperature and wind speed, extreme conditions were defined using the 99th percentile of the historical distribution across all observations in Sicily, capturing tail risk consistent with grid resilience literature (Jufri et al., 2019). For precipitation, due to the high frequency of zero-rain days and spatial heterogeneity in rainfall regimes, a percentile-based approach was deemed unsuitable. Instead, the threshold was defined as 80% of the RX1day (maximum daily precipitation) value for each locality, consistent with ISPRA guidelines (Fioravanti et al., 2013).

The resulting binary indicators, `extreme_heat_flag`, `extreme_rain_flag`, and `extreme_wind_flag`, were then used to flag days with severe weather at the local level. A day was considered ”extreme” if at least one of these flags was active. Figure 3 visualizes the empirical distributions of

---

[1] This excludes outages marked as “planned interventions” in the dataset.

the three variables alongside the selected thresholds.

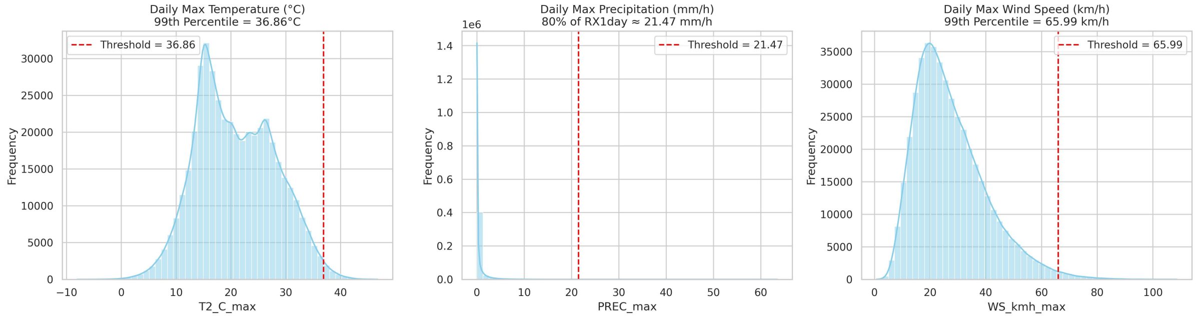

Figure 3: Distribution of key meteorological variables. The dashed red lines indicate the local extreme event threshold used to define extreme event flags.

As illustrated in Figure 3, the empirical distributions of precipitation and wind speed are highly right-skewed, with long tails capturing rare but intense events. To reduce this skewness and stabilize variance during the modeling phase, a log(1+x) transformation was applied to these variables, improving convergence and interpretability in machine learning models.

The thresholds and corresponding frequencies of extreme event days across the full dataset are reported in Table 3. These frequencies reflect the proportion of daily-locality observations classified as extreme by each weather dimension.

Table 3: Extreme Weather Thresholds and Relative Frequencies

Extreme Event Type	**Threshold**	**% of Days Triggered**
Extreme heat	36.86 °C	1%
Extreme precipitation	21.47 mm/h	0.04%
Extreme wind	65.99 km/h	1%

3.3.2 Outage Variables

An outage (”interruzione”) in the context of Terna’s classification system (Terna S.p.A., 2017) refers to the unavailability of one or more components of the power grid due to faults, protection triggers, or operational interventions that result in disconnection. These disconnections may be automatic or manual and do not necessarily entail a loss of service to end users. Outages are recorded across both the transmission and distribution networks and are characterized by a range of attributes including voltage level, duration, affected infrastructure, and underlying cause.

From a technical perspective, a power grid outage is often the consequence of a disturbance that causes voltage or frequency to fall outside safe operating limits. In normal conditions, grid frequency is tightly regulated around 50 Hz in Italy. Deviations as small as ±0.2 Hz can trigger automatic protection mechanisms. Likewise, voltage instability, resulting from overloads, faults, or sudden disconnections, can lead to cascading failures. When such thresholds are breached, circuit breakers isolate the affected section, and depending on the severity, under-frequency or under-voltage load shedding may be activated. These mechanisms prevent broader system collapse but result in localized or regional blackouts, which are later quantified using ENS metrics and interruption classifications.

Within this framework, the outage-related variables in the dataset provide a structured representation of power system disruptions. Central among these is the variable `ENS_ENR_netta` (Net Energy Not Supplied), which quantifies the magnitude of service loss in megawatt-hours (MWh). The `ENS_ENR_netta` quantifies the residual energy not delivered to end-users after rerouting or counterfeed measures. It reflects the operational impact of the outage and is reported only when mitigation capacity is assessable.

One important aspect of ENS is that many outage records report a value of zero. This occurs when:

- The event involved no actual interruption to end-user consumption (e.g., preventive operations, protection activations without service interruption),
- The affected component was not directly connected to a load (e.g., generation-side or system-defense disruptions),
- Effective countermeasures or system redundancies (such as loop-fed networks) avoided actual delivery shortfalls. This is systematically reflected in the variable `NF/NR/SD`, which encodes the nature of the outage impact. ”NF” corresponds to load-side disconnections with energy not supplied; ”NR” to generator-side disconnections (energy not withdrawn); and ”SD” to defense-related disconnections with potential ENS but usually mitigated by automated grid responses.

Each outage is categorized by its primary cause using the variable `codice_causa_AEEG_1_livello`, which follows a regulatory taxonomy defined by the Italian energy authority. This classification groups events into high-level causal categories:

- Resource inadequacy: imbalances between electricity supply and demand;
- Force majeure: exceptional natural events such as lightning or earthquakes and extreme weather events;
- External causes: disruptions caused by vegetation encroachment, animal interference, or third-party damage;
- Other causes: including mechanical failures, equipment malfunction, and operational errors;
- Programmed outages: pre-scheduled maintenance or grid interventions.

To ensure a focus on unpredictable and potentially weather-related events, programmed outages were excluded from the predictive modeling phase. This filtering enhances the statistical validity of the ENS analysis by avoiding confounded inferences due to deterministic scheduling

This study does not rely on Terna's "extreme weather" label, whose attribution rules are not publicly documented. Moreover, outages attributed to other causes, such as vegetation interference, overloads, or equipment faults, may also be weather-sensitive. Instead, weather effects are defined exogenously through high-resolution meteorological data to ensure transparency and replicability.

The field `tipo_interruzione` categorizes the nature of the outage by its duration and persistence rather than the component affected. Specifically, "L" denotes a long interruption, "B" a brief interruption, and "T" a transient event. This classification is essential to distinguish sustained disruptions, which are more likely to result in energy not supplied, from short or automatically recovered events that may not impact the delivery to the end user.

The dataset also records the voltage level at which each outage originated via the variable `tensione_rete_origine`. This classification follows the hierarchy defined by Terna (Terna S.p.A., 2017), which distinguishes between Extra High Voltage , High Voltage and Medium Voltage networks.

ENS and Severity Class Indicator To improve interpretability and facilitate operational prioritization, a new categorical variable `ENS_level` was created to summarize the severity of each outage based on its net energy not delivered (`ENS_ENR_netta`) and cause. Outages associated with planned actions and zero energy loss were labeled *Very Low*, reflecting preplanned actions without service impact. Other zero ENS events were labeled *Low*, while those with ENS between 0 and 1 MWh were classified as *Medium*. Outages greater than 1 MWh were categorized as *High*, reflecting significant disruptions. The distribution of ENS levels across all events is shown in Table 4.

Table 4: Distribution of ENS Severity Levels

ENS Level	**Proportion**
Very Low	55.2%
Low	28.7%
Medium	9.6%
High	6.5%

To capture the joint impact of duration and energy loss, a composite severity classification was constructed by combining the `ENS_level` with the `tipo_interruzione` variable. The matrix in Figure 4 illustrates this rule-based classification logic and shows the distribution of outages mapped to their respective severity classes. The derived variable `severity_class` was used as a multiclass target in the modeling phase.

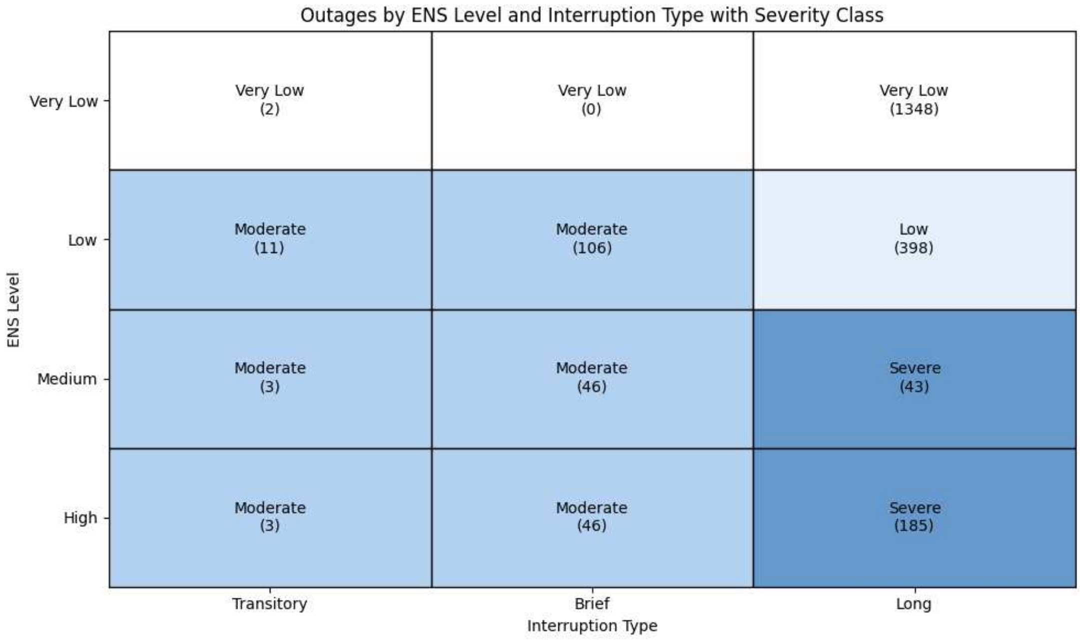

Figure 4: Outage Count by ENS Level and Interruption Type, Mapped to Severity Class

The resulting distribution of events across composite severity classes is displayed in Table 5. This variable not only enables the modeling of the occurrence of the outage but also supports the prediction of the severity of future disruptions, thus informing resilience planning and economic impact assessment. As already mentioned, the programmed outages will be excluded from further analysis.

Table 5: Outage Events by Composite Severity Class

Severity Class	**% of All Outage Events**
Not Severe - Programmed	56.92%
Very Low	12.46%
Low	16.26%
Moderate	7.56%
Severe	6.78%

Severity classifications improve the operational interpretability of outages. Events labeled *Very Low* result from preplanned maintenance or asset testing and do not result in actual energy loss. *Low* severity events also report zero MWh of ENS, but result from unplanned or reactive interventions, such as fault responses or protection activations, that were successfully mitigated by system redundancies or rerouting. While these events do not disrupt supply to end users, they do signal an underlying stress in the network that the TSO is actively monitoring and trying to prevent. Moving up the scale, *Moderate* severity events (1 MWh or less) indicate limited disturbances that may affect small users or peripheral substations. In comparison, *Severe* events, those exceeding 1 MWh of ENS, can correspond to significant interruptions. For example, 1 MWh of ENS is equivalent to the daily electricity consumption of about 30 average Italian households, while outages of 50 MWh or more could affect industrial sites, public infrastructure, or entire communities, with significant economic and social consequences.

While the *Severe* class encompasses a wide range of magnitudes, this grouping was chosen to maintain as much statistical balance as possible across classes after excluding programmed events. A more granular breakdown is provided in the Exploratory Data Analysis section.

The detailed structure of the dataset is presented in Appendix A.

3.4 Exploratory Data Analysis and Diagnostics

3.4.1 Overview

An initial descriptive analysis was conducted to assess variable distributions, structural completeness, and statistical patterns relevant to both predictive modeling and operational diagnostics. All analyses in this section are conducted on the filtered dataset excluding programmed outages, in alignment with the focus on unplanned, weather-relevant events.

Table 6 reports descriptive statistics for the main variables included in the analysis.

Table 6: Descriptive Statistics of Key Variables (Excluding Programmed Outages)

Variable	**Count**	**Mean**	**Std Dev**	**Median**	**Min**	**Max**
`T2_C_mean` (°C)	447,406	17.14	6.68	16.67	-9.35	38.44
`T2_C_max` (°C)	447,406	20.89	7.18	20.20	-8.03	46.74
`PREC_mean` (mm/h)	447,406	0.07	0.23	0.00	0.00	8.57
`PREC_max` (mm/h)	447,406	0.48	1.45	0.00	0.00	63.46
`WS_kmh_mean` (km/h)	447,406	14.29	8.99	11.85	0.56	87.01
`WS_kmh_max` (km/h)	447,406	27.45	12.77	24.99	0.74	108.60
`n_outages` (if outage)	1,097	5.83	8.93	3.00	1.00	166.00
`ENS_ENR_netta` (MWh)	1,097	7.89	74.05	0.00	0.00	2,052.20

Weather-related variables exhibit substantial variability across the ten-year period. Temperature ranges reflect seasonal fluctuations with relatively symmetric distributions, whereas precipitation and wind variables are dominated by frequent near-zero values and punctuated by rare, intense events.

Outage-related variables are reported only for the small fraction of observations that experienced at least one unplanned outage. These cases represent 1,097 daily events over a ten-year period. Among these, the number of outages per day ranges from 1 to 166, with a median of 3 events, highlighting the occurrence of multiple outage days requiring aggregation.

The `ENS_ENR_netta` variable has a highly concentrated distribution: the median is zero, indicating that most outages did not result in any measurable energy not delivered, while a small number of high-impact events account for extreme values exceeding 2,000 MWh. This long-tailed behavior justifies the use of a severity classification scheme and has implications for both model calibration and interpretation, as rare but consequential events disproportionately influence the variance of the results.

The top 1% of events by net energy not supplied (`ENS_ENR_netta`) reveal not only the scale of potential disruptions but also their strong ties to environmental stressors. As shown in Table 7, these extreme events all occurred on high-voltage lines (150–220 kV) and involved long interruptions, confirming their criticality within the grid's transmission backbone. Several of these disruptions coincide with days marked by extreme weather conditions. For instance, the most severe outage on January 19, 2017, in Caltavuturo, exceeding 2,000 MWh,occurred during an anomalous cold wave. During that same period, the Protezione Civile had issued an alert. The majority of causes are classified as "External" or "Other", indicating exposure to environmental hazards or mechanical failures, while the spatial clustering in the Madonie region and along the southeastern axis points to geographical concentrations of grid fragility.

Table 7: Top 1% ENS Events with Associated Weather Conditions

Date	**Locality**	**ENS** (MWh)	**# Out.**	**Type**	**Voltage** (kV)	**Cause**	**T$_{\text{max}}$** (°C)	**PREC$_{\text{max}}$** (mm/h)	**WS$_{\text{max}}$** (km/h)
2014-03-27	Raddusa	681.72	3	Long	150	External	13.37	2.07	55.64
2014-06-05	Porto Empedocle	147.81	1	Long	150	Other	20.03	0.00	34.51
2016-08-07	Caltavuturo	573.76	4	Long	150	External	24.37	0.00	17.96
2017-01-19	Caltavuturo	2052.20	10	Long	150	External	8.57	0.00	20.85
2017-03-08	Termini Imerese	454.11	6	Long	220	Other	13.62	4.23	43.26
2017-03-26	Gangi	206.16	11	Long	150	Other + Ext.	18.66	0.00	9.46
2017-05-22	Siracusa	197.32	3	Long	150	Other	20.26	0.00	25.07
2017-09-01	Catania	476.25	6	Long	150	Other	34.42	0.00	48.62
2022-11-22	Caltabellotta	146.30	14	Long	150	Other	16.29	1.32	58.60
2023-02-10	Castiglione di Sicilia	276.74	15	Long	150	Other + FM	-0.40	7.91	97.54
2023-06-30	Campofelice di Roccella	543.09	9	Long	150	Programmed + Other	28.97	0.00	14.10

Table 8 summarizes the key statistical properties of the variables under study. Outliers were observed primarily in precipitation and wind measurements, corresponding to real extreme weather events.

These fields exhibit expected structural missingness, since they are only populated when an outage occurs. This is handled naturally in the modeling framework described in Section 4.

These descriptive insights provided the basis for the data preprocessing steps described in the following section, including log-transformation and resampling strategies of variables to address the identified challenges.

Table 8: Summary of Data Diagnostic Findings

Aspect	Findings
Dataset structure	447,406 daily records across 124 localities and 9 provinces, 2014–2023
Temperature variables	Symmetric, bimodal distribution, no skewness issues
Precipitation variables	Highly right-skewed with extreme values, corresponding to real extreme events
Wind variables	Right-skewed distribution with significant outliers
Correlation analysis	Strong internal correlation in mean/max pairs; no multicollinearity risk concerning outage occurrence
Outliers	Detected mainly in precipitation and wind variables; outliers retained as valid extreme events
Normality test	Precipitation and wind variables non-normal; highly skewed distributions
Seasonality	Extreme weather events show clear seasonal patterns; outage occurrence is sporadic

Table 9: Missing Values and Class Imbalance Overview

Aspect	Findings
Missing values	Structural missingness in outage-specific variables, corresponding to absence of outage events
Outage occurrence imbalance	Approximately 0.25% of records correspond to days with outages; severe imbalance in target variable
Extreme events frequency	Seasonal peaks in extreme heat, rain, and wind events; no clear seasonality in outage occurrences

3.4.2 Normalization and Variable Transformation

To improve model stability, a log transformation was applied to precipitation and wind speed variables, which exhibit strong right skew and many near-zero observations. The same transformation was later applied to the `ENS_ENR_netta` variable during severity modeling to handle its heavy-tailed distribution, as shown in Figures 5, 6, and 7.

3.4.3 Handling Class Imbalance

The presence of class imbalance in the dataset is an inherent characteristic of the problem domain rather than a methodological flaw. Weather-related blackouts are rare, low-probability events with high-impact consequences. As such, it is expected that the majority of daily observations report no outage occurrence.

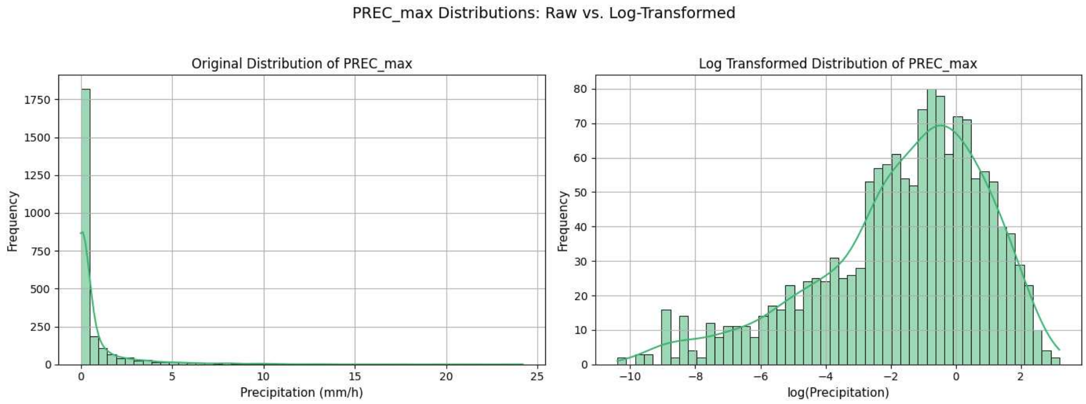

Figure 5: Comparison of the raw and log-transformed distributions of `PREC_max`.

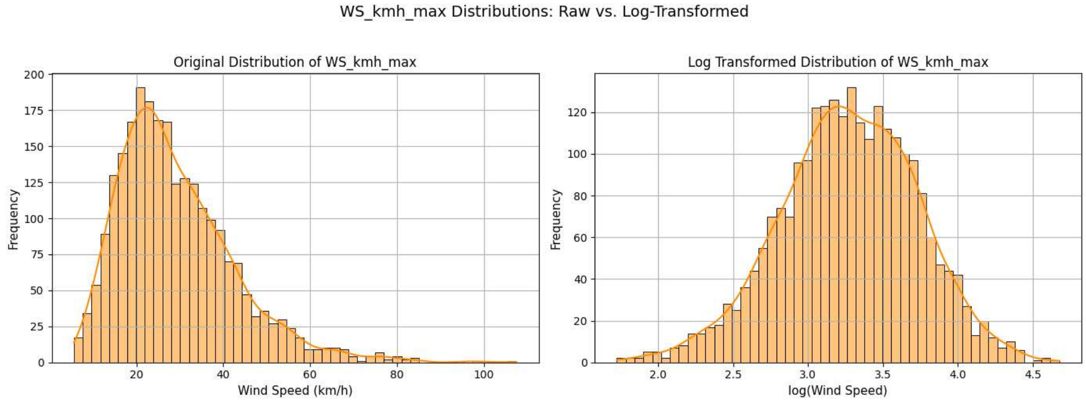

Figure 6: Comparison of the raw and log-transformed distributions of `WS_kmh_max`.

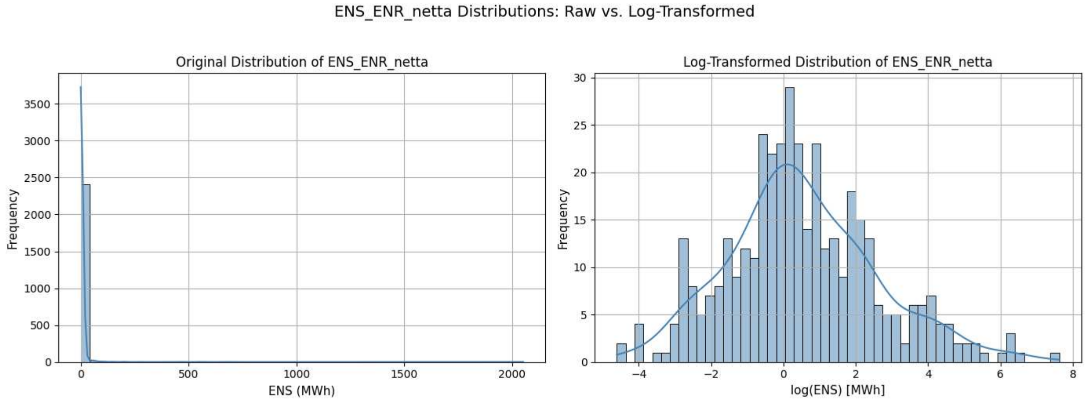

Figure 7: Comparison of the raw and log-transformed distributions of `ENS_ENR_netta`.

Rather than attempting to artificially balance the dataset during preprocessing, the class imbalance was addressed during model development and evaluation. A targeted downsampling strategy was used to reduce the dominance of examples in the majority class. This was done carefully, preserving a representative sample of non-failure cases to ensure generalization, while allowing the models to better learn the minority class signal.

In addition, modeling algorithms and scoring metrics were selected with imbalance sensitivity in mind. Algorithms such as XGBoost were used with built-in mechanisms for class weight adjustment, and evaluation focused on precision-recall performance rather than accuracy. This avoids misleading conclusions that can arise in highly skewed datasets and ensures that the model's ability to detect rare but critical failure events is properly assessed.

Together, these data preparation and exploration steps establish the foundation for the modeling pipeline described in the next section, which aims to predict outage occurrence, severity class, and the cost of energy not supplied.

4 Methodology and Results

This section presents the three-stage predictive modeling pipeline developed to analyze and forecast power outages in Sicily. The first stage involves a binary classification task to estimate the daily probability of an outage occurring in a given locality. The second stage, conditional on the occurrence of an outage, predicts its severity using multi-class classification. Finally, the third stage estimates the energy not supplied (ENS) associated with the event through regression, allowing a monetary quantification of its economic impact.

Each stage is implemented using supervised learning models chosen for their suitability to the task and interpretability. Particular attention is paid to class imbalance, fixed effect structure, temporal validation, and feature transformation.

This modular structure mirrors the real-world prioritization process: first identifying potential failures, then assessing their severity, and finally estimating economic consequences.

4.1 Stage 1: Binary Classification of Outage Occurrence

This first stage focuses on estimating the daily probability of a power outage in each Sicilian locality using meteorological data. The goal is to develop a predictive risk model that can support early warning and preventive resource allocation under extreme weather conditions.

Formally, the dependent variable is a binary indicator $\texttt{outage\_occurred}_{it} \in \{0, 1\}$, where i indexes the locality and t the day. This stage aims to establish whether extreme weather conditions can predict the likelihood of grid failure, a critical step in identifying vulnerabilities and anticipating operational risk under evolving climatic stressors.

The baseline model is a logistic regression, where the conditional probability of an outage is modeled as:

$$\mathbb{P}(\texttt{outage\_occurred}_{it} = 1 \mid \mathbf{x}_{it}) = \frac{\exp(\beta_0 + \beta_1 \cdot \texttt{T2\_C\_max}_{it} + \beta_2 \cdot \texttt{PREC\_max\_log}_{it} + \beta_3 \cdot \texttt{WS\_kmh\_max\_log}_{it} + \alpha_i)}{1 + \exp(\beta_0 + \beta_1 \cdot \texttt{T2\_C\_max}_{it} + \beta_2 \cdot \texttt{PREC\_max\_log}_{it} + \beta_3 \cdot \texttt{WS\_kmh\_max\_log}_{it} + \alpha_i)} \tag{1}$$

This is the canonical logistic function, as formalized in Taddy (2019), where α_i represents unobserved locality-specific fixed effects. The covariates include maximum temperature, and log-transformed maximum precipitation and wind speed. This formulation leverages temporal variation within each locality to extract predictive signals from daily weather dynamics.

The logistic model was complemented by random forest and XGBoost classifiers, which can capture complex interactions and non-linearities without requiring an explicit functional form. Several models were tested before finalizing the specifications.

4.1.1 Data Preprocessing

Due to the extremely low outage frequency (0.25%), the dataset is highly imbalanced, necessitating targeted rebalancing techniques in training to prevent model bias toward the majority class.

The dataset is temporally segmented to maintain the integrity of predictive evaluation: data prior to January 1, 2021, is used for model training, while data from 2021 onward is allocated for testing.

Categorical localities are accounted for via fixed effects in the logistic model or through one-hot encoding in the tree-based classifiers.

4.1.2 Model Evaluation Criteria

The performance of the models was evaluated using a set of metrics tailored to imbalanced classification tasks. The Area Under the ROC Curve (AUC) was identified as the primary metric, as it quantifies the model's effectiveness in discriminating between true outage days and non-outage days across all classification thresholds. Precision and recall were also reported to quantify the trade-off between detecting true outages and minimizing false alarms, i.e. false positives. Furthermore, top-k lift charts were utilized to assess the concentration of true outage events among the highest-risk predictions, thereby demonstrating the model's effectiveness in prioritization scenarios. Given the operational aim of identifying the most at-risk days, evaluation metrics were selected to reward ranking quality and detection coverage rather than overall accuracy.

Despite the implementation of tuning and rebalancing procedures, precision remained at a low level across all models. This reflects two key factors: first, it underscores the rarity of the target event; second, it indicates that weather alone cannot fully explain occurrences of outage. Unobserved factors, including grid infrastructure characteristics, vegetation interference, and operational disruptions, likely contribute substantially to the outcome; however, these factors are not captured in the current dataset.

4.1.3 Model Design

During model development, multiple configurations of covariates and spatial identifiers were evaluated. For the fixed effects structure, both province-level and locality-level identifiers were tested. While province-level effects provided greater parsimony and interpretability, they offered a coarser spatial resolution ($n = 9$). Locality-level fixed effects ($n = 124$) were found to more effectively capture spatial heterogeneity in exposure, infrastructure, and topographic risk.

Diagnostic tests were conducted to assess the implications of the high dimensionality associated with locality dummies. Variance Inflation Factor (VIF) analysis revealed no concerning multicollinearity, with all VIF values below 4. L2-regularized logistic regression (Ridge) confirmed that the coefficient estimates were stable and interpretable. Based on these results, locality fixed effects were retained as a statistically sound and operationally meaningful model for time-invariant spatial heterogeneity within the panel structure.

Meanwhile, the temporal autocorrelation of the meteorological predictors was evaluated using autocorrelation function (ACF) plots. Strong short-term persistence was observed in all three weather variables, especially temperature and wind speed. To account for this temporal structure, three-day rolling averages were computed for each weather feature. These smoothed covariates replaced the raw daily values in the model to improve its ability to capture sustained weather stress events and reduce noise from day-to-day fluctuations. This adjustment aligns the model more closely with the temporal dynamics of extreme weather systems.

Finally, the residuals of the logistic regression model were examined for temporal autocorrelation using ACF plots. The residual series, derived from a ridge-penalized logistic model, showed no statistically significant autocorrelation at lags up to 30 days. This confirms that the temporal structure of the data was adequately captured by the model specification, and that residual dependence across days is not a concern.

4.1.4 Comparison and Interpretation of Final Classifiers

The three final classifiers, logistic regression, random forest, and XGBoost, were selected based on their out-of-sample performance and interpretability.

Figure 8 shows the ROC curves, which visualize the trade-off between true positive rate (recall or sensitivity) and false positive rate (probability of false alarm = 1 – specificity) as the classification threshold varies.

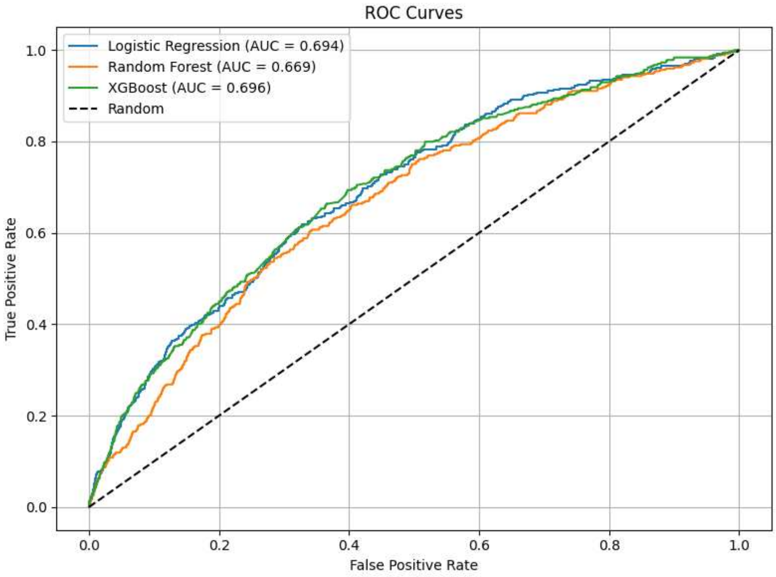

Figure 8: ROC curves for the final best three models on the test set.

Table 10 summarizes the out-of-sample performance metrics for the three final classifiers.

The ROC AUC scores indicate that all models have moderate ability to discriminate between outage and non-outage days. However, precision is extremely low across all models, with values around 0.005–0.008. In operational terms, this means that a system based on binary predictions

Model	**Precision**	**Recall**	**Specificity**	**F1-Score**
Logistic Regression	0.005	0.588	0.694	0.010
Random Forest	0.008	0.295	0.900	0.016
XGBoost	0.006	0.515	0.754	0.011

Table 10: Out-of-sample performance metrics of the three final classifiers on the test set.

would generate many false alarms.

Despite this, the recall scores show that the models are able to detect a substantial share of the actual outages, with logistic regression identifying over 58.8% of all true outage days. XGBoost and random forest perform slightly worse in recall, though still capture meaningful signal. Specificity scores suggest that the models are also capable of correctly identifying a large share of non-outage days. The low F1-scores reflect the severe class imbalance, where the rarity of the positive class naturally penalizes precision-heavy metrics.

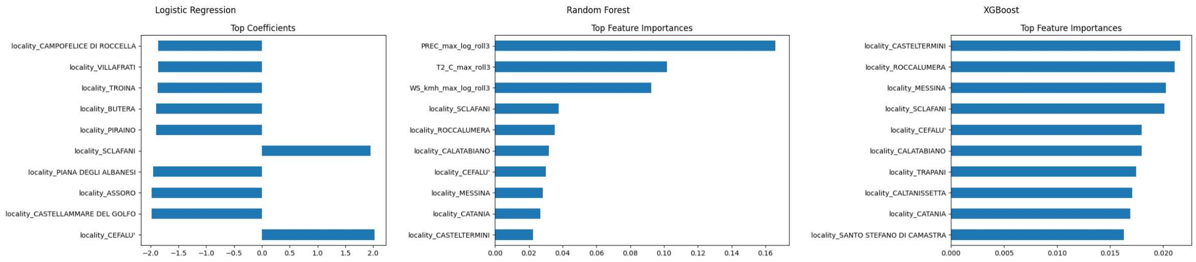

Figure 9: Top-10 most influential features across the three final models: logistic regression (coefficients), random forest (impurity-based importance), and XGBoost (gain-based importance).

To further interpret the inner workings of the final classifiers, Figure 9 compares the ten most influential features identified by each model. In the logistic regression and in the XGBoost models several specific municipalities emerge with large positive coefficients. The prominence of certain municipalities may reflect localized vulnerabilities in grid design, vegetation exposure, or historical maintenance patterns. These unobserved structural and environmental factors are not captured by the weather data but are partially absorbed through fixed effects.

In contrast, random forest prioritize weather covariates more prominently. Rolling features of the maximum temperature and of the log-transformed maximum precipitation, and wind speed rank among the top predictors, highlighting their contribution to outage likelihood.

Figure 10 shows the models' ability to prioritize days by outage risk. The lift curve shows that, at small percentiles, the lift is markedly high. This indicates that true outages are densely concentrated among the days with the highest predicted risk. As the inspection threshold broadens, however, the lift steadily declines, reflecting a transition to performance comparable to random selection. The steeper the initial drop, the stronger the model's ability to identify the most hazardous days at the beginning of the ranking process.

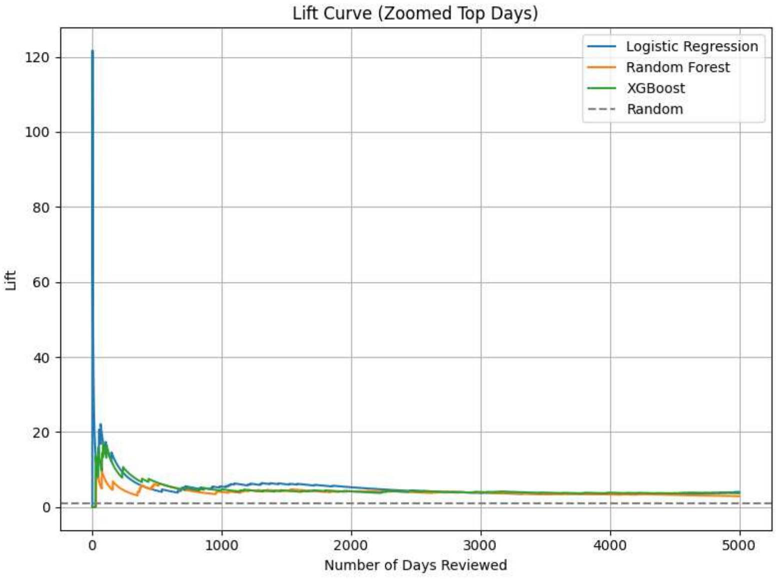

Figure 10: Lift curve showing model performance across the full ranking of predicted risk, to assess prioritization quality relative to random inspection.

This suggests that the models can stratify days by outage risk effectively, enabling grid operators to allocate preventive resources, such as targeted inspections, vegetation clearing, and equipment reinforcement, more efficiently during critical periods. Despite the rarity of outages and the multifactorial nature of grid failures, the models demonstrate an ability to extract actionable information from weather and locality data. This improves anticipatory risk management and supports developing climate-sensitive resilience strategies for the Sicilian grid.

These results highlight a core limitation of using weather and locality data alone: while the models effectively identify the riskiest periods, they struggle to produce highly precise binary predictions. As such, the model outputs should not be interpreted as definitive forecasts but as probabilistic risk scores that can guide prioritization. In practice, these scores can support operational decisions such as ranking days for inspection, pre-positioning response teams, or scheduling preventive maintenance during high-risk windows.

4.2 Stage 2: Multiclass Classification of Outage Severity

The second stage of the modeling pipeline addresses the classification of outage severity, conditional on an event having occurred. This step assigns each outage a severity category based on meteorological conditions and location factors. The resulting predictions inform emergency prioritization and resource allocation strategies.

Several modeling approaches were explored to identify the most effective configuration for this multiclass task. The models are trained only on the subset of the dataset where $\texttt{outage\_occurred}_{it} = 1$, ensuring that the classifier operates only on realized events. The dependent variable is the categorical indicator `severity_class` constructed as described in

Section 3.

$$\mathbb{P}(\texttt{severity\_class}_{it} = c) = \frac{\exp(\beta_{0c}+\beta_{1c}\texttt{T2\_C\_max\_roll3}_{it}+\beta_{2c}\texttt{PREC\_max\_log\_roll3}_{it}+\beta_{3c}\texttt{WS\_kmh\_max\_log\_roll3}_{it}+\alpha_i)}{\sum_{j\in C}\exp\left(\beta_{0j}+\beta_{1j}\texttt{T2\_C\_max\_roll3}_{it}+\beta_{2j}\texttt{PREC\_max\_log\_roll3}_{it}+\beta_{3j}\texttt{WS\_kmh\_max\_log\_roll3}_{it}+\alpha_i\right)} \tag{2}$$

This corresponds to a multinomial logistic regression model, where the softmax function is used to map linear predictors to class probabilities (Taddy, 2019). The index j runs over the set of severity classes $C = \{$*Very Low, Low, Moderate, Severe*$\}$. The term α_i represents locality-specific fixed effects, included as dummy variables, and is constant across classes.

This formulation assumes the *independence of irrelevant alternatives* (IIA), implying that the relative odds between any two severity levels are unaffected by the presence or characteristics of the other alternatives. While this assumption simplifies estimation, it may not fully reflect the ordinal or substitutable nature of adjacent severity classes. Nonetheless, it provides a tractable approach for modeling multiclass severity risk in the presence of unobserved locality heterogeneity.

Here, as in the model at Stage 1, the covariates refer to three-day rolling means of the raw weather variables, constructed to reflect short-term persistence in temperature, precipitation, and wind speed.

Again, the target variable is unbalanced. Of all unplanned outages, 28.9% are classified as *Very Low* severity, corresponding to short-duration events with no ENS. The *Low* severity category, which comprises the majority (37.7%), corresponds to long-duration outages without ENS. The remaining categories, *Moderate* (15.7%) and *Severe* (17.5%), capture outages with positive ENS, differentiated by duration and magnitude. These last two classes are particularly relevant from an operational and economic perspective, but are less frequent and more difficult to predict accurately.

4.2.1 Data Preprocessing

In this phase, the primary data filtering and construction of the target variable were performed during the creation of the `severity_class` indicator, as described in Section 3. Class imbalance was addressed during the model design phase.

The temporal split adopted in Stage 1 was maintained.

4.2.2 Model Evaluation Criteria

The evaluation strategy mirrors that of Stage 1. Model performance is assessed using precision, recall, specificity, and F1-score. For each severity class, specificity is defined as the proportion of non-events (i.e., days belonging to other classes) correctly classified as such in a one-vs-rest framework. This complements recall by indicating how well the model avoids false positives for each class. In operational terms, high recall ensures that most outages of a given severity are detected, while high specificity prevents unnecessary escalation or misallocation of resources to less critical events.

These metrics are particularly important for the *Moderate* and *Severe* classes, which are both less frequent and operationally significant. A trade-off naturally arises: increasing recall for severe outages, for example, may reduce specificity by misclassifying more non-severe days as high-risk. This tension is reflected in the ROC curves, where steeper initial slopes indicate stronger separation between true positives and false positives. Consequently, classifier selection should consider both the ability to detect critical events and the cost of over-alerting.

4.2.3 Model Design

Three classifiers were tested for the multiclass prediction of outage severity: multinomial logistic regression, random forest, and XGBoost.

Although the severity classes have an ordinal structure, the models applied in this stage treat the target as nominal. The labels were encoded in ordinal order (*Very Low* = 0, . . . , *Severe* = 3), which may enable models such as logistic regression to partially reflect the progression in severity. Future work could extend this framework to explicitly ordinal classification techniques, better aligning with the operational significance of severity misclassification.

Several different classifiers were explored prior to identifying the best-performing ones. Localities were encoded as one-hot vectors for the fixed effects models and as cluster labels in the K-means variant, thereby capturing regional heterogeneity either explicitly (fixed effects) or implicitly (through cluster-based groupings).

To address class imbalance, inverse-frequency weighting was applied for logistic regression and random forest models (via `class_weight="balanced"`), while XGBoost employed the `scale_pos_weight` heuristic.

Overall, models leveraging full locality fixed effects consistently outperformed their cluster-based counterparts, better capturing the nuanced spatial patterns in outage severity. Similarly, unweighted versions of each model tended to produce more stable and balanced results than those trained with reweighting. Although class weighting modestly improved recall for rarer classes, it also reduced precision and macro-F1 scores. Given the prioritization objectives of this task, the final comparison focuses on the unweighted versions of logistic regression, random forest, and XGBoost trained with full locality identifiers, as these configurations achieved the best balance between discriminative performance and operational relevance.

4.2.4 Comparison and Interpretation of Final Classifiers

Figure 11 presents the ROC curves for each model and severity class. All models exhibit acceptable discrimination for the *Very Low* and *Low* categories, with performance declining somewhat for the rarer *Moderate* and *Severe* outages. Logistic regression performs slightly better on these rarer classes, suggesting that while weather and locality information helps discriminate mild events, it is not sufficient to robustly distinguish the most impactful ones. This highlights the challenges of anticipating high-ENS events solely through environmental signals.

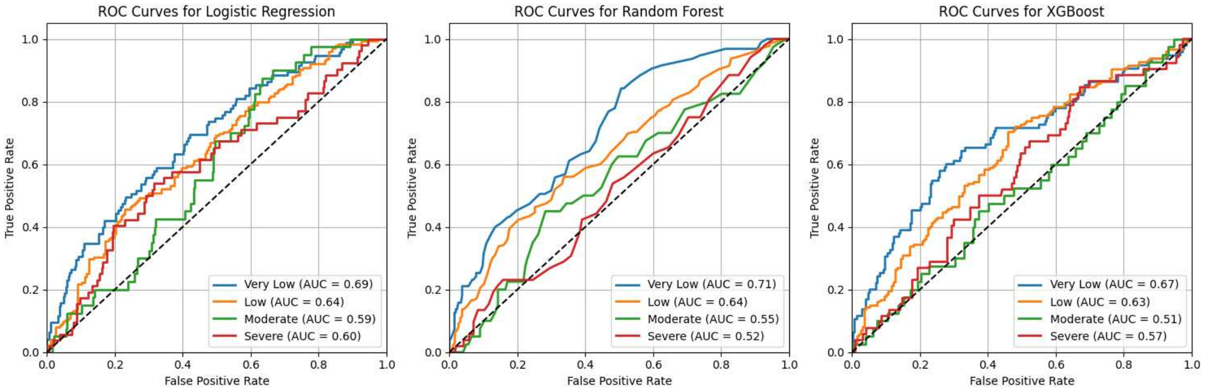

Figure 11: ROC curves for severity classification by model and class.

Table 11 presents class-wise precision, recall, specificity, and F1-scores. All models perform reliably on the dominant categories. Logistic regression slightly outperforms the others in identifying *Severe* events, achieving a recall of 0.31. XGBoost improves performance on the *Moderate* class, though neither ensemble method consistently outperforms the simpler logistic model. Specificity values are relatively stable across models, higher for the rarer classes and lower for the dominant classes, reflecting the models' tendency to correctly exclude infrequent classes more easily than common ones.

Model	**Class**	**Precision**	**Recall**	**Specificity**	**F1-score**
Logistic Regression	Very Low	0.37	0.37	0.77	0.37
	Low	0.58	0.50	0.66	0.53
	Moderate	0.14	0.15	0.88	0.14
	Severe	0.22	0.31	0.82	0.26
Random Forest	Very Low	0.42	0.46	0.77	0.44
	Low	0.56	0.54	0.61	0.55
	Moderate	0.12	0.15	0.87	0.14
	Severe	0.24	0.19	0.90	0.22
XGBoost	Very Low	0.38	0.38	0.77	0.38
	Low	0.56	0.55	0.60	0.56
	Moderate	0.13	0.15	0.87	0.14
	Severe	0.24	0.23	0.88	0.24

Table 11: Class-wise precision, recall, specificity and F1-scores for the final multiclass severity classifiers.

Figure 12 displays the confusion matrices. The most common misclassifications occur between adjacent severity levels, particularly between *Low* and *Very Low*. Logistic regression

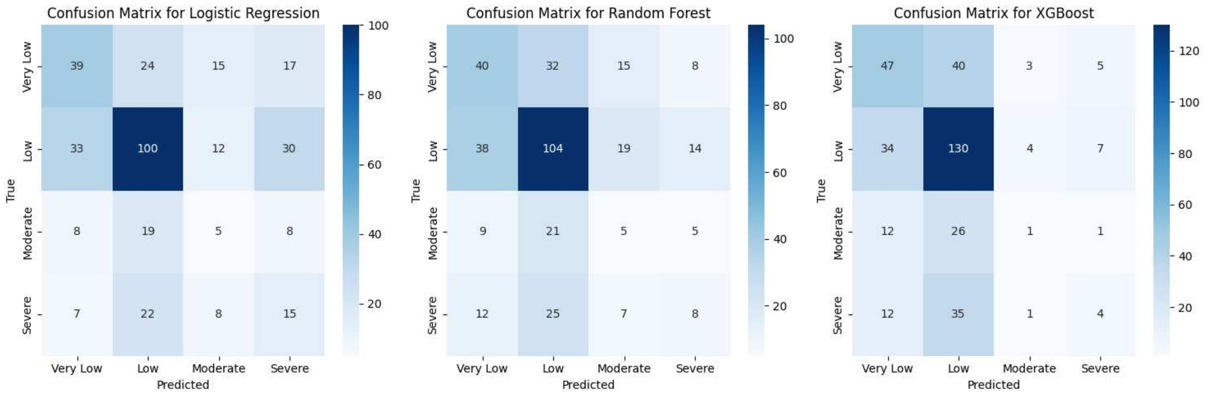

Figure 12: Confusion matrices for each model.

exhibits better sensitivity to *Severe* cases but tends to confuse *Moderate* with both higher and lower categories. Random forest and XGBoost are more reliable in identifying dominant classes but frequently mislabel rarer ones, confirming a tendency to optimize for majority classes unless explicitly reweighted.

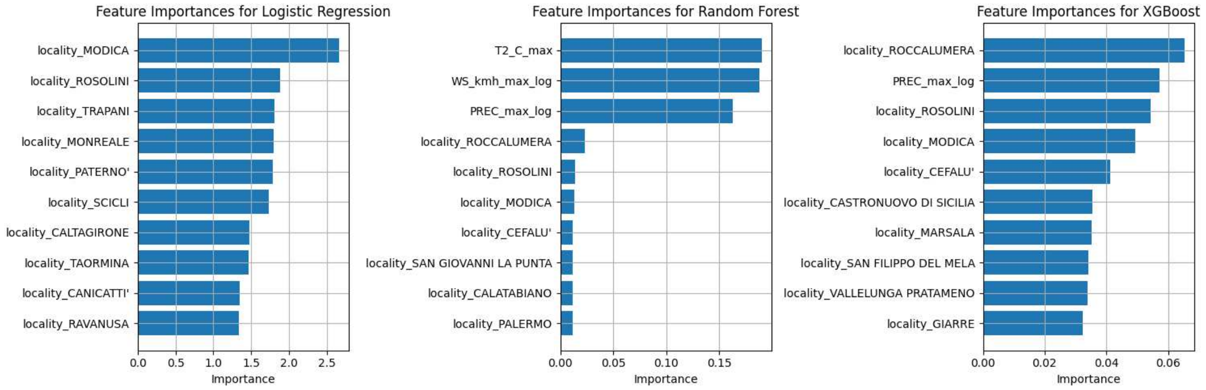

Figure 13: Top 10 feature importances by model.

Figure 13 highlights the most influential predictors. While random forest prioritizes meteorological inputs, logistic regression and XGBoost place greater emphasis on specific municipalities (e.g., Modica, Rosolini), suggesting localized infrastructure vulnerabilities.

Lift curves in Figure 14 assess how well models concentrate true events at the top of the prediction ranking. Logistic regression offers more stable prioritization for *Severe* outages across a broader inspection window. Although XGBoost achieves sharp early gains for the *Moderate* class, its performance quickly converges to baseline, limiting its utility in proactive response scenarios.

In addition to ranking and classification performance, calibration plots were used to evaluate the reliability of predicted probabilities for the *Moderate* and *Severe* classes (Figure 15). For the *Moderate* class, all models tend to underpredict observed event frequencies at higher probability levels. Logistic regression displays more consistent predictions but remains system-

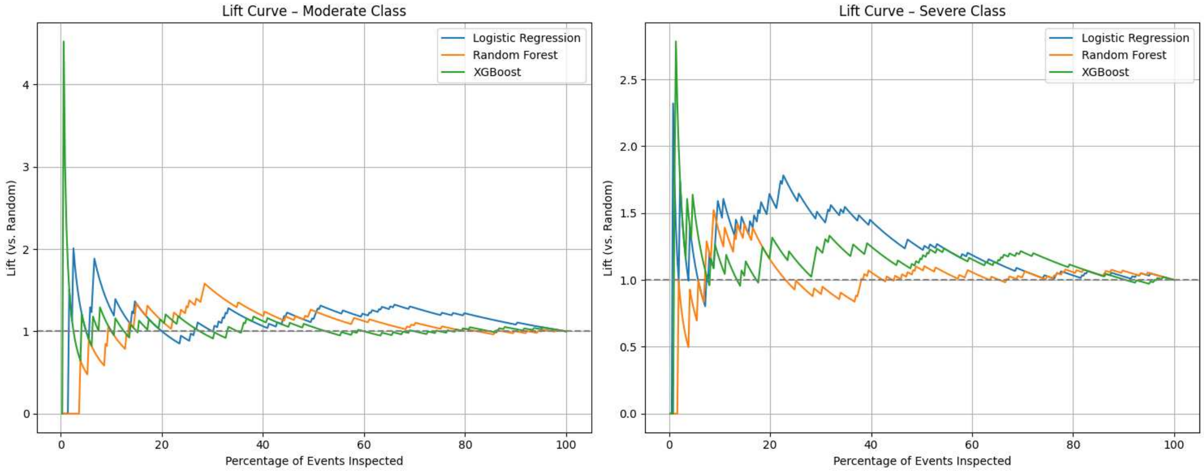

Figure 14: Lift curves for *Moderate* (left) and *Severe* (right) classes.

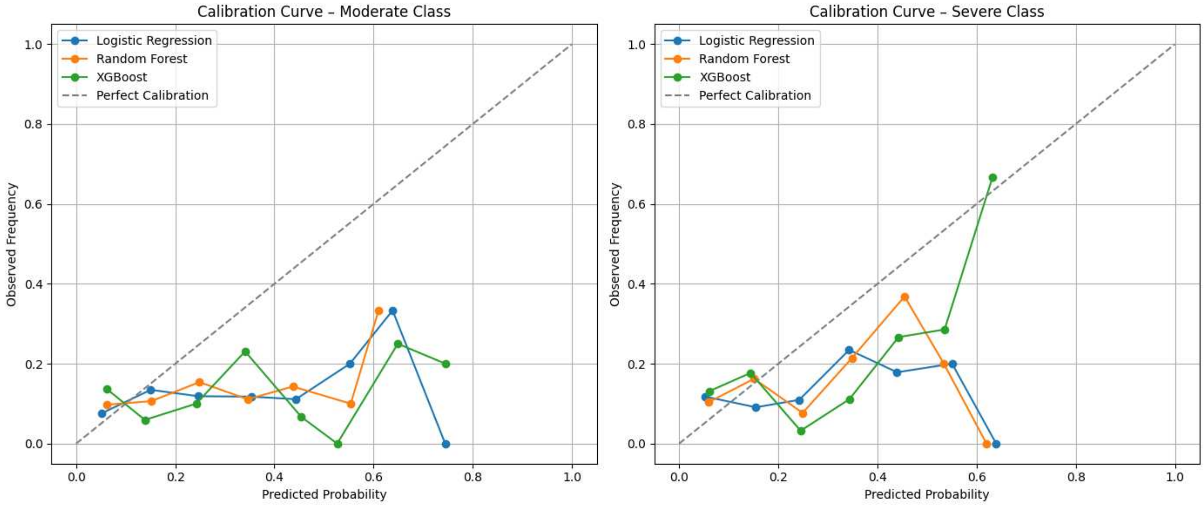

Figure 15: Calibration curves for *Moderate* (left) and *Severe* (right) severity classes.

atically underconfident. For the *Severe* class, XGBoost exhibits overconfident predictions in the upper probability bins, while logistic regression and random forest produce more conservative estimates, remaining closer to the ideal diagonal.

These patterns indicate that although the models are useful for risk stratification, their raw probability outputs should be interpreted cautiously. In operational contexts, post-hoc calibration may be required to support threshold-based decision-making.

Overall, while none of the models excels uniformly, logistic regression provides the most stable and interpretable performance, particularly for high-severity events where early intervention is critical. Despite limited precision due to data sparsity, the model's ability to flag high-risk cases underscores its value in operational planning for climate-sensitive grid resilience.

4.3 Stage 3: Regression-Based Forecast of Economic Impact

The final stage of the modeling pipeline focuses on estimating the economic consequences of power outages by predicting the *Energy Not Supplied* (ENS), measured in MWh. ENS represents the magnitude of service disruption and serves as the primary proxy for quantifying outage impact in economic terms, consistent with Italian regulatory frameworks for dispatching cost allocation and service quality evaluation (Terna S.p.A., 2024a, 2025a). This estimation supports ex-ante prioritization of high-impact outages, guiding resilience investment and informing risk-based planning.

The regression models developed in this stage are trained exclusively on the subset of unplanned outages with strictly positive ENS values. This modeling restriction is justified on both statistical and operational grounds. First, the ENS distribution is zero-inflated, with over 80% of observations taking a value of zero, corresponding to faults without a direct impact on the final users. Second, under the regulatory criteria enforced by Terna, only outages with positive ENS result in economic compensation or penalties, making them the relevant subset for financial forecasting (Terna S.p.A., 2024a). By conditioning on realized impact, the models target the question: *"Given that an outage caused energy loss, how severe was it in economic terms?"*

To ensure robust estimation while maintaining interpretability, the modeling strategy combines ordinary least squares (OLS) regression with locality-level fixed effects and two regularized methods: Ridge and Lasso. These approaches control for overfitting introduced by high-dimensional fixed effects and allow for the identification of key meteorological and spatial drivers of economic loss. The target variable, $\log(\texttt{ENS})$, is used to reduce skewness and enable semi-elastic interpretations of coefficients. Predictions are exponentiated back to the original scale and aggregated to compute forecasted annual ENS values under observed and synthetic climate scenarios.

To translate predicted ENS into economic terms, the analysis adopts the incentive-based valuation defined by ARERA, which imposes a financial penalty of €27,000 for each megawatt-hour of energy not supplied above an annual threshold of 763 MWh. This benchmark, in place for 2024, defines the boundary between acceptable and sanctionable performance. Although held constant in this simulation, it is reasonable to expect that future regulatory targets may become more stringent, as historical trends show progressive tightening of service quality criteria by ARERA (Terna S.p.A., 2024b). This assumption reinforces the strategic importance of scenario-based risk modeling as a tool for long-term system planning.

The framework adopted here also aligns with Terna's broader strategy to develop a sustainable and resilient national transmission network, as stated in its 2023 development plan. This includes investments aimed at improving climate adaptation, integrating renewable energy sources, and reducing the frequency and severity of faults. The model supports SDGs 7, 9, and 13 by providing a data-driven mechanism to evaluate the economic and operational implications of climate variability, enabling informed resilience policy and infrastructure prioritization.

Formally, the estimated regulatory cost exposure is computed as:

$$\text{Regulatory Impact}_{\text{EUR}} = \max(0, \widehat{\mathtt{ENS}}_{\text{annual}} - 763) \times 27{,}000 \tag{3}$$

where $\widehat{\mathtt{ENS}}_{\text{annual}}$ denotes the predicted total ENS for a given year. This formulation ensures that only the exceedance beyond the regulatory threshold contributes to the cost, consistent with the incentive structure currently applied to transmission system operators.

4.3.1 Data Preprocessing

The regression model for estimating ENS is trained on the subset of the dataset where unplanned outages occurred and resulted in a strictly positive ENS value. This selection removes events associated that caused no actual service disruption. These events, while operationally relevant, do not contribute to economic losses and would otherwise dilute the predictive signal with structural zeros. After filtering, approximately 16% of outage events remain, representing the cases of direct economic relevance.

The predictor variables include the rolling averages of three days of meteorological covariates. This smoothing approach is consistent with prior modeling stages and helps to capture short-term weather persistence relevant to damage mechanisms.

The final modeling dataset contains 394 observations and balances parsimony with spatial specificity through the inclusion of locality-level fixed effects.

For regularized regressions, locality indicators were one-hot encoded and combined with weather features to form the design matrix. This preprocessing pipeline ensures comparability across model types while preserving the interpretability of coefficients and model outputs.

4.3.2 Model Evaluation Criteria

The performance of the regression models predicting the logarithm of Energy Not Supplied (ENS) was evaluated using both in-sample goodness-of-fit and cross-validated predictive accuracy. The primary in-sample diagnostic is the coefficient of determination (R^2), which quantifies the proportion of variance in $\log(\mathtt{ENS})$ explained by the covariates.

Given the risk of overfitting in the presence of high-dimensional categorical variables such as locality dummies, regularized regression methods were employed.

In addition to explanatory and predictive metrics, model reliability was assessed through the root mean squared error (RMSE), which serves as a basis for constructing prediction intervals. These RMSE values were also used in constructing 95% prediction confidence intervals around scenario forecasts, by back-transforming log errors to the original scale.

Overall, evaluation criteria were selected to balance interpretability, predictive power, and robustness to overfitting, with particular emphasis on the stability of weather-related coefficients and the operational relevance of spatial heterogeneity in outage severity.

4.3.3 Model Design

The baseline specification is an OLS model that includes meteorological covariates as well as locality-level fixed effects. These fixed effects absorb unobserved spatial heterogeneity in infrastructure exposure, vegetation proximity, and protection schemes, all of which can influence outage severity. The estimated equation takes the form:

$$\log(\texttt{ENS}_{it}) = \beta_0 + \beta_1 \texttt{T2\_C\_max\_roll3}_{it} + \beta_2 \texttt{PREC\_max\_log\_roll3}_{it} + \beta_3 \texttt{WS\_kmh\_max\_log\_roll3}_{it} + \alpha_i + v_{it} \tag{4}$$

where α_i denotes fixed effects for each locality and v_{it} is the idiosyncratic error term. This model specification achieves an R^2 of 0.354 and an adjusted R^2 of 0.119, indicating moderate explanatory power but suggesting potential overfitting. The high number of fixed effects relative to the sample size, combined with a low F-statistic ($F = 1.51$, $p = 0.007$), points to multicollinearity and inflated variance in coefficient estimates.

To address these limitations, Ridge and Lasso regularized regression models, were introduced. Ridge regression adds an L_2 penalty that shrinks coefficients toward zero, reducing sensitivity to collinear predictors. Lasso regression incorporates an L_1 penalty that also enables automatic variable selection by setting some coefficients exactly to zero. Both models are trained on standardized predictors and tuned via five-fold cross-validation.

The Ridge model retained all predictors, applying global shrinkage to reduce coefficient variance. In contrast, the Lasso model performed both shrinkage and variable selection across all predictors by setting a subset of coefficients exactly to zero.

The resulting models strike a balance between interpretability and predictive performance, providing robust estimates of outage impact suitable for downstream cost forecasting.

4.3.4 Model Interpretation

The regression results suggest that meteorological factors and locality-specific effects explain a limited but non-negligible portion of the variability in outage severity, measured as $\log(\texttt{ENS})$. As shown in Table 12, the Ridge model achieves an R^2 of 0.203 with a RMSE of 1.808, while the Lasso model yields an R^2 of 0.360 and an RMSE of 1.621. These metrics are computed on the full subset, consistent with the explanatory and simulation-oriented objectives of this modeling stage. The models are not intended for predictive deployment, but rather to support scenario analysis and isolate structural drivers of severity across space and weather conditions.

While the overall fit remains moderate, the Lasso model proves valuable in isolating key explanatory factors. Among the continuous predictors, the log-transformed rolling maximum wind speed variable ($\log(\texttt{WS\_kmh\_max\_roll3})$) is the only weather feature retained. Its estimated coefficient of approximately 0.244 suggests that a 1% increase in the three-day rolling maximum wind speed is associated with a 0.244% increase in expected ENS, conditional on an outage.

Table 12: Performance of Ridge and Lasso regressions.

Model	R^2	RMSE
Ridge	0.203	1.808
Lasso	0.360	1.621

This result reinforces the importance of sustained wind extremes in driving the severity of service disruptions.

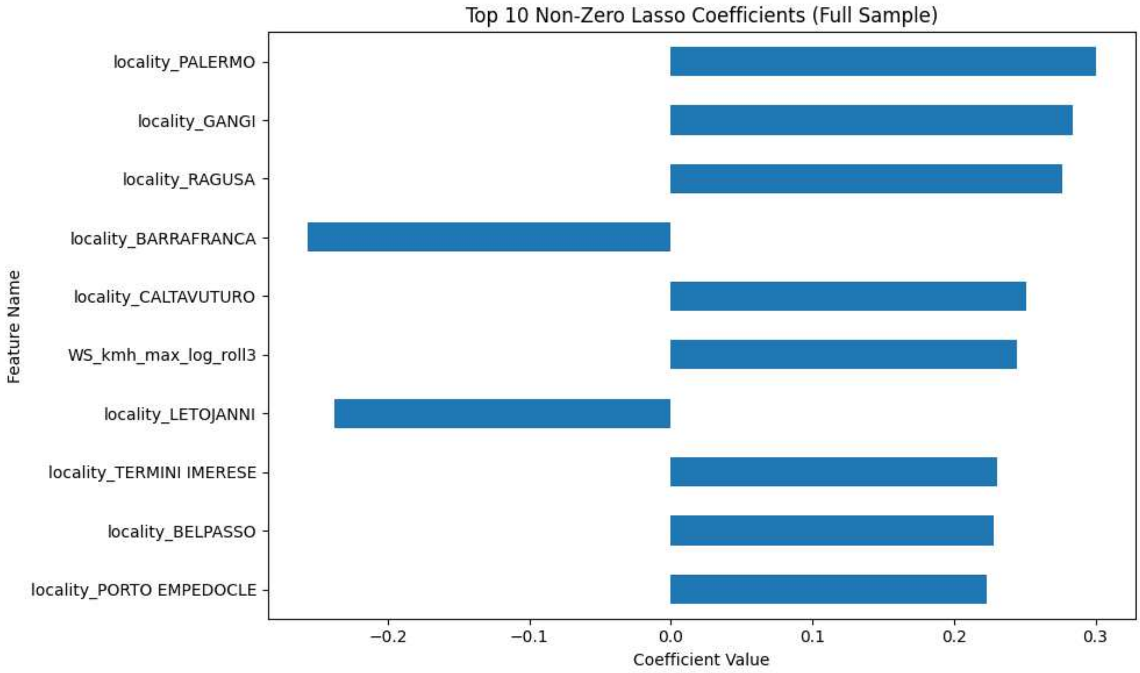

Figure 16: Top 10 coefficients selected by the Lasso regression.

Figure 16 displays the top coefficients selected by Lasso. Alongside wind, several localities are associated with disproportionate severity, these patterns suggest location-specific vulnerability possibly linked to transmission bottlenecks, terrain effects, or exposure. Conversely, negative coefficients for municipalities may reflect grid redundancy, fault-tolerant design, or effective operational response.

Overall, the regression models serve as interpretable simulators rather than precise forecasters. They are particularly useful for generating ENS-based cost projections under altered climate scenarios and for identifying areas of the grid with persistently high or low outage severity exposure.

The limited model fit is expected given that actual ENS outcomes depend on real-time dispatch configurations, fault topology, demand loads, and protection system responses, variables that are not included in the present data. Nonetheless, the identified relationships provide a first-order approximation of risk intensity and support economic scenario modeling under evolving climate conditions.

4.3.5 Scenario-Based Forecasting and Economic Projections

The final stage of the analysis quantifies the potential regulatory costs of power outages under intensified wind scenarios, using predictive models trained on historical outage data. The focus is on the variable `WS_kmh_max_log_roll3`, a three-day rolling log-transformed maximum wind speed. For each scenario (+10%, +20%, and +50% wind increase), the corresponding wind intensification was first applied in the original (non-log) scale, then re-log-transformed and standardized to match the model's input space.

Both Ridge and Lasso models were used to forecast the natural logarithm of ENS, which was then converted back into MWh using the exponential function. Forecasts were aggregated over the 2021–2023 period and averaged to obtain annual ENS estimates. Regulatory costs were computed according to ARERA guidelines: only the share of annual ENS exceeding 763 MWh was monetized at a penalty rate of €27,000 per MWh.

Table 13 reports the resulting annual economic exposures in thousands of euros (k€), along with 95% simulation intervals.

Table 13: Annual forecasted regulatory economic impact under wind intensification scenarios (k€)

Scenario	Ridge Forecast (k€)	Ridge Range (k€)	Lasso Forecast (k€)	Lasso Range (k€)
Baseline	0	0–218,549	0	0–550,848
+10% Wind	0	0–232,305	0	0–658,693
+20% Wind	0	0–245,552	0	0–774,827
+50% Wind	0	0–282,784	1.654	0–1,171,689

Figure 17 shows the forecasted annual ENS under each scenario. The y-axis is log-scaled to visualize the non-linear escalation in risk, particularly for the Lasso model. Ridge predictions remain below the threshold across all scenarios, indicating low regulatory exposure. In contrast, the Lasso model projects a marked increase in unserved energy as wind intensity grows, with point forecasts surpassing the 763 MWh threshold in the +50% scenario.

These scenario-based forecasts provide quantitative evidence for justifying investments in grid hardening and preventive maintenance, with the potential to avert regulatory costs that could exceed €1 million for the TSO, under extreme wind events intensification.

Confidence intervals widen significantly with intensification, reflecting increased model uncertainty and the exponential amplification of log-residual variability. According to Terna's provisional 2023 operational data, regulated ENS stood at 507 MWh, safely below the ARERA threshold. This confirms that threshold violations remain rare and are generally indicative of exceptional stress.

The simulations in this thesis therefore represent credible yet extreme stress-test scenarios,

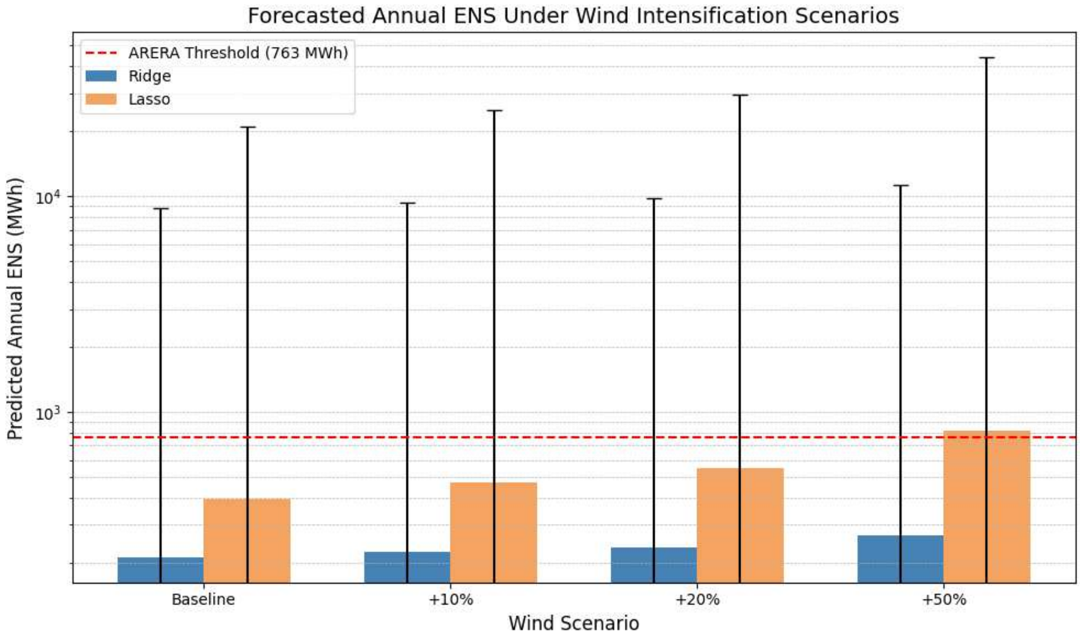

Figure 17: Forecasted annual ENS under wind intensification scenarios (log scale). The red dashed line indicates ARERA's regulatory cap of 763 MWh.

illustrating the potential financial exposure under future climate intensification. These insights reinforce the importance of integrating predictive scenario modeling into grid planning and climate resilience strategies.

5 Discussion

The results of this study highlight the partial yet actionable value of weather-based predictive models for assessing power outage risk and impact in Sicily's infrastructure-sparse environment. The structured pipeline captures the operational layers of the power system and reveals key constraints in data quality, model granularity, and causal inference.

The first-stage classifiers demonstrated that meteorological extremes and spatial identifiers contain predictive signal, particularly when leveraged through locality fixed effects. Logistic regression outperformed more complex models in terms of interpretability and lift performance. However, the persistent lack of precision across all classifiers reflects the inherent limitations of relying solely on environmental exposure to infer outage occurrence. Key causal mechanisms, such as vegetation proximity, asset aging, grid topology, and protection scheme behavior, remain unobserved. Fixed effects partially compensate for these omissions, but at the cost of generalizability beyond the training localities. Despite these limitations, the models' ability to rank days by outage risk represents a significant step forward in anticipatory risk management, providing a practical early-warning framework for resilience planning.

In the second stage, severity classification introduced an important operational layer by distinguishing between low- and high-consequence events. The model's ability to identify *Severe* events with reasonable recall, despite class imbalance, supports its use in event triage and prioritization of field inspections and protective interventions. However, calibration plots revealed systematic underconfidence, and confusion matrices suggest that severity levels are difficult to distinguish based on weather inputs alone. The models do not fully capture the behavior of protective systems (e.g., Telescatto triggers and load shedding), which can decouple meteorological stress from economic impact. Additionally, severity misclassifications between adjacent categories point to both limitations in the categorical target design and to the likely role of internal system dynamics that are not reflected in the feature set. Nonetheless, these severity models contribute valuable insights for developing risk-informed operational protocols, strengthening the preparedness of grid operators under extreme weather conditions.

The third-stage regression analysis aimed to quantify ENS for financially significant events. Consistent with operational insights from Terna, wind speed emerged as the only robust meteorological predictor. However, the models only explained a modest amount of variance. The zero-inflated nature of ENS and the small number of high-loss events limited the regression's statistical power. This highlights a structural challenge: economic impacts depend not only on the occurrence of faults, but also on how they propagate through real-time dispatch decisions, demand levels, and redundancy configurations. The wide prediction intervals in scenario-based simulations reflect this residual uncertainty, which limits the precision of the models as policy tools. However, the simulations highlight the thesis's practical contribution as a quantitative stress-testing framework for climate-exacerbated risks. This framework helps decision-makers identify financially critical thresholds and assess the potential scale of regulatory exposure.

Several limitations warrant further attention. First, the static nature of the locality fixed effects prevents insights from being generalized or extrapolated to new or changing infrastructure configurations. Second, the daily aggregation level masks sub-daily dynamics critical to grid instability, such as diurnal load curves, fault propagation sequences, and auto-recovery mechanisms. Third, the absence of asset-level and topological features constrains causal interpretability. Fourth, weather data from reanalysis, while spatially consistent, may smooth out localized extremes, underestimating true exposure in some cases. Finally, the modeling framework remains observational; it does not test structural causal relationships between weather and outages, limiting its applicability for regulatory accountability or climate attribution studies.

Future research should pursue several directions to address these gaps. Integrating asset-level data, such as maintenance logs, vegetation invasion reports, and real-time measurements, could substantially improve both classification and regression accuracy. Temporal disaggregation to hourly or sub-hourly resolution would allow for modeling of fast-evolving disturbances and automated defense responses. Causal inference frameworks, including instrumental variable approaches or structural hazard models, could improve interpretability and support regulatory attribution. Furthermore, unsupervised learning techniques, such as clustering or anomaly detection, may uncover latent spatial or temporal patterns of grid fragility not visible in supervised settings. Deep learning architectures, particularly recurrent networks for temporal modeling or graph neural networks for topological learning, could be explored to capture complex, nonlinear interactions across space and time, though their operational interpretability remains a key concern. An especially promising avenue involves applying image-based deep learning to visual inspections captured by Terna's helicopter-mounted cameras. Finally, using convolutional neural networks (CNNs) to detect physical anomalies, such as corrosion, conductor sag, or vegetation encroachment, could automate infrastructure condition assessment and link visual features directly to outage risk.

Overall, this study demonstrates that, even under strong data constraints, interpretable supervised learning models can provide actionable insights for anticipatory risk management in climate-sensitive grid operations. However, realizing the full potential of predictive resilience analytics will require more granular, transparent, and real-time infrastructure data.

6 Conclusion

This thesis proposed a structured, modular framework to predict and assess the vulnerability of the Sicilian power grid to extreme weather events, integrating supervised learning with economic consequence estimation. Motivated by the growing operational challenges posed by climate change and the criticality of regional infrastructure, the study addresses a gap in the literature concerning the integration of meteorological stressors and power system disruptions in the Italian context.

The three-stage modeling pipeline, comprising outage occurrence prediction, conditional severity classification, and ENS-based economic forecasting, demonstrates that interpretable, data-driven methods can support anticipatory risk management even under severe data limitations. Logistic regression consistently provided stable and actionable results in both binary and multiclass settings, outperforming more complex alternatives in calibration and lift-based prioritization. The final regression models identified maximum wind speed as the dominant meteorological driver of economic losses, validating operational experience and highlighting exposure escalation under future climate scenarios.

Despite these contributions, the study remains constrained by the absence of real-time grid condition data, sub-daily resolution, and asset-level metadata. The explanatory power of meteorological predictors is necessarily limited in the face of complex infrastructural and operational dynamics. Nevertheless, the findings provide a structured foundation for integrating environmental risk into grid resilience planning and suggest that even coarse-grained environmental data can give meaningful operational insights when interpreted through well-designed statistical models.

This work contributes to the growing body of research on climate-resilient power systems by offering an interpretable, spatially explicit, and operationally relevant approach to outage prediction and impact modeling. It complements existing studies by focusing on a Mediterranean island grid with unique exposure characteristics, thus extending the applicability of resilience analytics to underrepresented geographies.

Looking forward, the integration of more granular grid monitoring data, unsupervised pattern discovery, and image-based diagnostics using aerial surveillance imagery may unlock new layers of predictive performance and causal understanding. As climate extremes intensify, such innovations will be essential to support grid modernization, targeted investment, and sustainable energy system design.

Acknowledgments

I am deeply grateful to Dott. Renato Licciardello and his team at Electric Power S.r.l. for the opportunity to visit their facilities and for their invaluable support in arranging a technical meeting with Terna S.p.A. I extend special thanks to Ing. Giuseppe Chiarenza for hosting me at the Terna substation in Misterbianco, for his openness in addressing my inquiries, and for providing crucial documentation that significantly enhanced the depth of this study.

My sincere appreciation also goes to Dott. Carlo Barozzi for granting access to the outage dataset and for his guidance during the complex phases of data integration and preprocessing. The successful completion of this work would not have been possible without the generous contributions of time and expertise from these professionals, which greatly enriched both the empirical and contextual dimensions of the thesis.

I would also like to thank my supervisor, Prof. Pedro Afonso Fernandes, for his invaluable support and mentorship throughout the writing of this thesis. I am equally thankful to Prof. Filipa Reis and the entire faculty of the Master's in Business Analytics at Católica Lisbon for their academic guidance and for fostering an enriching learning environment. This intensive yet rewarding journey has equipped me with essential analytical skills and has been a period of profound personal growth.

With the completion of this thesis, I mark the end of my formal academic journey and the beginning of a new chapter—one that I hope will be filled with continued learning, meaningful experiences, and new opportunities.

References

ANSA. Aeolian islands cut off by bad weather, floods at catania, 2025. URL https://www.ansa.it/english/news/2025/01/17/aeolian-islands-cut-off-by-bad-weather-floods-at-catania_5f75656f-a6cf-45bc-9d15-9794d85534ca.html. Accessed: 20-Feb-2025.

E. Avolio, G. Castorina, R. C. Torcasio, and S. Federico. A multi hazard extreme weather event in southern italy: Assessment and sensitivity tests of the wrf model. *Atmospheric Research*, 315, 4 2025. ISSN 01698095. doi: 10.1016/j.atmosres.2024.107827.

R. Bonanno, M. Lacavalla, and S. Sperati. A new high-resolution meteorological reanalysis italian dataset: Merida. *Quarterly Journal of the Royal Meteorological Society*, 145:1756–1779, 4 2019. ISSN 1477870X. doi: 10.1002/qj.3530.

K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. W. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barrett, G. Blanco, W. W. Cheung, S. Connors, F. Denton, A. Diongue-Niang, D. Dodman, M. Garschagen, O. Geden, B. Hayward, C. Jones, F. Jotzo, T. Krug, R. Lasco, Y.-Y. Lee, V. Masson-Delmotte, M. Meinshausen, K. Mintenbeck, A. Mokssit, F. E. Otto, M. Pathak, A. Pirani, E. Poloczanska, H.-O. Pörtner, A. Revi, D. C. Roberts, J. Roy, A. C. Ruane, J. Skea, P. R. Shukla, R. Slade, A. Slangen, Y. Sokona, A. A. Sörensson, M. Tignor, D. van Vuuren, Y.-M. Wei, H. Winkler, P. Zhai, Z. Zommers, J.-C. Hourcade, F. X. Johnson, S. Pachauri, N. P. Simpson, C. Singh, A. Thomas, E. Totin, A. Alegría, K. Armour, B. Bednar-Friedl, K. Blok, G. Cissé, F. Dentener, S. Eriksen, E. Fischer, G. Garner, C. Guivarch, M. Haasnoot, G. Hansen, M. Hauser, E. Hawkins, T. Hermans, R. Kopp, N. Leprince-Ringuet, J. Lewis, D. Ley, C. Ludden, L. Niamir, Z. Nicholls, S. Some, S. Szopa, B. Trewin, K.-I. van der Wijst, G. Winter, M. Witting, A. Birt, and M. Ha. Ipcc, 2023: Climate change 2023: Synthesis report. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change [core writing team, h. lee and j. romero (eds.)]. ipcc, geneva, switzerland. Technical report, Intergovernmental Panel on Climate Change, 7 2023. URL https://www.ipcc.ch/report/ar6/syr/.

R. J. Campbell. Crs report for congress weather-related power outages and electric system resiliency specialist in energy policy weather-related power outages and electric system resiliency congressional research service weather-related power outages and electric system resiliency congressional research service. Technical report, 2012. URL www.crs.gov.

Copernicus Data Space Ecosystem. Sicily during heatwave, 2023. URL https://dataspace.copernicus.eu/gallery/2023-7-20-sicily-during-heatwave. Accessed: February 20, 2025.

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 2023*, pages 671–677. Institute of Electrical and Electronics Engineers Inc., 2023. ISBN 9798350348378. doi: 10.1109/ICCEP57914.2023.10247465.

P. Di Gloria, S. Paradiso, M. Pede, V. M. E. Sorrentino, C. Vergine, F. Massaro, A. Vasile, and G. Zizzo. On the impact of renewable generation on the sicilian power system in near-future scenarios: A case study. *Energies*, 17, 7 2024. ISSN 19961073. doi: 10.3390/en17133352.

D. Dwivedi, P. K. Yemula, and M. Pal. Evaluating the planning and operational resilience of electrical distribution systems with distributed energy resources using complex network theory. *Renewable Energy Focus*, 46:156–169, 9 2023. ISSN 1755-0084. doi: 10.1016/J.REF.2023.06.007.

G. Fioravanti, E. Piervitali, F. Desiato, W. Perconti, and P. I. Fraschetti. Variazioni e tendenze degli estremi di temperatura e precipitazione in italia. Technical report, ISPRA, Stato dell'Ambiente 37/2013, 7 2013.

A. Forestieri, E. Arnone, S. Blenkinsop, A. Candela, H. Fowler, and L. V. Noto. The impact of climate change on extreme precipitation in sicily, italy. *Hydrological Processes*, 32:332–348, 1 2018. ISSN 10991085. doi: 10.1002/hyp.11421.

A. M. Fowler and K. J. Hennessy. Potential impacts of global warming on the frequency and magnitude of heavy precipitation. Technical report, 1995.

F. Giorgi. Climate change hot-spots. *Geophysical Research Letters*, 33, 4 2006. ISSN 00948276. doi: 10.1029/2006GL025734. Mediterranean is themost prominent hot-spot emerging from the RCCI. ”The greatest contribution to the MED RCCI is given by a large decrease in mean precipitation and an increase in precipitation variability during the dry (warm) season.”.

IEEE Std 1366. Ieee guide for electric power distribution reliability indices. *IEEE Std 1366-2012 (Revision of IEEE Std 1366-2003)*, pages 1–43, 2012. doi: 10.1109/IEEESTD.2012.6209381.

F. H. Jufri, V. Widiputra, and J. Jung. State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies, 4 2019. ISSN 03062619.

A. Kenward and U. Raja. Blackout: Extreme weather, climate change and power outages. Technical report, Climate Central, 2014. URL `www.climatecentral.org`.

K.-T. Ma, Y. Luo, T. Kwan, and Y. Wu. Environmental loads and vessel motions. *Mooring System Engineering for Offshore Structures*, pages 41–62, 2019. doi: 10.1016/B978-0-12-818551-3.00003-X.

A. Moschetti and C. Barozzi. A data preparation pipeline for the analysis of the influence of weather events on power grid failures. Master's thesis, Politecnico di Milano, 2024.

M. Panteli and P. Mancarella. Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies, 6 2015. ISSN 03787796.

Protezione Civile Sicilia. Emergenza ondate di calore e blackout elettrici nel catanese: Le raccomandazioni e il lavoro di monitoraggio della protezione civile siciliana, 2023. URL https://www.protezionecivilesicilia.it/it/12263-23-luglio-2023--emergenza-ondate-di-calore-e-blackout-elettrici-nel-catanes
asp. Accessed: 3 March 2025.

G. Scicchitano, G. Scardino, C. Monaco, A. Piscitelli, M. Milella, F. D. Giosa, and G. Mastronuzzi. Comparing impact effects of common storms and medicanes along the coast of south-eastern sicily. *Marine Geology*, 439, 9 2021. ISSN 00253227. doi: 10.1016/j.margeo.2021.106556.

L. J. Slater, B. Anderson, M. Buechel, S. Dadson, S. Han, S. Harrigan, T. Kelder, K. Kowal, T. Lees, T. Matthews, C. Murphy, and R. L. Wilby. Nonstationary weather and water extremes: A review of methods for their detection, attribution, and management, 7 2021. ISSN 16077938.

A. Stankovski, B. Gjorgiev, L. Locher, and G. Sansavini. Power blackouts in europe: Analyses, key insights, and recommendations from empirical evidence. *Joule*, 7:2468–2484, 11 2023. ISSN 25424351. doi: 10.1016/j.joule.2023.09.005.

M. Taddy. *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*. McGraw-Hill Education, 2019.

Terna S.p.A. Guida tecnica codifica allegato a54 classificazione e registrazione delle interruzioni degli utenti direttamente e indirettamente connessi alla rtn. Technical report, Terna S.p.A., 2017. Technical Guide A54: Classification and Recording of Interruptions for Users Directly and Indirectly Connected to the National Transmission Grid (RTN).

Terna S.p.A. IO461DC – Regole operative per l'attribuzione dei costi dei servizi di Dispacciamento, 2022. Operational Rules for the Allocation of Dispatching Service Costs. Internal Instruction, Revision 01, November 17, 2022.

Terna S.p.A. Piano di sviluppo 2023 - stato del sistema elettrico. Technical report, Terna S.p.A., Rome, Italy, 2023. URL https://download.terna.it/terna/Piano_Sviluppo_2023_Stato_Sistema_Elettrico.pdf. 2023 Development Plan – Status of the Electrical System.

Terna S.p.A. a03IO503DC – Regole operative per l'esercizio dei collegamenti Sicilia-Calabria e dell'anello calabro, 2024a. Operating Rules for the Sicily–Calabria Interconnections and the Calabrian Ring. Internal Instruction, Revision 02, January 15, 2024.

Terna S.p.A. Una rete sostenibile, 2024b. URL `https://www.terna.it/it/sostenibilita/rete-sostenibile`. Accessed: 18-May-2025.

Terna S.p.A. Qualità del servizio – Meccanismo di regolazione del servizio di trasmissione (2024-2027), 2025a. Transmission Service Regulation Mechanism for 2024–2027. Internal Regulation Document, Revision February 2025.

Terna S.p.A. RdE Telescatto Area Occidentale Sicilia, 2025b. Western Sicily Area Telescatto (Remote Trip) Specification. Internal Specification, Revision 00, February 5, 2025.

F. Yang, M. Koukoula, S. Emmanouil, D. Cerrai, and E. N. Anagnostou. Assessing the power grid vulnerability to extreme weather events based on long-term atmospheric reanalysis. *Stochastic Environmental Research and Risk Assessment*, 37:4291–4306, 11 2023. ISSN 14363259. doi: 10.1007/s00477-023-02508-y.

A Appendix A: Description of Dataset Variables (Grid and Weather)

This appendix documents the variables used in the empirical analysis, drawn from two primary sources: (i) outage records and grid metadata provided by Terna SpA, and (ii) daily weather data at the locality level from MERIDA atmospheric reanalysis. All data is structured as a panel with daily observations per locality.

A.1 Geographical Metadata

- `lat_x`, `lon_x`, `alt`: Latitude, longitude, and altitude of the user node.
- `località_x`, `provincia_x`, `regione`, `area`: Locality, province, region, and broader macro-area.
- `data`: Date.

A.2 Power Grid and Outage Variables

Event Metadata

- `n_disalimentazione`: Annual progressive identifier for each outage event.
- `durata_disalimentazione`: Outage duration (mm:ss).
- `n_incidente_rilevante`: Indicator for relevant incidents (ENS net > 250 MWh).

Cause and Classification

- `tipo_interruzione`: Categorical description of the outage type (e.g., short, long, or transient).
- `codice_causa_AEEG_1_livello`, `codice_causa_AEEG_2_livello`: Regulatory codes classifying outage causes at two levels.
- `NF_NR_SD`: Type of disconnection:
  - `NF`: Load disconnection.
  - `NR`: Generation disconnection.
  - `SD`: Disconnection due to defense systems.

Energy Metrics

- `ENS_ENR_lorda`: Gross energy not supplied (MWh).
- `ENS_ENR_netta`: Net energy not supplied (MWh), adjusted for temporary restorations or alternative feeds.

Network and Infrastructure Metadata

- `elemento_rete_origine`, `titolare_rete_origine`: Originating network element and its owner.
- `tensione_rete_origine`, `tensione_impianto`: Voltage levels (kV) of origin and user site.
- `codice_utente`: Identifier of the affected user site.
- `configurazione_rete`: Network configuration (e.g., radial, meshed).
- `titolare`, `denominazione_impianto`: Ownership and name of the impacted facility.
- `direttamente_connesso`, `altri_collegamenti`, `funzione`: Flags indicating direct connection and functional role in the grid.

A.3 Weather Variables (MERIDA)

Daily Meteorological Indicators

- `T2_C_mean`: Daily mean temperature at 2 meters height (°C).
- `T2_C_max`: Daily maximum temperature (°C).
- `PREC_mean`: Daily mean precipitation (mm/h).
- `PREC_max`: Daily maximum precipitation (mm/h).
- `WS_kmh_mean`: Daily mean wind speed (km/h).
- `WS_kmh_max`: Daily maximum wind speed (km/h).
- `extreme_heat_flag`: Binary indicator of heat extreme.
- `extreme_rain_flag`: Binary indicator of rainfall extreme.
- `extreme_wind_flag`: Binary indicator of wind extreme).

Temporal Resolution and Matching

- Weather variables are aggregated at the daily level and matched to outages via locality and date, allowing for modeling of environmental exposure at the time of grid stress.

A Appendix B: Key Code Summary

This appendix presents essential components of the modeling and forecasting pipeline implemented in this thesis.

The code is grouped into three stages: outage occurrence prediction, severity classification, and economic impact estimation. [2]

---

[2] AI-based tools such as OpenAI's ChatGPT were employed to assist with code refinement and technical clarity. Analytical decisions were subject to full human supervision.

Stage 1: Outage occurrence prediction

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import ast
import warnings
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import StratifiedKFold, RandomizedSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import (roc_auc_score, precision_recall_curve, auc, classi
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.calibration import calibration_curve
from scipy.stats import randint, uniform
from statsmodels.stats.outliers_influence import variance_inflation_factor
import xgboost as xgb
warnings.filterwarnings('ignore')

df = pd.read_csv('/content/drive/MyDrive/Thesis/merged_outages_weather_PREP.csv


# Mapping dictionary for cause codes
cause_mapping = {
    '1CD': 'Resource Insufficiency',
    '2FM': 'Force Majeure',
    '3CE': 'External Cause',
    '4AC': 'Other Cause',
    '5DP': 'Programmed Power Outage',
    'NaN': 'No outage'
}

# Function to expand coded causes into descriptive labels
def expand_cause(cause_list_str):
    if pd.isna(cause_list_str) or not isinstance(cause_list_str, str):
        return 'Unknown'
    try:
        cause_list = ast.literal_eval(cause_list_str)
        descriptions = [cause_mapping.get(code, f'Unknown Code: {code}') for cc
        return ' and '.join(descriptions)
    except Exception:
        return 'Parsing Error'
```

```
df['cause_label'] = df['codice_causa_AEEG_1_livello'].apply(expand_cause)

# Remove rows associated with programmed outages
df = df[df['cause_label'] != 'Programmed Power Outage']

# Overview of the target variable
print(df['outage_occurred'].value_counts(normalize=True))

# Time handling
df['date'] = pd.to_datetime(df['date'])

# Train/Test split: training before 2021, test from 2021 onward
train_df = df[df['date'] < '2021-01-01']
test_df = df[df['date'] >= '2021-01-01']

print("Train/Test split complete.")
print("Train size:", train_df.shape)
print("Test size:", test_df.shape)

# Address Class Imbalance in Training Set
train_pos = train_df[train_df['outage_occurred'] == 1]
train_neg = train_df[train_df['outage_occurred'] == 0]

ratio = 3
n_neg_samples = min(len(train_neg), ratio * len(train_pos))
train_neg_sampled = train_neg.sample(n=n_neg_samples, random_state=42)

train_balanced = pd.concat([train_pos, train_neg_sampled]).sample(frac=1, rando
print("Balanced train size:", train_balanced.shape)
print("New outage rate:", train_balanced['outage_occurred'].mean())

# Define Features and Targets
feature_cols = ['T2_C_max', 'PREC_max_log', 'WS_kmh_max_log', 'locality']
target_col = 'outage_occurred'

X_train = train_balanced[feature_cols]
y_train = train_balanced[target_col]
X_test = test_df[feature_cols]
y_test = test_df[target_col]
```

```
# One-Hot Encode Fixed Effects
# Province fixed effects (Model A)
X_train_A = train_balanced[['T2_C_max', 'PREC_max_log', 'WS_kmh_max_log', 'prov
X_test_A = test_df[['T2_C_max', 'PREC_max_log', 'WS_kmh_max_log', 'province']]
X_train_A = pd.get_dummies(X_train_A, columns=['province'], drop_first=True)
X_test_A = pd.get_dummies(X_test_A, columns=['province'], drop_first=True)
X_train_A, X_test_A = X_train_A.align(X_test_A, join='left', axis=1, fill_value

# Locality fixed effects (Model B)
X_train_B = pd.get_dummies(X_train, columns=['locality'], drop_first=True)
X_test_B = pd.get_dummies(X_test, columns=['locality'], drop_first=True)
X_train_B, X_test_B = X_train_B.align(X_test_B, join='left', axis=1, fill_value

# Baseline Model
# Logistic Regression
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import roc_auc_score, average_precision_score, classificat

def fit_and_evaluate_model(X_tr, y_tr, X_te, y_te, label="Model"):
    print(f"\n=== Evaluating {label} ===")
    model = LogisticRegression(class_weight='balanced', max_iter=1000, solver='
    model.fit(X_tr, y_tr)
    y_pred = model.predict(X_te)
    y_proba = model.predict_proba(X_te)[:, 1]

    print("ROC AUC:", roc_auc_score(y_te, y_proba))
    print("PR AUC:", average_precision_score(y_te, y_proba))
    print("Classification Report:\n", classification_report(y_te, y_pred))
    return model

model_A = fit_and_evaluate_model(X_train_A, y_train, X_test_A, y_test, label="L
model_B = fit_and_evaluate_model(X_train_B, y_train, X_test_B, y_test, label="L
```

```
# Choosing model B predictors
X_train = X_train_B
X_test = X_test_B

# Multicollinearity and Ridge Coefficient Stability Check
# Standardize X for VIF and Ridge
scaler = StandardScaler()
X_scaled_v1 = pd.DataFrame(scaler.fit_transform(X_train), columns=X_train.column

# VIF Calculation
vif_data = pd.DataFrame()
vif_data["Feature"] = X_scaled_v1.columns
vif_data["VIF"] = [variance_inflation_factor(X_scaled_v1.values, i) for i in ran
print("Top 10 VIF Scores:\n", vif_data.sort_values(by="VIF", ascending=False).he

# Ridge Logistic Regression
ridge_model = LogisticRegression(penalty='l2', solver='liblinear', C=1.0, max_it
ridge_model.fit(X_scaled_v1, y_train)

# Coefficients Summary
coef_df = pd.DataFrame({'Feature': X_scaled_v1.columns, 'Coefficient': ridge_mod
coef_df = coef_df.sort_values(by='Coefficient', key=abs, ascending=False)
print("Top 10 Ridge Coefficients:\n", coef_df.head(10))

# Evaluate on test set
X_test_scaled = pd.DataFrame(scaler.transform(X_test), columns=X_test.columns)
y_pred_ridge = ridge_model.predict(X_test_scaled)
print("Ridge Logistic Test Classification Report:\n", classification_report(y_te
```

```
# Trying different models

# Hyperparameter Tuning Functions
def tune_logistic(X_train, y_train):
    C_values = [0.01, 0.1, 1, 10, 100]
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    results = []

    for C in C_values:
        model = LogisticRegression(C=C, class_weight='balanced', max_iter=1000,
        pr_aucs = []
        for train_idx, val_idx in cv.split(X_train, y_train):
            model.fit(X_train.iloc[train_idx], y_train.iloc[train_idx])
            y_probs = model.predict_proba(X_train.iloc[val_idx])[:,1]
            precision, recall, _ = precision_recall_curve(y_train.iloc[val_idx]
            pr_aucs.append(auc(recall, precision))
        results.append((C, np.mean(pr_aucs)))

    best_C = max(results, key=lambda x: x[1])[0]
    final_model = LogisticRegression(C=best_C, class_weight='balanced', max_ite
    final_model.fit(X_train, y_train)
```

```
    return final_model

def tune_random_forest(X_train, y_train):
    param_dist = {
        'n_estimators': randint(100, 500),
        'max_depth': [5, 10, 20, None],
        'min_samples_split': [2, 5, 10],
        'max_features': ['sqrt', 'log2', None]
    }
    rf = RandomForestClassifier(class_weight='balanced', random_state=42, n_jo
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    rf_search = RandomizedSearchCV(rf, param_dist, scoring='average_precision',
    rf_search.fit(X_train, y_train)
    return rf_search.best_estimator_

def tune_xgboost(X_train, y_train):
    n_pos = y_train.sum()
    n_neg = len(y_train) - n_pos
    scale_weight = n_neg / n_pos

    param_dist = {
        'n_estimators': [100, 200, 300],
        'max_depth': [3, 5, 7],
        'learning_rate': uniform(0.01, 0.2),
        'subsample': [0.6, 0.8, 1.0],
        'colsample_bytree': [0.6, 0.8, 1.0]
    }
    xgb_model = xgb.XGBClassifier(
        objective='binary:logistic',
        scale_pos_weight=scale_weight,
        use_label_encoder=False,
        eval_metric='aucpr',
        random_state=42,
        n_jobs=-1
    )
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    xgb_search = RandomizedSearchCV(xgb_model, param_dist, scoring='average_pr
    xgb_search.fit(X_train, y_train)
    return xgb_search.best_estimator_

# Train Models
log_model = tune_logistic(X_train, y_train)
rf_model = tune_random_forest(X_train, y_train)
xgb_model = tune_xgboost(X_train, y_train)

# Evaluate Models
def evaluate_models(models, X_test, y_test):
    results = {}
    for label, model in models.items():
        y_scores = model.predict_proba(X_test)[:, 1]
        roc = roc_auc_score(y_test, y_scores)
```

```
        precision, recall, _ = precision_recall_curve(y_test, y_scores)
        prc = auc(recall, precision)
        results[label] = {'ROC AUC': roc}
    return results

models = {'Logistic Regression': log_model, 'Random Forest': rf_model, 'XGBoost
results = evaluate_models(models, X_test, y_test)

# Trying different models

# Hyperparameter Tuning Functions
def tune_logistic(X_train, y_train):
    C_values = [0.01, 0.1, 1, 10, 100]
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    results = []

    for C in C_values:
        model = LogisticRegression(C=C, class_weight='balanced', max_iter=1000,
        pr_aucs = []
        for train_idx, val_idx in cv.split(X_train, y_train):
            model.fit(X_train.iloc[train_idx], y_train.iloc[train_idx])
            y_probs = model.predict_proba(X_train.iloc[val_idx])[:,1]
            precision, recall, _ = precision_recall_curve(y_train.iloc[val_idx]
            pr_aucs.append(auc(recall, precision))
        results.append((C, np.mean(pr_aucs)))

    best_C = max(results, key=lambda x: x[1])[0]
    final_model = LogisticRegression(C=best_C, class_weight='balanced', max_ite
    final_model.fit(X_train, y_train)
    return final_model

def tune_random_forest(X_train, y_train):
    param_dist = {
        'n_estimators': randint(100, 500),
        'max_depth': [5, 10, 20, None],
        'min_samples_split': [2, 5, 10],
        'max_features': ['sqrt', 'log2', None]
    }
    rf = RandomForestClassifier(class_weight='balanced', random_state=42, n_job
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    rf_search = RandomizedSearchCV(rf, param_dist, scoring='average_precision',
    rf_search.fit(X_train, y_train)
    return rf_search.best_estimator_

def tune_xgboost(X_train, y_train):
    n_pos = y_train.sum()
    n_neg = len(y_train) - n_pos
    scale_weight = n_neg / n_pos

    param_dist = {
        'n_estimators': [100, 200, 300],
```

```
        'max_depth': [3, 5, 7],
        'learning_rate': uniform(0.01, 0.2),
        'subsample': [0.6, 0.8, 1.0],
        'colsample_bytree': [0.6, 0.8, 1.0]
    }
    xgb_model = xgb.XGBClassifier(
        objective='binary:logistic',
        scale_pos_weight=scale_weight,
        use_label_encoder=False,
        eval_metric='aucpr',
        random_state=42,
        n_jobs=-1
    )
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    xgb_search = RandomizedSearchCV(xgb_model, param_dist, scoring='average_pre
    xgb_search.fit(X_train, y_train)
    return xgb_search.best_estimator_

# Train Models
log_model = tune_logistic(X_train, y_train)
rf_model = tune_random_forest(X_train, y_train)
xgb_model = tune_xgboost(X_train, y_train)

# Evaluate Models
def evaluate_models(models, X_test, y_test):
    results = {}
    for label, model in models.items():
        y_scores = model.predict_proba(X_test)[:, 1]
        roc = roc_auc_score(y_test, y_scores)
        precision, recall, _ = precision_recall_curve(y_test, y_scores)
        prc = auc(recall, precision)
        results[label] = {'ROC AUC': roc}
    return results

models = {'Logistic Regression': log_model, 'Random Forest': rf_model, 'XGBoost
results = evaluate_models(models, X_test, y_test)

# Results
for model_name, scores in results.items():
    print(f"{model_name} - ROC AUC: {scores['ROC AUC']:.3f}")
```

```
# Enhance the Models
# Rolling Feature Engineering
window = 3
for col in ['T2_C_max', 'PREC_max_log', 'WS_kmh_max_log']:
    df[f'{col}_roll{window}'] = df.groupby('locality')[col].transform(lambda x:

# === Locality Clustering ===
locality_summary = df.groupby('locality').agg({
    'outage_occurred': 'mean',
```

```
    'T2_C_max': 'max',
    'PREC_max_log': 'max',
    'WS_kmh_max_log': 'max'
}).rename(columns={'outage_occurred': 'outage_rate'})

scaler = StandardScaler()
X_clust = scaler.fit_transform(locality_summary)
kmeans = KMeans(n_clusters=4, random_state=42)
locality_summary['cluster'] = kmeans.fit_predict(X_clust)
df = df.merge(locality_summary['cluster'], on='locality', how='left')

# Prepare Train/Test Split
train_cutoff = '2021-01-01'
train_df = df[df['date'] < train_cutoff].copy()
test_df = df[df['date'] >= train_cutoff].copy()

# Downsample train set (5:1 negative:positive)
train_pos = train_df[train_df['outage_occurred'] == 1]
train_neg = train_df[train_df['outage_occurred'] == 0]
train_neg_sampled = train_neg.sample(n=5*len(train_pos), random_state=42)
train_balanced = pd.concat([train_pos, train_neg_sampled]).sample(frac=1, rando

features = ['T2_C_max_roll3', 'PREC_max_log_roll3', 'WS_kmh_max_log_roll3', 'lo

X_train = pd.get_dummies(train_balanced[features], columns=['locality'], drop_f
X_test = pd.get_dummies(test_df[features], columns=['locality'], drop_first=Tru
X_test = X_test.reindex(columns=X_train.columns, fill_value=0)
y_train = train_balanced['outage_occurred']
y_test = test_df['outage_occurred']

from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import StratifiedKFold, RandomizedSearchCV
from sklearn.metrics import precision_recall_curve, auc
from scipy.stats import uniform, randint
import xgboost as xgb
import numpy as np

# Hyperparameter Tuning Functions

def tune_logistic(X_train, y_train):
    C_values = [0.01, 0.1, 1, 10, 100]
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    results = []

    for C in C_values:
        model = LogisticRegression(C=C, class_weight='balanced', max_iter=1000,
        pr_aucs = []
        for train_idx, val_idx in cv.split(X_train, y_train):
            model.fit(X_train.iloc[train_idx], y_train.iloc[train_idx])
            y_probs = model.predict_proba(X_train.iloc[val_idx])[:, 1]
```

```
            precision, recall, _ = precision_recall_curve(y_train.iloc[val_idx]
            pr_aucs.append(auc(recall, precision))
        results.append((C, np.mean(pr_aucs)))

    best_C = max(results, key=lambda x: x[1])[0]
    final_model = LogisticRegression(C=best_C, class_weight='balanced', max_ite
    final_model.fit(X_train, y_train)
    return final_model

def tune_random_forest(X_train, y_train):
    param_dist = {
        'n_estimators': randint(100, 500),
        'max_depth': [5, 10, 20, None],
        'min_samples_split': [2, 5, 10],
        'max_features': ['sqrt', 'log2', None]
    }
    rf = RandomForestClassifier(class_weight='balanced', random_state=42, n_job
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    rf_search = RandomizedSearchCV(rf, param_dist, scoring='average_precision',
    rf_search.fit(X_train, y_train)
    return rf_search.best_estimator_

def tune_xgboost(X_train, y_train):
    n_pos = y_train.sum()
    n_neg = len(y_train) - n_pos
    scale_weight = n_neg / n_pos

    param_dist = {
        'n_estimators': [100, 200, 300],
        'max_depth': [3, 5, 7],
        'learning_rate': uniform(0.01, 0.2),
        'subsample': [0.6, 0.8, 1.0],
        'colsample_bytree': [0.6, 0.8, 1.0]
    }
    xgb_model = xgb.XGBClassifier(
        objective='binary:logistic',
        scale_pos_weight=scale_weight,
        use_label_encoder=False,
        eval_metric='aucpr',
        random_state=42,
        n_jobs=-1
    )
    cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
    xgb_search = RandomizedSearchCV(
        xgb_model,
        param_dist,
        scoring='average_precision',
        cv=cv,
        n_iter=20,
        random_state=42,
        n_jobs=-1
```

```
    )
    xgb_search.fit(X_train, y_train)
    return xgb_search.best_estimator_

# Train Final Models
log_model = tune_logistic(X_train, y_train)
rf_model = tune_random_forest(X_train, y_train)
xgb_model = tune_xgboost(X_train, y_train)

models = {'Logistic Regression': log_model, 'Random Forest': rf_model, 'XGBoost
```

```
from sklearn.metrics import precision_score, recall_score, f1_score, confusion_

def compute_metrics(models, X_test, y_test, threshold=0.5):
    results = []
    for name, model in models.items():
        y_proba = model.predict_proba(X_test)[:, 1]
        y_pred = (y_proba >= threshold).astype(int)
        precision = precision_score(y_test, y_pred)
        recall = recall_score(y_test, y_pred)
        f1 = f1_score(y_test, y_pred)

        # Compute specificity from confusion matrix
        tn, fp, fn, tp = confusion_matrix(y_test, y_pred).ravel()
        specificity = tn / (tn + fp)

        results.append((name, round(precision, 3), round(recall, 3), round(f1,

    return pd.DataFrame(results, columns=["Model", "Precision", "Recall", "F1-S

metrics_df = compute_metrics(models, X_test, y_test)
print(metrics_df.sort_values(by='F1-Score', ascending=False).head(3))
```

```
import matplotlib.pyplot as plt
from sklearn.metrics import roc_curve, precision_recall_curve, auc

def plot_roc(models, X_test, y_test):
    plt.figure(figsize=(6, 5))
    for name, model in models.items():
        y_scores = model.predict_proba(X_test)[:, 1]
        fpr, tpr, _ = roc_curve(y_test, y_scores)
        roc_auc = auc(fpr, tpr)
        plt.plot(fpr, tpr, label=f"{name} (AUC = {roc_auc:.2f})")
    plt.plot([0, 1], [0, 1], 'k--')
    plt.title("ROC Curve")
    plt.xlabel("False Positive Rate")
    plt.ylabel("True Positive Rate")
    plt.legend()
    plt.tight_layout()
    plt.savefig("model_roc_curves.png", dpi=300)
    plt.show()

plot_roc(models, X_test, y_test)
```

Stage 2: Severity Class Prediction

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
from sklearn.metrics import classification_report
from collections import defaultdict
import warnings
```

```
# === Rolling Feature Engineering ===
df_model = pd.read_csv('/content/drive/MyDrive/Thesis/merged_outages_weather_PR
window = 3
for col in ['T2_C_max', 'PREC_max_log', 'WS_kmh_max_log']:
    df_model[f'{col}_roll{window}'] = df_model.groupby('locality')[col].transfo

# Create severity_class
def create_severity_class(row):
    ens = row['ENS_level']
    int_type = row['interruption_type_clean']

    if ens == 'High' and int_type == 'Long':
        return 'Severe'
```

```
    elif  ens in ['High',"Medium"] and int_type in ['Brief', 'Transitory']:
        return 'Moderate'
    elif ens == 'Low' and int_type == 'Long':
        return 'Low'
    elif ens == 'Low' and int_type in ['Brief', 'Transitory']:
        return 'Very Low'
    else:
        return 'Not Severe - Programmed'

# Apply
df_model['severity_class'] = df_model.apply(create_severity_class, axis=1)

# order severity class: Not Severe - Programmed , Very Low, Low, Moderate, Seve
df_model['severity_class'] = pd.Categorical(df_model['severity_class'], categor

# Exclude Programmed Outages
df_model = df_model[df_model['severity_class'] != 'Not Severe - Programmed']

# Encode target
le = LabelEncoder()
df_model['severity_encoded'] = le.fit_transform(df_model['severity_class'])

# Train/Test Split (time-aware)
df_model['date'] = pd.to_datetime(df_model['date'])
train_df = df_model[df_model['date'] < '2021-01-01']
test_df = df_model[df_model['date'] >= '2021-01-01']

# Results storage
results_all = defaultdict(lambda: defaultdict(dict))

# Creating Locality Clusters
locality_summary = df_model.groupby('locality').agg({
    'outage_occurred': 'mean',       # outage_rate
    'T2_C_max': 'max',                # maximum temperature exposure
    'PREC_max_log': 'max',            # maximum precipitation exposure (log)
    'WS_kmh_max_log': 'max'           # maximum wind exposure (log)
}).rename(columns={
    'outage_occurred': 'outage_rate'
}).reset_index()

# Rolling Feature Engineering
window = 3
for col in ['T2_C_max', 'PREC_max_log', 'WS_kmh_max_log']:
    df_model[f'{col}_roll{window}'] = df_model.groupby('locality')[col].transfo

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
X_locality = scaler.fit_transform(locality_summary[['outage_rate', 'T2_C_max',

# Apply Kmeans
```

```
from sklearn.cluster import KMeans

kmeans = KMeans(n_clusters=4, random_state=42)
locality_summary['locality_cluster'] = kmeans.fit_predict(X_locality)

df_model = df_model.merge(locality_summary[['locality', 'locality_cluster']], c


# Functions

def prepare_features(feature_version, train_df, test_df):
    weather_cols = ['T2_C_max_roll3', 'PREC_max_log_roll3', 'WS_kmh_max_log_rol

    if feature_version == 'full_locality':
        features = weather_cols + ['locality']
    elif feature_version == 'cluster_locality':
        features = weather_cols + ['locality_cluster']
    else:
        raise ValueError("Invalid feature set.")

    # One-hot encode categorical locality features
    X_train = pd.get_dummies(train_df[features], drop_first=True)
    X_test = pd.get_dummies(test_df[features], drop_first=True)

    # Align columns (in case train and test have slightly different locality du
    X_train, X_test = X_train.align(X_test, join='left', axis=1, fill_value=0)

    y_train = train_df['severity_encoded']
    y_test = test_df['severity_encoded']

    return X_train, X_test, y_train, y_test


def prepare_models(imbalance):
    models = {}

    if imbalance:
        models['LogReg'] = LogisticRegression(max_iter=1000, class_weight='bala
        models['RandomForest'] = RandomForestClassifier(n_estimators=100, class
        models['XGBoost'] = XGBClassifier(n_estimators=100, learning_rate=0.1,
    else:
        models['LogReg'] = LogisticRegression(max_iter=1000, solver='lbfgs')
        models['RandomForest'] = RandomForestClassifier(n_estimators=100, rando
        models['XGBoost'] = XGBClassifier(n_estimators=100, learning_rate=0.1,

    return models


def train_and_evaluate(models, X_train, X_test, y_train, y_test, feature_versio
    for model_name, model in models.items():
        model.fit(X_train, y_train)
```

```
        y_pred = model.predict(X_test)

        report = classification_report(y_test, y_pred, target_names=le.classes_

        # Compute per-class specificity
        specificity_per_class = {}
        for class_index, class_label in enumerate(le.classes_):
            # One-vs-rest
            y_true_binary = (y_test == class_index).astype(int)
            y_pred_binary = (y_pred == class_index).astype(int)
            tn, fp, fn, tp = confusion_matrix(y_true_binary, y_pred_binary).rav
            specificity = tn / (tn + fp) if (tn + fp) > 0 else 0.0
            specificity_per_class[class_label] = round(specificity, 3)

        # Add specificity to the report
        for label in le.classes_:
            if label in report:
                report[label]["specificity"] = specificity_per_class[label]

        # Store report
        results_all[feature_version][imbalance_flag][model_name] = report

        # Print summary
        print(f"\n===== Feature Set: {feature_version} | Model: {model_name} |
        print(classification_report(y_test, y_pred, target_names=le.classes_))
        print("Specificity per class:")
        for label, spec in specificity_per_class.items():
            print(f"{label}: {spec}")


for feature_version in ['full_locality', 'cluster_locality']:
    X_train, X_test, y_train, y_test = prepare_features(feature_version, train_

    for imbalance_flag in [False, True]:
        models = prepare_models(imbalance_flag)
        train_and_evaluate(models, X_train, X_test, y_train, y_test, feature_ve
```

```
# Best three models
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
from sklearn.metrics import classification_report, roc_curve, auc, confusion_ma
from sklearn.preprocessing import label_binarize
from sklearn.multiclass import OneVsRestClassifier
```

```
from sklearn.preprocessing import StandardScaler
from collections import defaultdict
import warnings

warnings.filterwarnings("ignore")

# Define the correct class order
class_order = ['Very Low', 'Low', 'Moderate', 'Severe']

# Encode target
le = LabelEncoder()
df_model['severity_class'] = pd.Categorical(df_model['severity_class'], categor
df_model['severity_encoded'] = df_model['severity_class'].cat.codes

# Train/Test Split (time-aware)
df_model['date'] = pd.to_datetime(df_model['date'])
train_df = df_model[df_model['date'] < '2021-01-01']
test_df = df_model[df_model['date'] >= '2021-01-01']

# Prepare features (full locality)
weather_cols = ['T2_C_max', 'PREC_max_log', 'WS_kmh_max_log']
features = weather_cols + ['locality']

X_train = pd.get_dummies(train_df[features], drop_first=True)
X_test = pd.get_dummies(test_df[features], drop_first=True)
X_train, X_test = X_train.align(X_test, join='left', axis=1, fill_value=0)

y_train = train_df['severity_encoded']
y_test = test_df['severity_encoded']

# Models
models = {
    'Logistic Regression': LogisticRegression(max_iter=1000, solver='lbfgs'),
    'Random Forest': RandomForestClassifier(n_estimators=100, random_state=42),
    'XGBoost': XGBClassifier(n_estimators=100, learning_rate=0.1, use_label_enc
}

# Train models and collect results
model_preds = {}
model_probs = {}
trained_models = {}

for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    y_prob = model.predict_proba(X_test)
    print(f"\n===== {name} =====")
    print("Hyperparameters:", model.get_params())
    print("\nClassification Report:")
    print(classification_report(y_test, y_pred, target_names=class_order))
    model_preds[name] = y_pred
```

```
    model_probs[name] = y_prob
    trained_models[name] = model

    # Plot ROC curves
    classes = class_order
    y_test_bin = label_binarize(y_test, classes=[0, 1, 2, 3])
    plt.figure(figsize=(10, 8))
    for i in range(len(classes)):
        fpr, tpr, _ = roc_curve(y_test_bin[:, i], y_prob[:, i])
        roc_auc = auc(fpr, tpr)
        plt.plot(fpr, tpr, lw=2, label=f'{classes[i]} (AUC = {roc_auc:.2f})')
    plt.plot([0, 1], [0, 1], 'k--')
    plt.xlim([0.0, 1.0])
    plt.ylim([0.0, 1.05])
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title(f'ROC Curves for {name}')
    plt.legend(loc='lower right')
    plt.grid()
    plt.show()

    # Plot Confusion Matrix
    cm = confusion_matrix(y_test, y_pred, labels=[0, 1, 2, 3])
    plt.figure(figsize=(8,6))
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=class_order,
    plt.xlabel('Predicted')
    plt.ylabel('True')
    plt.title(f'Confusion Matrix for {name}')
    plt.show()

    # Feature Importance
    if name == 'Logistic Regression':
        importance = np.abs(model.coef_[0])
        feature_names = X_train.columns
        imp_df = pd.DataFrame({'feature': feature_names, 'importance': importar
        imp_df = imp_df.sort_values(by='importance', ascending=False).head(10)
        plt.figure(figsize=(10, 6))
        plt.barh(imp_df['feature'], imp_df['importance'])
        plt.gca().invert_yaxis()
        plt.title(f'Top 10 Coefficients for {name}')
        plt.xlabel('Coefficient Magnitude')
        plt.grid()
        plt.show()
    elif name in ['Random Forest', 'XGBoost']:
        importance = model.feature_importances_
        feature_names = X_train.columns
        imp_df = pd.DataFrame({'feature': feature_names, 'importance': importar
        imp_df = imp_df.sort_values(by='importance', ascending=False).head(10)
        plt.figure(figsize=(10, 6))
        plt.barh(imp_df['feature'], imp_df['importance'])
        plt.gca().invert_yaxis()
```

```
        plt.title(f'Top 10 Feature Importances for {name}')
        plt.xlabel('Importance')
        plt.grid()
        plt.show()
```

Stage 3: Economical Impact Forecast

```
import statsmodels.formula.api as smf
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import RidgeCV, LassoCV
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
import statsmodels.api as sm

df = pd.read_csv('/content/drive/MyDrive/Thesis/merged_outages_weather_PREP_seve

# Rolling features
window = 3
for col in ["T2_C_max", "PREC_max_log", "WS_kmh_max_log"]:
    df[f"{col}_roll{window}"] = df.groupby("locality")[col].transform(lambda x:

# Filter to ENS > 0 outage events
df_pos_ens = df[(df["outage_occurred"] == 1) & (df["ENS_ENR_netta"] > 0)].copy()

# Log-transform the target variable
df_pos_ens["log_ENS"] = np.log(df_pos_ens["ENS_ENR_netta"])

# Baseline
# Fit the OLS regression with locality fixed effects
model = smf.ols(
    formula="log_ENS ~ T2_C_max_roll3 + PREC_max_log_roll3 + WS_kmh_max_log_roll
    data=df_pos_ens
).fit()

print(model.summary())
```

```
# Regularised Models

import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
from sklearn.linear_model import RidgeCV, Lasso
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Filter and transform
df_pos_ens = df[(df["outage_occurred"] == 1) & (df["ENS_ENR_netta"] > 0)].copy()
df_pos_ens["log_ENS"] = np.log(df_pos_ens["ENS_ENR_netta"])

# Feature selection and encoding
features = ["T2_C_max_roll3", "PREC_max_log_roll3", "WS_kmh_max_log_roll3", "loc
df_model = df_pos_ens[["log_ENS", "date"] + features].dropna()
df_model = pd.get_dummies(df_model, columns=["locality"], drop_first=True)

# Prepare X and y
X = df_model.drop(columns=["log_ENS", "date"])
y = df_model["log_ENS"]

# Pipelines for Ridge and Lasso (on full data)
ridge = Pipeline([
    ("scaler", StandardScaler()),
    ("model", RidgeCV(alphas=np.logspace(-3, 3, 100), scoring='neg_mean_squared_
])

lasso = Pipeline([
    ("scaler", StandardScaler()),
    ("model", Lasso(alpha=0.01, max_iter=10000))
])

# Fit models on full data
ridge.fit(X, y)
lasso.fit(X, y)

# Predict and evaluate
ridge_pred = ridge.predict(X)
lasso_pred = lasso.predict(X)

ridge_r2 = r2_score(y, ridge_pred)
lasso_r2 = r2_score(y, lasso_pred)
ridge_rmse = np.sqrt(mean_squared_error(y, ridge_pred))
lasso_rmse = np.sqrt(mean_squared_error(y, lasso_pred))

# Print results
print(f"Ridge R²: {ridge_r2:.3f}")
print(f"Ridge RMSE: {ridge_rmse:.3f}")
print(f"Lasso R²: {lasso_r2:.3f}")
print(f"Lasso RMSE: {lasso_rmse:.3f}")

# Coefficient inspection (Lasso only)
lasso_coef = pd.Series(lasso.named_steps['model'].coef_, index=X.columns)
lasso_nonzero = lasso_coef[lasso_coef != 0].sort_values(key=np.abs, ascending=Fa
print("\nTop 10 non-zero Lasso coefficients:")
print(lasso_nonzero.head(10))

# Coefficient bar plot
if not lasso_nonzero.empty:
```

```
if not lasso_nonzero.empty:
    plt.figure(figsize=(10, 6))
    lasso_nonzero.head(10).plot(kind='barh')
    plt.xlabel("Coefficient Value")
    plt.ylabel("Feature Name")
    plt.title("Top 10 Non-Zero Lasso Coefficients (Full Sample)")
    plt.gca().invert_yaxis()
    plt.tight_layout()
    plt.show()
else:
    print("Lasso selected no features. Skipping plot.")

# Residual plot
residuals_ridge = y - ridge_pred
residuals_lasso = y - lasso_pred

plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
sns.scatterplot(x=ridge_pred, y=residuals_ridge)
plt.axhline(0, color='red', linestyle='--')
plt.title("Ridge (Full Data): Residuals vs. Predicted")
plt.xlabel("Predicted")
plt.ylabel("Residuals")

plt.subplot(1, 2, 2)
sns.scatterplot(x=lasso_pred, y=residuals_lasso)
plt.axhline(0, color='red', linestyle='--')
plt.title("Lasso (Full Data): Residuals vs. Predicted")
plt.xlabel("Predicted")
plt.ylabel("Residuals")

plt.tight_layout()
plt.show()
```

```
# Scenario
import numpy as np
import pandas as pd

# Constants
YEARS = 3
ENS_THRESHOLD = 763     # MWh per year
ENS_PRICE = 27000       # €/MWh
scenarios = [0, 0.10, 0.20, 0.50]

# Placeholder for results
cost_rows = []

# Extract scaler and feature index for wind
scaler = ridge.named_steps['scaler']
ws_idx = list(X.columns).index("WS_kmh_max_log_roll3")
ws_mean = scaler.mean_[ws_idx]
```

```
ws_std = scaler.scale_[ws_idx]

# Reconstruct true wind values (pre-log)
log_ws_std = X["WS_kmh_max_log_roll3"]
log_ws = log_ws_std * ws_std + ws_mean
ws_original = np.exp(log_ws)

# Loop through wind scenarios
for s in scenarios:
    # Apply intensification in original space
    ws_intensified = ws_original * (1 + s)

    # Log-transform, then standardize
    log_ws_modified = np.log(ws_intensified)
    ws_scaled_std = (log_ws_modified - ws_mean) / ws_std

    X_scenario = X.copy()
    X_scenario["WS_kmh_max_log_roll3"] = ws_scaled_std

    # Predict log(ENS)
    ridge_preds_log = ridge.predict(X_scenario)
    lasso_preds_log = lasso.predict(X_scenario)

    # Convert to ENS (MWh)
    ridge_mean = np.exp(ridge_preds_log).sum() / YEARS
    lasso_mean = np.exp(lasso_preds_log).sum() / YEARS

    # Confidence intervals (log-normal)
    ridge_low = np.exp(ridge_preds_log - 1.96 * ridge_rmse).sum() / YEARS
    ridge_high = np.exp(ridge_preds_log + 1.96 * ridge_rmse).sum() / YEARS
    lasso_low = np.exp(lasso_preds_log - 1.96 * lasso_rmse).sum() / YEARS
    lasso_high = np.exp(lasso_preds_log + 1.96 * lasso_rmse).sum() / YEARS

    # Apply ARERA penalty rule
    ridge_cost = max(0, ridge_mean - ENS_THRESHOLD) * ENS_PRICE / 1e3
    lasso_cost = max(0, lasso_mean - ENS_THRESHOLD) * ENS_PRICE / 1e3
    ridge_low_cost = max(0, ridge_low - ENS_THRESHOLD) * ENS_PRICE / 1e3
    ridge_high_cost = max(0, ridge_high - ENS_THRESHOLD) * ENS_PRICE / 1e3
    lasso_low_cost = max(0, lasso_low - ENS_THRESHOLD) * ENS_PRICE / 1e3
    lasso_high_cost = max(0, lasso_high - ENS_THRESHOLD) * ENS_PRICE / 1e3

    # Store results
    cost_rows.append({
        "Scenario": "Baseline" if s == 0 else f"+{int(s*100)}% Wind",
        "Ridge Forecast (k€)": round(ridge_cost),
        "Ridge Range (k€)": f"{round(ridge_low_cost)} – {round(ridge_high_cost)
        "Lasso Forecast (k€)": round(lasso_cost),
        "Lasso Range (k€)": f"{round(lasso_low_cost)} – {round(lasso_high_cost)
    })

# Output
```

```
cost_df = pd.DataFrame(cost_rows)
print("\nForecasted Regulatory Economic Impact (ARERA Rule):")
print(cost_df)


# Scenario ENS in MWh
import numpy as np
import pandas as pd

# Constants
YEARS = 3  # 2021–2023
scenarios = [0, 0.10, 0.20, 0.50]

# Placeholder for results
ens_rows = []

# Extract scaler parameters
scaler = ridge.named_steps['scaler']
ws_idx = list(X.columns).index("WS_kmh_max_log_roll3")
ws_mean = scaler.mean_[ws_idx]
ws_std = scaler.scale_[ws_idx]

# Reconstruct original wind speed
log_ws_std = X["WS_kmh_max_log_roll3"]
log_ws = log_ws_std * ws_std + ws_mean
ws_original = np.exp(log_ws)

# Loop through intensification scenarios
for s in scenarios:
    # Intensify wind in original (unlogged) domain
    ws_intensified = ws_original * (1 + s)

    # Log-transform, then re-standardize
    log_ws_intensified = np.log(ws_intensified)
    ws_scaled_std = (log_ws_intensified - ws_mean) / ws_std

    # Build scenario dataset
    X_scenario = X.copy()
    X_scenario["WS_kmh_max_log_roll3"] = ws_scaled_std

    # Predict log(ENS)
    ridge_preds_log = ridge.predict(X_scenario)
    lasso_preds_log = lasso.predict(X_scenario)

    # Convert to ENS in MWh
    ridge_ens = np.exp(ridge_preds_log)
    lasso_ens = np.exp(lasso_preds_log)

    # Aggregate annually
    ridge_mean = ridge_ens.sum() / YEARS
    lasso_mean = lasso_ens.sum() / YEARS
```

```
    # 95% log-normal confidence intervals
    ridge_low = np.exp(ridge_preds_log - 1.96 * ridge_rmse).sum() / YEARS
    ridge_high = np.exp(ridge_preds_log + 1.96 * ridge_rmse).sum() / YEARS
    lasso_low = np.exp(lasso_preds_log - 1.96 * lasso_rmse).sum() / YEARS
    lasso_high = np.exp(lasso_preds_log + 1.96 * lasso_rmse).sum() / YEARS

    # Store results
    ens_rows.append({
        "Scenario": "Baseline" if s == 0 else f"{int(s * 100)}%",
        "Ridge ENS": round(ridge_mean),
        "Ridge Low": round(ridge_low),
        "Ridge High": round(ridge_high),
        "Lasso ENS": round(lasso_mean),
        "Lasso Low": round(lasso_low),
        "Lasso High": round(lasso_high)
    })

# Compile final DataFrame
ens_df = pd.DataFrame(ens_rows)
print("\nForecasted Annual ENS Under Wind Intensification Scenarios (MWh):")
print(ens_df)
```