Extreme weather disturbance of electrical substations: a geospatial prediction and impact assessment model for Portugal

Tim Bajramaj

Dissertation written under the supervision of professor

Prof. Nicolo Bertani

Dissertation submitted in partial fulfilment of requirements for the M.Sc. in Business Analytics, at the Universidade Católica Portuguesa

May 2026

Abstract

Title of thesis: Extreme weather disturbance of electrical substations: a geospatial prediction and impact assessment model for Portugal
Author: Tim Bajramaj

English Extreme weather events are an increasing threat to electrical distribution infrastructure, yet data-driven approaches to substation-level disturbance prediction remain scarce, particularly at the most granular level of the high-to-medium voltage interface. This thesis develops and deploys a modular, open-data framework for predicting and contextualising weather-driven disturbances of substations on the Portuguese mainland.

A three-type sampling strategy is used to train a calibrated Gradient Boosting classifier, which achieves 81.3% accuracy, 70.9% recall, and a ROC-AUC of 0.866 on held-out test set of unseen installations. Daily and cumulative precipitation, temperature extremes, and geographic latitude emerge as the dominant disturbance predictors, with isolated 15 kV substations exhibiting disproportionate vulnerability. A self-developed line-length-proportional allocation mechanism assigns parish-level population, enterprise activity, and critical infrastructure exposure to individual substations across all 2,882 parishes, producing a bounded Criticality Index that ranks installations by their socio-economic consequence in case of disturbance. Both models are operationalized in Grid Sentinel PT, an open-source dashboard delivering automated, daily-updated, three-day-ahead disturbance risk forecasts for every substation on the Portuguese mainland. The framework is designed as a transferable blueprint for other European distribution operators.

Português Os eventos meteorológicos extremos representam uma ameaça crescente para a infraestrutura elétrica de distribuição, porém as abordagens baseadas em dados para a previsão de perturbações ao nível das subestações permanecem escassas, em particular na interface entre alta tensão e média tensão. Esta tese desenvolve e implementa um modelo modular, baseado em dados abertos, para prever e contextualizar perturbações meteorológicas em subestações do continente português.

Uma estratégia de amostragem em três tipos é utilizada para treinar um classificador Gradient Boosting calibrado, que alcança uma precisão de 81,3%, uma taxa de deteção de 70,9% e uma ROC-AUC de 0,866 num conjunto de teste com instalações não observadas durante o treino. A precipitação diária e acumulada, as temperaturas extremas e a latitude geográfica emergem como os principais preditores de perturbação, sendo as subestações isoladas de 15 kV as que apresentam maior vulnerabilidade. Um mecanismo de alocação proporcional ao comprimento de linha, desenvolvido nesta tese, distribui a população, a atividade empresarial e a exposição a infraestrutura crítica ao nível de freguesia pelas subestações, abrangendo as 2 882 freguesias do continente, e produz um Índice de Criticidade (IC) delimitado que classifica as instalações pelo seu impacto socioeconómico em caso de perturbação. Ambos os modelos são operacionalizados no Grid Sentinel PT, uma plataforma de código aberto que fornece previsões de risco automatizadas, atualizadas diariamente para um horizonte de três dias, para cada subestação do continente português. O modelo é concebido como um referencial transferível para outros operadores de distribuição europeus.

Keywords Weather-driven outage prediction, Distribution network resilience, Gradient boosting classifier, Socio-economic criticality assessment, Geospatial risk allocation

Acknowledgements

I would like to express my sincere gratitude to Nicolò, a key shaper of this thesis, whose guidance and expertise influenced this work in more ways than I had expected. His support throughout this journey has been invaluable, especially the informal chats before and after settling on the thesis topic.

I am also deeply grateful to Católica and all the valuable learnings I was able to apply throughout this work. My time in Lisbon and St. Gallen will never be forgotten.

To my friends and colleagues: thank you for the espresso chats, sometimes longer, sometimes shorter. Thanks for Italian-Style office-lunches and the numerous padel-games, in between lectures and submissions. Thank you for making this Master's journey a once-in-a-lifetime experience.

Most importantly, to my family and my girlfriend, whose unconditional and never-ending support has carried me further than they will ever know. Being able to complete each step of this journey, knowing you will always have my back — no matter what comes — gives me incredible strength and confidence. I am humbled every day to work on topics that truly matter to me. Babi, Nonä, Ari, Oma, Ludwig, and Mima — thank you for everything.

Contents

List of Figures

List of Tables

List of abbreviations

CAOP	Carta Administrativa Oficial de Portugal - Official Map Administration Portugal
CI	Criticality Index
DIM	Disturbance Impact Model
DPM	Disturbance Prediction Model
DSO	Distribution System Operator
EDA	Exploratory Data Analysis
ERSE	Entidade Reguladora dos Serviços Energéticos - Energy Services Regulatory Authority
FN	False Negative
FP	False Positive
GB	Gradient Boost Model
GS	Grid Sentinel PT
HV	High voltage
IDW	Inverse-Distance-Weighting
INE	Instituto Nacional de Estatística - Statistics Portugal
IPMA	Instituto Português do Mar e da Atmosfera - The Portuguese Institute for Sea and Atmosphere
LGBM	Light Gradient Boosting Machine
LV	Low voltage
ML	Machine learning
MV	Medium voltage
MWU	Mann–Whitney U
OSM	Open-Street-Map

POI Point-of-Interest

ROC Receiver Operating Characteristic

ROC-AUC Receiver Operating Characteristic Area Under The Curve

RF Random Forest

RQ Research Question

TN True Negative

TP True Positive

TSO Transmission Grid Operator

WMO World Meteorological Organization

XGB Extreme Gradient Boosting

1 Introduction

Relevance of the Topic

Over 800,000 residents of central and Northern Portugal were left without electricity in January of 2026. The root cause, affecting approx. 9% of the Portuguese mainland population was Storm Kristin. Before hitting the Spanish part of the Iberian peninsula, it hit Portugal with wind gusts above 200 km/h, heavy rain and even snowfall (Euronews, 2026).

However, Storm Kristin is not a singular event, as it is closely followed by additional storms Leonardo and Marta, roughly ten and thirteen days, apart respectively, again causing flooding and evacuation of thousands of citizens. Next to a dozen of fatalities, the series of storms caused an estimated financial damage of EUR 775 million (The Guardian, 2026).

Given the strong increase of storm patterns on the Iberian peninsula, resilience of society and infrastructure become more important than before. Electricity, since its introduction, has a transforming character for nations, industries, and its people. It is a key element to accelerate the industrialization process, thus countries rely on the increasing commercial energy demand to raise its per capita income. This is enabled by process improvements, cost reductions, labour time savings and health improvements brought by electricity. In short: Electricity plays a crucial role in economic growth and prosperity (UNDP, 2000).

The broad consensus of science that rising temperatures and climate change will further impact the stability of electrical grids in the future are not negligible, as recent events show drastically. For Portugal only, the number of extreme weather events doubled from two to four events within four years from 2014 to 2018 (Almeida et al., 2019).

Thus, securing the electrical infrastructure as the backbone of society and economy is already today a core task of grid operators. Alongside a decentralized energy system, increasing extreme weather events are challenging grid operators across the world. This thesis therefore aims to contribute to this core task of Portuguese grid operators.

Problem Setting, Objective and Approach

The electricity grid's age, in line with the increasing digitalization of networks allow grid operators to adopt new work planning methods - stepping away from empirical to data-driven decision making (Almeida et al., 2019). The usage of data-driven impact forecast models in grid operations is limited. Especially the lacking focus on substations, with just-in-time open source data. In combination with increasing frequency of extreme weather events with immense effect on the broad society, this thesis aims to examine this limitation for Portugal.

The objective of this thesis is to create transparency of disturbances of substations for the broad society, grid operators, and companies. Those should be enabled to identify and prioritize grid bottlenecks based on socio-economic impact of disturbances. This will be based on an open-data approach, which can serve as a blueprint for other European countries.
Following a rigorous scientific approach the following Research Questions (RQs) are aimed to be answered within this thesis:

1. What extreme weather events and patterns lead to disturbances of substations in between the High voltage (HV) and Medium voltage (MV) grid on the Portuguese mainland?
2. What Machine learning (ML) algorithm is best suited for predicting disturbance events of substations, taking geo-spatial components into account?
3. How can substations be ranked based on their socio-economic and critical infrastructure importance?
4. How can a disturbance-prediction model be deployed to create real-life impact for grid operators, the society and companies?

The stated approach is integrated in a modular approach answering the stated research questions. After Chapter 2 stating contextual knowledge on the Portuguese power grid, weather events, and machine learning, Chapter 3 describes the data collection process and its proceeding transformation steps, including feature engineering, followed by an Exploratory Data Analysis (EDA) of key variable distributions. In Chapter 4, the two models are developed and evaluated: the Disturbance Prediction Model (DPM) is benchmarked and fine-tuned, the Criticality Index (CI) of the Disturbance Impact Model (DIM) scoring framework is constructed, and both are operationalized in Grid Sentinel PT (GS). Lastly, Chapter 5 discusses the results and limitations of each model component and provides a blueprint for transferring the framework to other European distribution operators, before summarizing the answers to the stated RQs.

Results

The analysis of the disturbance records reveals a pronounced seasonal concentration of events between October and February, most likely driven by Atlantic storm systems delivering intense precipitation and wind. Isolated 15 kV substations emerge as disproportionately vulnerable, accounting for the majority of recorded disturbances across the study period.
A calibrated Gradient Boost Model (GB) classifier achieves 81.3% accuracy and 70.9% recall on the held-out test set, with a Receiver Operating Characteristic (ROC)-AUC of 0.866, nearly doubling the random baseline in average precision. Daily and cumulative precipitation, temperature extremes, and geographic latitude rank as the dominant disturbance predictors, confirming that both weather intensity and spatial exposure jointly govern substation vulnerability.
A self-developed allocation mechanism proxies parish-level population, enterprise activity, and critical infrastructure exposure to substations. A resulting CI ranks all substations on a bounded $[0, 1]$ scale, with installations in the Lisbon and Porto metropolitan areas achieving the highest scores, reflecting their dense concentration of residents and life-safety-critical facilities.
Both models are operationalized in GS, a publicly accessible risk dashboard that delivers automated, daily-updated, three-day-ahead disturbance risk forecasts for every substation on the Portuguese mainland, providing grid operators, companies, and the public with an actionable, real-time decision-support tool.

2 Contextual Knowledge and Related Work

2.1 Contextual Fundamentals

Within this Chapter fundamental concepts of the Portuguese power grid (2.1.1), meteorology and extreme weather events in Portugal (2.1.2), and machine learning (2.1.3) will be stated. The Chapter is concluded with an overview of related work in the field (2.2) and the resulting gap this thesis aims to fill.

2.1.1 Portuguese Power Grid

Stakeholders and grid topology

As core infrastructure for modern societies, an overview of active stakeholders in the Portuguese power grid is necessary to understand the limitations later introduced within this work. The following list, based on (CIGRE, 2020; Heuck et al., 2007) briefly describes those stakeholders, which are summarized in Figure 2.1:

- **Transmission Grid Operator (TSO)**: REN, the only Portuguese TSO, operates the highest voltage levels and is responsible for balancing of demand and supply, frequency control and system restoration
- **Distribution System Operator (DSO)**: DSOs are responsible for managing, operating the lower-voltage grids, extending it to new supply locations and maintaining the lines, transformer stations and ancillary facilities and ensuring the quality of service
- **Regulator**: Energy Services Regulatory Authority (ERSE) is responsible for the economic regulation of the electricity distribution. Further, ERSE sets quality criteria for final consumers
- **Prosumers**: Final consumers of electricity, either residential or commercial users. With increasing build-out of distributed renewable energies, the definition of consumers shifts

to prosumers, given an increasing producer character of end-consumers

- Several other parties, e.g., energy producers, traders and retailers

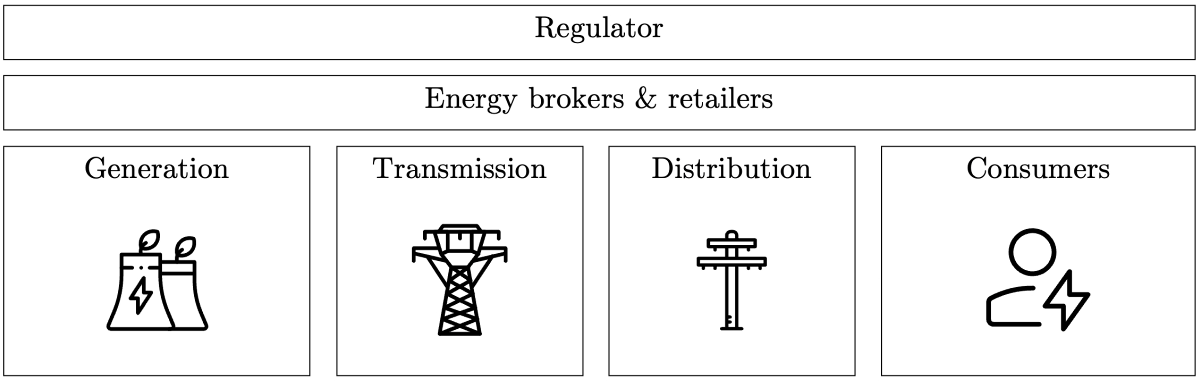


Figure 2.1: Active parties in the Portuguese power grid

Further, alongside the parties, explaining the underlying hardware is helpful to understand the need for substations and potential weaknesses across the grid. As visible in the appendix (A.2), the power grid is split into different voltage levels to satisfy the different objectives of each level. HV levels (> 150kV) are used to transport electricity across larger distance with minimal power loss possible, whereas Low voltage (LV) grids transport electricity in directly useable voltages, proximate for final consumers. In between each voltage level substations are required to transform the electricity (Heuck et al., 2007). As example, the substation ARGANIL, opened in 2023, visible in figure 2.2 combines the HV and MV levels. It supplies electricity to ~4-5,000 residential customers (E-REDES, 2023).

Figure 2.2: Visualization of Substation ARGANIL; (E-REDES, 2023)

E-REDES, the largest Portuguese DSO operates > 72,000 substations. Given the large coverage of E-REDES across the entire mainland and its operation of HV and LV distribution grids, only

E-REDES stations on the Portuguese mainland will be considered. Further, this thesis only considers primary substations between MV and HV level. Thus, secondary substations (LV), based on their constructional differences will be excluded.

Disturbance Risk of Substations

Based on the rare occasion of full electricity outage events, a typical pattern in infrastructure failures, a broader definition rather than outages will be used. Thus, in the context of this thesis, disturbances events, that include full outages but also voltage deviations, and flicker events of substations will be used.
According to (Sharma et al., 2023), in line with (IEEE, 2014), various root-causes for disturbances exist. The first level differentiates outages in pre-planned (maintenance) and unplanned events. Unplanned events, can be further split into the five root-cause categories, sorted by relevance for disturbances:

- **Natural events**: Natural-based disturbances caused by weather, wildfire and vegetation factors
- **Equipment**: Equipment failures including electric failures, damages, wear and tear
- **Power supply**: Outages of the interconnected power supply based on grid overload, and failure of transmission
- **Public**: Human-based disturbances mostly caused by foreign objects within the grid (e.g., excavators), human accidents and non-utility fires
- **Others**: Utility-based factors such as construction, maintenance, and accidents

2.1.2 Meteorology, Weather and Extreme Events

As stated in Chapter 1, storms on Portuguese mainland are accompanied by heavy rainfall. Based on the irregular distribution of orography and the resulting spatial variability, the rainfall is also unevenly distributed. Influential factors are the Atlantic ocean proximity, latitudinal location and different weather types. Remarkable is the precipitation spread (mean annual precipitation) from over 2,000 mm in the north-west to over 500 mm in the Douro River valley over to the south-eastern. Further, a strong annual cycle characterizes the rainfall, with 40% of annual values occurring in winter months. While, summer months only account for 6% of annual rainfall, the remaining precipitation occurs in the transition months, with again, spatial variability in north and south (Santos et al., 2017).

Next to strong rainfalls, wind gusts are characteristic of storms on the mainland. Due to its exposure to the North Atlantic extratropical storm tracks, with storm activity concentrated in the October–March period that is strongly modulated by the North Atlantic Oscillation (Santos et al., 2017; Soares et al., 2017). Recent events such as Hurricane Leslie (2018) and Storm Kristin (2026), which produced gusts exceeding 200 km/h over the Leiria and Coimbra districts, illustrate that rapidly intensifying cyclones can generate wind loads well beyond standard design values (Instituto Português do Mar e da Atmosfera, 2026c).
Given the objective of disturbance identification, rather than full outages, the extreme wind threshold is set lower at 15 m/s, aligning with Beaufort scale 7, “near gale”, reflecting potential disruptions for consumers.

Portugal’s temperature regime reflects a Csa and Csb climate regime, according to the Köppen–Geiger classification (temperate with dry, warm/hot summer), with again a strong north to south and coast/interior gradient, similar to rain, (Instituto Português do Mar e da Atmosfera, 2026a). Mean summer maxima range from around 25 °C on the northwestern coast to above 33 °C in the inland Alentejo and Douro valley, while winter minima rarely drop below 0 °C on the coast (Santos et al., 2017). Snowfall is confined mostly to elevations above 800–1000 m in the Serra da Estrela and northern mountain ranges and is therefore of limited relevance for the majority of substations in the grid (Instituto Português do Mar e da Atmosfera, 2026a). Absolute extremes reach above 45 °C, with the national record of 47.3 °C observed in Amareleja in August 2003 (Instituto Português do Mar e da Atmosfera, 2026b).

Given the importance of composite events, relevant disturbance systems will be explained, in line with the extreme events as defined by ERSE in the Appendix (A.3).

Atlantic cyclone Based on Portugal’s location, extra-tropical cyclones play a major role in its climate. These types of low-pressure storms, predominantly powered by baroclinic processes, are characterized by strong winds moving from west to east in mid-latitudes combined with rain, thunderstorms and intense winds (Redondo Gonçalves, 2024).
Associated storms with atlantic cyclones, also part of the disturbance dataset, to be introduced in chapter 3.1.1, are Ana (December 2017), Emma (February 2018), Gisele (March 2018), and Helena (January 2019) (Redondo Gonçalves, 2024).
The hazardous impact of atlantic cyclones can be reflected in urban flooding events, land slides and damage of grid infrastructure by falling trees etc.

Heat waves The Portuguese Institute for Sea and Atmosphere (IPMA) definition of heat waves is used, which builds on the World Meteorological Organization (WMO) definition: for a period of at least 5 consecutive days 5 degrees higher than avg. maximum temperature in the referred period. IPMA sets the threshold to 6 days instead of the WMO's 5 days. 2024 was the hottest year on record with eight registered heat waves, again showcasing the demand for further grid resilience. As visible in table 2.1, the majority of heat waves occur in the summer months, although not exclusively (Oliveira et al., 2025).

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Heat Waves	1	3	6	8	16	13	12	17	7	9	2	1

Table 2.1: Composite heat waves per month; Source: (Oliveira et al., 2025)

The increased occurrence of heat waves, as risk itself, further drives the risk of wild fires remarkably, when combined with drought conditions. Given the climate change and resulting change in the precipitation regime, wildfire risks are expected to increase. Recent wildfire events, occurring in 2003, 2005, and 2017 are closely linked to extreme weather conditions. Notably, 97% of recorded extreme wildfires occurred during heat waves, and 90% during drought (Oliveira et al., 2025).

Extreme weather events for the Portuguese Power Grid

Within the study period 2015–2024, 24 extreme weather events, as defined by ERSE could be identified with significant impact on the Portuguese power grid. While the full list of events can be seen in the Appendix A.3, the overview in Table 2.2 shows main disturbance reasons.

Cluster	**Events**	**Total Affected installations**
Storm	17	1066
Wildfire	3	73
Cyclone	2	225
Other	2	76
Total	**24**	–

Table 2.2: Summary of disturbance events by meteorological cluster.

2.1.3 Machine Learning

As ML plays an important role in this dissertation, the following sections briefly describe key concepts on algorithms used in this work.

Binary Classifier: Relevant Algorithms

Before listing relevant binary classifiers, a brief classification of the underlying task of this thesis is required. In supervised learning, an algorithm is given a set of examples, consisting of input features and a known outcome, and learns a rule that maps inputs to outputs. (Hastie et al., 2009). If the target is a category (e.g. "disturbance" vs. "no disturbance"), the task is called classification, whereas a continuous number (e.g. wind speed) as target, is a regression task (Bishop, 2006).

In binary classification, although the goal is to assign observations to discrete categories, algorithms generate continuous predictions $\hat{Y}$. To map these continuous scores back to discrete class labels, a threshold is required to serve as a decision rule. While a threshold of 0.5 is the standard value for defining a linear decision boundary between two classes, this cutoff can be adjusted to navigate the trade-off between sensitivity and specificity (Hastie et al., 2009).

In this work, given meteorological and infrastructure-related features, the model predicts whether a substation experiences a weather-induced disturbance ($y = 1$) or not ($y = 0$). During training, the model adjusts its internal parameters to minimize prediction errors on the training data. Its quality is then assessed on separate, held-out data to verify that it generalizes to unseen situations rather than merely memorizing the training examples (Hastie et al., 2009).

Several ensemble and linear methods were evaluated for the binary classification task. A brief description of each algorithm is provided below.

Random Forest (RF) trains many decision trees in parallel on bootstrap samples, reducing prediction variance via majority vote. RF captures non-linear relationships and is robust to outliers, though less interpretable than a single tree (Breiman, 2001; Hastie et al., 2009).

Gradient Boosting builds trees sequentially, each correcting the residual errors of the previous ensemble (Friedman, 2001):

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \nu\, h_m(\mathbf{x}), \tag{2.1}$$

where $\nu \in (0, 1]$ is the learning rate controlling each tree's contribution. This sequential selfcorrection yields strong predictive performance, though at greater overfitting risk than RF and sensitivity to hyperparameter tuning (Hastie et al., 2009).

Extreme Gradient Boosting (XGB) extends GB with regularization against overfitting (Chen and Guestrin, 2016). Engineering optimizations including approximate split-finding and native missing-value handling make it efficient on large datasets, establishing it as a standard in natural hazard impact modelling (Arora and Ceferino, 2023).

Light Gradient Boosting Machine (LGBM) achieves faster training via gradient-based one-side sampling and leaf-wise tree growth, requiring careful depth control to avoid overfitting (Ke et al., 2017)

Neural networks were excluded due to the limited disturbance event sample size (chapter 4), limited interpretability for non-expert users was an additional consideration. Logistic regression was excluded due to non-linear relationships within weather-related topics.

Binary Classifier: Performance Metrics

Performance is evaluated via the standard confusion matrix (TP, TN, FP, FN), from which the following metrics are derived (Fawcett, 2006): a True Positive (TP) correctly predicts a disturbance, a True Negative (TN) correctly predicts no disturbance, a False Positive (FP) incorrectly raises an alarm, and a False Negative (FN) misses an actual disturbance.
From these four counts, several summary metrics can be computed. Accuracy measures the share of all predictions that are correct:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{2.2}$$

However, accuracy can be misleading when classes are imbalanced. For example, if only 2 % of all observations are disturbance events, a model that always predicts 'no disturbance' would achieve 98 % accuracy while being entirely useless for the task at hand (Fawcett, 2006).
Precision and recall address this limitation by focusing on the positive class. Whereas precision measures what fraction of predicted disturbances are genuine disturbances,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2.3}$$

recall, also called sensitivity, answers how many positives did the model detect out of all disturbances:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{2.4}$$

The Receiver Operating Characteristic Area Under The Curve (ROC-AUC) provides a threshold-independent summary of model performance. It plots the TP rate against the FP rate at every possible decision threshold, and the area under this curve ranges from 0.5 (no better than random guessing) to 1.0 (perfect separation of classes) (Fawcett, 2006). While ROC-AUC is a widely used metric, it can paint an overly optimistic picture when one class heavily outnumbers the other.

The F_β score combines precision and recall into a single number, with a parameter β that controls the relative importance of each:

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{(\beta^2 \cdot \text{Precision}) + \text{Recall}}. \tag{2.5}$$

When $\beta = 1$, both metrics are weighted equally (F_1 score). Setting $\beta = 2$ yields the F_2 score, which places roughly twice as much emphasis on recall as on precision (van Rijsbergen, 1979). In this thesis, the F_2 score is adopted as the primary evaluation metric. Based on the operational circumstances, missing a real substation disturbance (FN) is more costly than a FP that merely triggers precautionary measures at comparatively low cost, given that this thesis is not necessarily a service-team deployment planner. The F_2 score directly reflects this asymmetry by rewarding models that detect as many actual disturbance events as possible, even if this comes with some additional false alarms.

2.2 Recent Research on Disturbance Prediction of Substations

The following paragraphs summarize similar work and research on the field.

Yang et al. (2023) evaluate the vulnerability of the Connecticut distribution grid to meteorological extremes using an integrated framework combining event severity classification with ML-based outage prediction. The study draws on 40 years of ERA5 reanalysis (1981–2020) and 16 years of utility outage data from Eversource Energy, classifying severity into four levels via a Quantile Weighted Distance technique. A hybrid linear combination of RF and GB models, validated by leave-one-storm-out cross-validation across 196 events, yields accuracies of 0.84 and 0.95 for high-impact and extreme-severity events respectively. The scope is, however, limited to heavy rain and strong wind, excluding winter storms and thunderstorms (Yang et al., 2023).

Tervo et al. (2021) predict the damage potential of extratropical storms on the Finnish power grid using an object-based framework, identifying storm polygons where wind gusts exceed 15 m/s and characterizing them with 35 meteorological and non-meteorological features. ERA5 data is combined with a 1.6 km national forest inventory and outage records (2010–2018). To counter extreme class imbalance (0.3%–12% damaging events), five ML algorithms were trained on SMOTE-balanced data. Support Vector Classifiers achieved the best macro-average F1-score of 0.60, with average wind speed and forest characteristics as the dominant predictors (Tervo et al., 2021).

Minnegalieva & Gainullina (2023) present a generalized additive model for substation outage prediction using basic weather variables: wind, temperature, and precipitation, encoded as binary inputs. A purely temporal split (2019–2021 training, 2022 test) and the use of substation name as a unique identifier suggest significant overfitting, casting doubt on the reported accuracy (0.92) and F1-score (0.91). Feature scope and generalizability remain limited (Minnegalieva and Gainullina, 2023).

Martins et al. (2024) combine historical weather data and elevation to build a reference dataset evaluated by an Extra Trees Classifier, producing vulnerability, likelihood, and risk indices over 12×12 km grid cells. The 80/20 split criteria are unspecified, and no training performance metrics are reported (Martins et al., 2024).

E-REDES & SmartWatt (2019–2021) developed a proprietary outage forecasting platform discretising MV and HV line data into 100 m segments with interpolated weather variables. Given outage rates of approximately 3%, a Bayesian-kernel knowledge base was constructed for risk estimation, with neural networks predicting four risk indices across 21 operational areas, updated every 12 hours for a three-day horizon (Almeida et al., 2019; Soares et al., 2021).

The preceding review reveals five consistent limitations motivating this work.
First, models target area-level forecasting for workforce deployment rather than individual substation prediction; geospatial features are underexploited and country-wide models are rare (Yang et al., 2023; Tervo et al., 2021; Martins et al., 2024).Second, predictor sets are narrow or proprietary, restricting reproducibility and transferability (Minnegalieva and Gainullina, 2023; Almeida et al., 2019). Third, sampling strategies rarely support out-of-sample generalization: temporal splits (Minnegalieva and Gainullina, 2023), unspecified criteria (Martins et al., 2024),

and storm-level leave-one-out schemes (Yang et al., 2023) all fail to validate spatial transferability. Fourth, class imbalance is addressed inconsistently, some studies apply SMOTE (Tervo et al., 2021) while others report high metrics without any documented balancing strategy (Minnegalieva and Gainullina, 2023), raising overfitting concerns. Finally, no open-data, reproducible model exists for substation-level weather-risk prediction in Portugal or any European grid. This work addresses that gap with a framework designed for transfer to other European countries where outage data is available.

3 Data Collection, Transformation and Distribution

This Chapter presents per data category, the data collection, data transformation including feature engineering and relevant feature distribution. Thus, the following Chapters are structured in substation data (3.1), weather and hazard data (3.2), and socio-economic data (3.3).
Based on the objective of this thesis, open-source data is preferred, to allow a potential rebuild of these steps for other European countries. Although, feature engineering is a methodological concept, it will be undertaken here to give a coherent overview of data input, data transformation and data output.

3.1 Substation Data

3.1.1 Data Collection

Substation data was obtained from two complementary datasets published by E-REDES on their open data portal (E-REDES, 2026b), supplemented by disturbance records (E-REDES, 2026a) and elevation data from Google Places (Google, 2026).

Primary Substation Layer Metadata for substations were retrieved programmatically via the E-REDES OpenDataSoft web page (E-REDES, 2026b). For each record the server returns the substation name, its geographic coordinates, and a unique `installation_code`, which serves as the primary key throughout the entire pipeline.
In a second scraping request, the supply-area multi-string associated with each substation was fetched. Results were collected and assembled into a `geopandas GeoDataFrame`, a DataFrame with predefined geospatial features, with Point geometry (station location) and an additional multiline string attribute (supply area boundary). Although relevant for this thesis, no further public data on substations could be gained.

From total 399 installations, only 397 installations are carried through all downstream analyses and model training, however in later stages two further stations were excluded due to incomplete geometries.

Disturbance Records Supply-disturbance records, as shown in Table 3.1 were sourced from the E-REDES quality-of-supply dataset as CSV file (E-REDES, 2026a). Each row describes a continuous-disturbance event with start dates, end dates, and observations field (populated for confirmed exceptional events ERSE), and an `installationcode` referencing the affected asset. Disturbances were derived from concrete deviations from the optimal state, of THD, voltage or frequency, as stated by E-REDES.

Installation Code	Busbar	NUTS III	Start Date	Fl. L1	Fl. L2	Fl. L3	Observations
1805S5011300	30 kV	Douro	2021-01-21	98.7	100.0	99.9	Evento Excecional - Depressão Hortense
0103S5006000	15 kV - I2	Região de Aveiro	2018-10-13	99.8	99.4	100.0	Evento Excecional - Tempestade LESLIE
...	...	...	...	...	...	...	...
1824S5009400	15 kV - I2	Viseu Dão-Lafões	2017-12-10	98.1	100.0	100.0	Evento Excecional - Tempestade Ana

Table 3.1: Dataset structure: Substation Disturbance (selection); Source: (E-REDES, 2026a)

The broader definition of disturbances, beyond the extreme-defined values by ERSE, serves to identify any kind of disturbances, while keeping the data availability high, to train models accordingly.

Elevation Data Elevations, in meters above sea level, were obtained via the Google Maps Elevation API (Google, 2026). Coordinates for each installation code were parsed into the API call request and the returned elevation value in metres was appended to the `GeoDataFrame`.

3.1.2 Data Transformation and Feature Engineering

A disturbance label was derived from the disruption records: a binary flag `label` was set to 1 for every installation–date pair associated with a confirmed disturbance event, to 0 otherwise. Where installations with at least one disturbance flag form the *disturbed* subset, remaining installations form the *non-disturbed* pool. The full sampling approach will be explained in depth in Chapter 4.1.1.

Feature Engineering

A set of static spatial features was derived for every substation including distance features, density features and interaction terms of spatial components and substation metadata[1].

Distance features The following reference-point distances were calculated per installation, substation to selected points or lines:

- **Distance to coast (km):** The minimum Euclidean distance from the substation to any of three representative coastal points: Algarve, Lisbon or Porto
- **Distance to geographic centre (km):** The distance to the approximate centroid of mainland Portugal (−8.0° E, 39.5° N)
- **Distance to nearest urban center (km):** The minimum distance to any of eight major cities Lisbon, Porto, Coimbra, Braga, Guimarães, Aveiro, Faro, Viseu
- **Coastal Binary Flag (0/1):** A binary coastal flag was set to 1 for substations within 20 km of the coast
- **Inter-substation proximity (km):** The distance to closest next substation was computed[2], taking the second-nearest neighbor distance to exclude self-matches

Density features A regular 10 km × 10 km grid was superimposed over the full substation coverage. Each installation was assigned to its enclosing cell. Substations-per-grid records the total count of substations in that cell. Additionally, substation-density-5km counts the number of other substations within a 5 km radius of each installation[3].
These density features, in line with the spatial laws of proximity, allow the conclusion of similar behavior of proximate assets (Louro and Ferreira, 2022).

Elevation categories The continuous `elevation_m` attribute was discretized into four categories: lowland (< 200 m), mid-elevation (200–500 m), highland (≥ 500 m), and mountain (≥ 1 000 m), allowing to take different risk tiers (e.g., flood-prone, landslide-prone, . . .) into account. The corresponding binary indicators were added alongside a normalized elevation value[4] used in interaction terms.

[1] using the Portuguese national projected coordinate system (EPSG:3763)
[2] via a *k*-d tree (`scipy.spatial.cKDTree`)
[3] using*k*-d tree ball query
[4] (`elevation_normalized` = `elevation_m` /1000)

Installation Code	Label	Subst. Density (5 km)	Elev. Cat.	Elev. (m)	Elev. Norm.	Dist. Subst. (km)	Dist. Urban (km)
1106S5800800	1	5	0	8.938	0.009	2.617	8.115
0702S5040700	0	0	1	238.194	0.238	21.525	103.501
...	...	...	...	...	...	...	...
0505S5006800	0	0	1	213.239	0.213	20.721	101.354

Table 3.2: Dataset structure: Elevation and density features

Voltage interaction features To capture the structural vulnerability of lower-voltage distribution assets, three interaction features were derived from the substation nominal voltage:

- **15 kV indicator (0/1):** Binary flag set to 1 for substations operating at a 15 kV nominal voltage level
- **Rural 15 kV exposure:** The product of `is_15kv` and the inverted, clipped local substation density ($1 - \min(\text{density}_{5\,\text{km}}, 1)$), approximating 15 kV assets in low-density (rural) areas.
- **Isolated 15 kV distance:** The product of `is_15kv` and the distance to the nearest neighbouring substation, highlighting radially-exposed 15 kV installations.

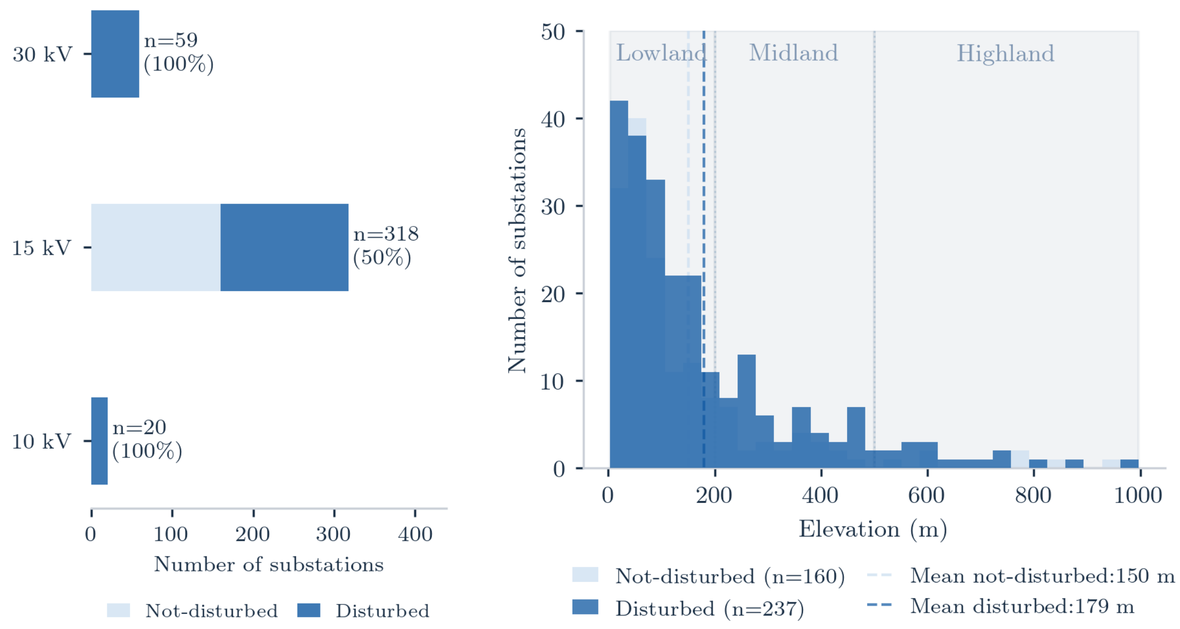
These interaction terms allow the model to separate voltage-class effects from pure geographic isolation, consistent with the proximity-based reasoning used for density features.

3.1.3 Data Distribution - EDA

Substation patterns

Considered installations span three voltage levels, as visible in Figure 3.1a. The 15 kV class dominates ($n = 318$), with approximately 50% experiencing a disturbance, indicating that voltage class alone does not determine risk. Notably, both the 30 kV ($n = 59$) and 10 kV ($n = 20$) classes show a 100% disturbance rate, likely reflecting their larger supply areas and disproportionate exposure to ERSE-classified exceptional events.

The elevation distribution in Figure 3.1b shows a modest but consistent difference: disturbed substations average 179 m versus 150 m for undisturbed ones, consistent with inland and highland assets facing more severe wind and precipitation exposure. Both groups are concentrated below 400 m, with right-skewed distributions.



(a) Voltage class distribution

(b) Elevation distribution

Figure 3.1: Substation pattern distributions by disturbance status: (a) voltage class and (b) elevation.

Spatial patterns

As shown in Figure 3.2, coverage is nationwide without geographic gaps. Disturbed substations (dark blue, $n = 237$) concentrate along the western coastal corridor between Lisbon and Porto, a region of high substation density and elevated Atlantic storm exposure. Non-disturbed installations (light blue, $n = 160$) are more evenly dispersed across interior and southern regions, where extreme weather seems to be less frequent.

The spatial split across subsets is broadly balanced (train: 279, validation: 59, test: 61 installations). Table 3.3 confirms the north-south gradient: disturbance rates reach 75.2% in the Norte region versus 43.0% in the south, consistent with the greater frequency of Atlantic depressions described in Section 2.1.2.

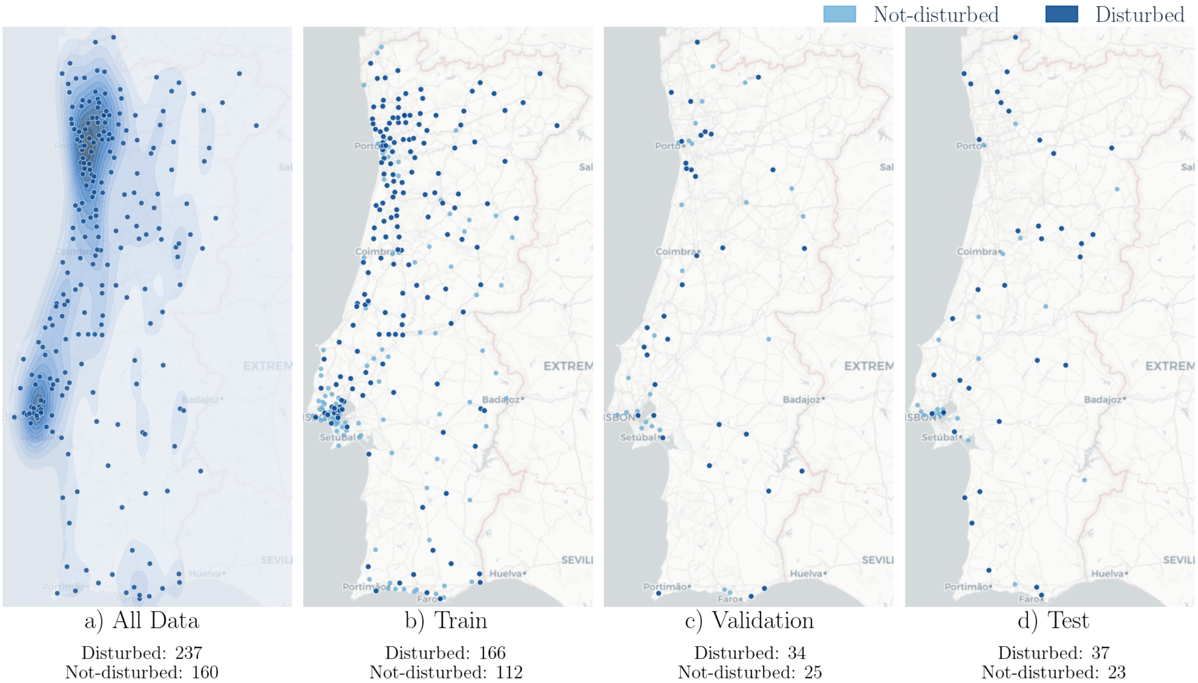


Figure 3.2: Spatial distribution in disturbed vs not-disturbed and per dataset

Category	**Total**	**Disturbed**	**Disturbance rate (%)**
North (NUTS II Norte)	105	79	75.2
Centre (NUTS II Centro)	120	84	70.0
South (NUTS II Alentejo / Algarve)	172	74	43.0
Coastal ($\leq$ 20 km)	98	37	37.8
Inland ($>$ 20 km)	299	200	66.9
Lowland ($<$ 200 m)	290	163	56.2
Mid-elevation (200–500 m)	82	58	70.7
Highland ($\geq$ 500 m)	25	16	64.0
Isolated 15 kV	318	158	49.7

Table 3.3: Disturbance rates by geographic classification

Temporal patterns

Events concentrate in autumn and winter (weeks 40–52 and 1–10), aligning with the Atlantic storm season, while summer months (weeks 22–35) show near-absence of failures, with isolated occurrences likely reflecting wildfire-related disturbances (Figure 3.3).

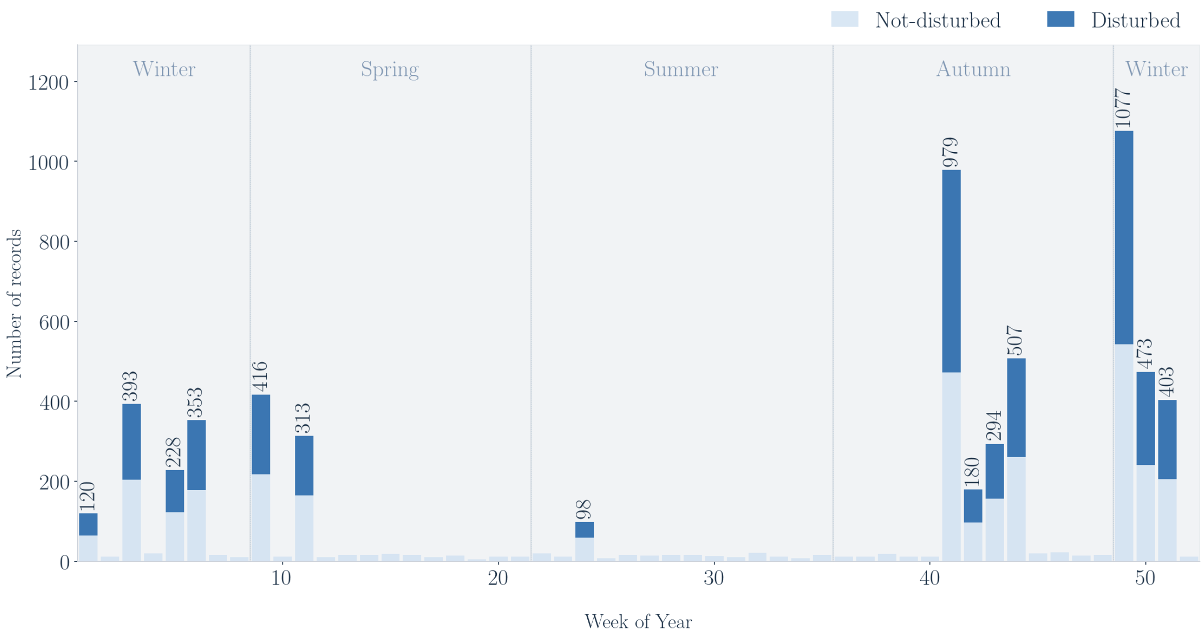


Figure 3.3: Composite yearly distribution of disturbed vs not-disturbed installations

3.2 Weather and Hazard Features

The following subchapters cover the core input for this thesis, historical weather used to assess disturbance events, static hazards linked with it, and forecast weather data for the inference runs.

3.2.1 Data Collection

Weather Data

Historical daily weather data was obtained from the E-OBS gridded observational dataset (Cornes et al., 2018), version 31.0e, provided by the Copernicus Climate Change Service (C3S) at 0.1° spatial resolution (≈9 km). Seven meteorological variables were retrieved for the period 2015–2024 visible in Table 3.4.

Variable	E-OBS Parameter	Unit	Description
Wind speed	fg	$\frac{m}{s}$	Daily mean wind speed at 10 m
Shortwave radiation	qq	$\frac{W}{m^{-2}}$	Downwelling surface shortwave flux
Relative humidity	hu	%	Daily mean relative humidity at 2 m
Precipitation	rr	mm	Daily total precipitation
Mean temperature	tg	°C	Daily mean air temperature at 2 m
Minimum temperature	tn	°C	Daily minimum air temperature at 2 m
Maximum temperature	tx	°C	Daily maximum air temperature at 2 m

Table 3.4: Weather variables from the E-OBS dataset

Missing value imputation The E-OBS gridded product contains structural missing values at grid points located outside the observation network's coverage, predominantly at coastal stations. Prior to substation-level interpolation, missing cells were imputed using a vectorized expanding-radius spatial fill. For each variable and date, three successive passes searched the outer ring at radii $r = 1, 2, 3$ (corresponding to reaches of 0.1°, 0.2°, and 0.3°[5]), averaging all available neighbours in that ring before proceeding outward. This approach, visible in Table 3.5, resolved the majority of missing cells: across the variables, radius-3 coverage reduced residual NAs to below 11,000 cells each (< 0.1% of the full spatio-temporal grid) while maintaining the distribution shape (Figure 3.4).

	NaN count by imputation pass				% remaining
Variable	Initial $r = 0$	After $r = 1$	After $r = 2$	After $r = 3$	
fg	1,125,435	791,095	366,819	7,430	0.7%
hu	1,281,964	872,127	427,427	12,601	1.0%
rr	1,234,714	854,802	423,748	10,959	0.9%
tg	1,234,714	854,802	423,748	10,959	0.9%
tn	1,234,714	854,802	423,748	10,959	0.9%
tx	1,234,714	854,802	423,748	10,959	0.9%
qq	2,533,908	2,230,437	1,851,672	1,438,917	56.8%

Table 3.5: Spatial weather imputation process

Shortwave radiation (qq) exhibits structurally denser coastal gaps that persist after the three-pass spatial fill, leaving 1.44 million cells (approx. 4.2%) unresolved. A temporal forward-and-backward fill of ±3 days was available as a fallback but was not applied, as these residual cells fall outside the Portuguese mainland bounding box and do not map to any substation location.

[5] 9, 18, and 27 km respectively

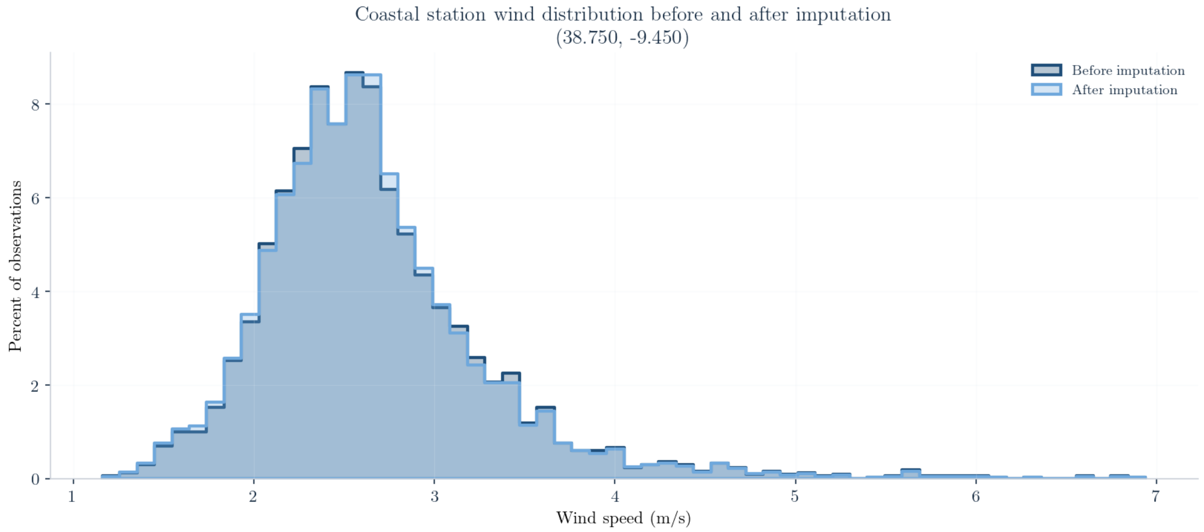


Figure 3.4: Histogram of wind gusts before and after imputation

For GS, weather forecasts were obtained from the Open-Meteo Forecast API using the `jma_seamless` model. The same seven variables were retrieved per grid point, enabling identical feature alignment with the historical reanalysis pipeline, as stated as potential deployment weakness by (Tervo et al., 2021). Further necess pre-processing steps are described in Chapter 4.3.

Hazard Data

Natural-hazard risk indicators at commune level were collected from the ThinkHazard! platform (World Bank, 2021). Commune-level URLs were enumerated programmatically, and for each commune a structured API response was scraped containing risk classifications for nine hazard dimensions (e.g., river flood, urban flood, coastal flood, landslide, wildfire).

District	Commune	Commune coordinates	River flood	Urban flood	Land-slide	Wild-fire
Viseu	Mortágua	POLYGON ((-8.1 40.4, . . .))	Very low	Low	Medium	High
Viseu	São João da Pesqueira	POLYGON ((-7.4 41.2, . . .))	Very low	Very low	Medium	High
Viseu	Mangualde	POLYGON ((-7.5 40.6, . . .))	Very low	Very low	Low	High

Table 3.6: ThinkHazard dataset structure for Viseu

Each classification is expressed as a categorical risk level and accompanied by a commune-level polygon geometry. A total of approx. 26,000 geo-referenced hazard features were obtained, covering the complete mainland.

3.2.2 Data Transformation and Feature Engineering

Weather Data

Grid alignment and tabular conversion All datasets were spatially and temporally subset to the mainland bounding box [6] and the modelling window. Since the seven source grids do not share identical coordinate values, all variables were re-interpolated onto the `fg` grid using bilinear interpolation. The aligned datasets were merged into a single dataset and converted to a flat `pandas` DataFrame with one row per grid point per day. Each grid point was assigned a four-character location identifier (`location_id`) derived from its latitude–longitude pair, ensuring a compact and collision-resistant station key.

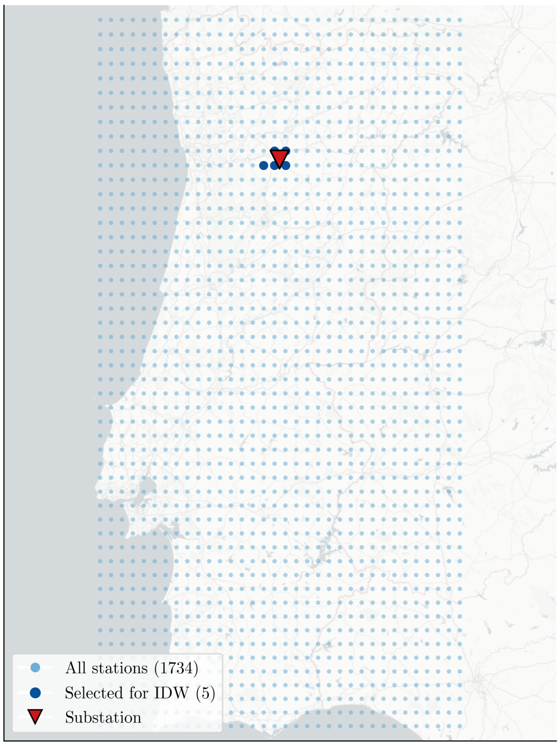
Substation-level interpolation As weather observations are available only at the 0.1° grid nodes (Figure 3.5a), not at the exact substation coordinates, substation-specific weather was interpolated using an Inverse-Distance-Weighting (IDW) approach (Emmendorfer and Dimuro, 2020), utilizing the laws of proximity (Tobler, 1970). For each substation–date combination, the five nearest grid points were identified using a Euclidean distance search.

Substation-level values were then estimated by a truncated IDW with power parameter $p = 2$, to allow closer stations to have more influence than farther ones, a common standard in weather interpolation (Ghomlaghi et al., 2022; Tan et al., 2021):

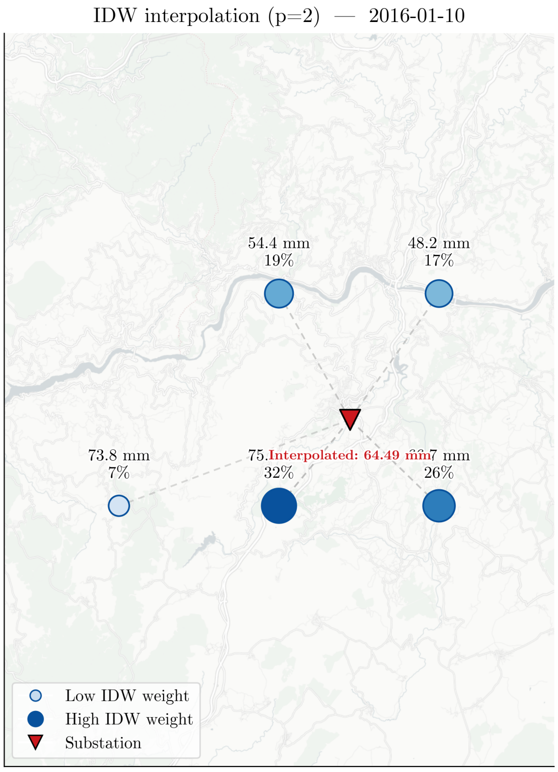
$$\hat{v} = \frac{\sum_{i=1}^{5} w_i\, v_i}{\sum_{i=1}^{5} w_i}, \quad w_i = d_i^{-2}, \tag{3.1}$$

where v_i is the observed value at grid point i and d_i is the Euclidean distance to the substation. A visual representation of the interpolation, using equation 3.1 can be found in figure 3.5b. Given computational constraints and physical logicality, a truncated IDW was chosen. As proximate weather stations are more relevant than farther ones, only the five closest were considered. Binary flags were excluded and recalculated after interpolation, as interpolating binary indicators would distort their physical meaning.

[6](latitude: 37.05°–42.15° N, longitude: 9.45°–6.05° W)



(a) E-OBS grid nodes



(b) Example IDW Interpolation Calculation

Figure 3.5: IDW interpolation visualization

Extreme-weather indicators Further, extreme-weather indicators were introduced to capture deviations that are anomalous relative to the local climate and season. Addressing identified climate-related differences, each weather variable was normalized using a station- and season-specific threshold:

$$z_{v,s} = \frac{v - \mu_{v,s}}{\sigma_{v,s}}, \tag{3.2}$$

where $\mu_{v,s}$ and $\sigma_{v,s}$ are the seasonal mean and standard deviation for variable v and season s, estimated from the full historical baseline at the nearest grid station. With 29 distinct event dates falling predominantly between October and February, this pronounced seasonality justifies the seasonal z-score normalization applied during feature engineering: absolute wind and precipitation thresholds carry materially different physical significance across seasons and regions. The resulting extreme indicator columns encode the local anomaly magnitude. A binary flag was activated whenever the absolute z-score exceeded a seasonally calibrated threshold, ensuring that a cold spell in southern Portugal and a cold spell in northern Portugal receive equivalent scores relative to their respective local norms. Seasons were defined as: winter (December–February), spring (March–May), summer (June–August), and autumn (September–November).

Engineered weather features Additionally, several derived features were constructed from weather variables, as described in Table 3.7. Absolute-threshold indicators flag conditions that carry universal physical significance regardless of location. Interaction terms capture compound hazard, such as simultaneous high wind and heavy precipitation (`storm_index`), or freezing temperatures combined with strong winds (`ice_wind_hazard`). A composite `weather_stress_index` aggregates wind, precipitation, and temperature anomaly contributions into a single severity score bounded to the range $[0, 1]$.

Temporal lookback features Additionally, to capture cumulative weather in the period preceding an event, a 14-day lookback window was applied. For each substation–date record, weather data for the 14 days prior to the event date were interpolated. Rolling aggregations were then computed over 3-day, 7-day, and 14-day windows, yielding features such as cumulative precipitation, maximum wind speed, and consecutive event-day counts for rain days. Only as 'event date' labeled rows were retained in the final dataset.

Feature Group	Examples
Absolute threshold flags	`is_very_hot` ($t_x > 40\,°\mathrm{C}$)
	`is_heavy_rain` ($rr > 20\,\mathrm{mm}$)
	`is_storm_wind` ($fg > 15\,\frac{m}{s}$)
Interaction / compound	`ice_wind_hazard` ($\mathbf{1}_{t_n<2\,°\mathrm{C}} \times fg$)
	`wind_rain_product`
Stress index	`weather_stress_index`
Temporal rolling (3/7/14d)	`rr_sum_Xd`
	`fg_max_Xd`
	`consecutive_rain_days`
	`weather_stress_mean_7d`
Trend indicators	`rr_trend_7d`
	`temp_trend_7d`
Cyclical temporal encoding	`month_sin`
	`week_sin`

Table 3.7: Selected engineered weather features

Hazard Data

Spatial join and encoding For its usage, the commune-level hazard polygons were joined to the substation point layer[7]. Where a substation fell within overlapping polygons, only the first match was retained for reproducibility. The five risk levels were encoded as ordinal integers on a 0–5 scale,*very high* = 5, *very low* = 1, and *No data / missing* = 0.

Hazard aggregation features Additionally, static summary statistics were derived across all hazard columns, including sum, mean, max, and count of dimensions with score ≥ 4. In addition, three binary indicators were created: `is_multi_hazard_site` (at least two high-risk dimensions), `flood_exposure` (at least two high-risk flood-related dimensions), and `is_flood_prone` (combined flood-exposure score > 6).

Weather-hazard interaction features Lastly, interaction terms to represent the activation of static hazard exposure by dynamic weather conditions (Table 3.8) were constructed. For example, `rain_flood_risk` multiplies daily precipitation by the flood exposure score, capturing the joint effect of heavy rain at a site that is structurally prone to flooding. Similarly, `fire_weather` flags

[7] via a point-in-polygon spatial join(`gpd.sjoin`, predicate *within*), after re-projecting both layers to a common CRS

days where maximum temperature exceeds 30°C, relative humidity falls below 30%, and wildfire exposure is elevated, accounting for the wildfire risk, as stated in Section 2.1.2. Elevation-based modifiers are applied where relevant.
A composite hazard activation score aggregates the five primary activation components (flood, landslide, wildfire, coastal, heat) into a bounded [0, 10] score, with a binary flag set once this score exceeds 5.

Feature Group	Example
Static hazard aggregations	`total_hazard_risk`, `is_multi_hazard_site`
Weather×hazard activations	`rain_flood_risk` ($rr\times$ flood_exposure), `fire_weather` ($\mathbf{1}_{t_x>30}\times\mathbf{1}_{hu<30}\times$ wildfire_enc.), `rain_landslide_risk`, `heat_exposure_stress`
Compound & seasonal scenarios	`compound_weather_hazard`, `winter_flood_risk`, `summer_fire_risk`
Overall activation score	`hazard_activation_score` $\in [0, 10]$, `is_hazard_activated`

Table 3.8: Engineered weather-hazard interaction features

3.2.3 Data Distribution – EDA

Weather Features

Weather conditions on disturbed vs. not-disturbed days At installation level, wind speed is significantly lower for disturbed installations than for undisturbed ones ($\bar{x}_d = 2.80\,\frac{m}{s}$ vs. $\bar{x}_{nd} = 3.05\,\frac{m}{s}$; Mann–Whitney U (MWU), $p < 0.001$), while relative humidity follows the same pattern ($\bar{x}_d = 82.4\,\%$ vs. $\bar{x}_{nd} = 85.7\,\%$; $p < 0.001$). However, as shown in geospatial patterns, disturbed installations cluster significantly further north and east than undisturbed ones with wetter and windier environments.

Variable	Unit	$\bar{x}$ (installation level) not-disturbed	$\bar{x}$ (installation level) disturbed	p-value	sig.
Wind speed	$\frac{m}{s}$	3.05	2.80	< 0.001	***
Precipitation	mm	15.68	16.45	0.153	n.s.
Mean temp.	°C	15.31	15.08	0.114	n.s.
Rel. humidity	%	85.66	82.41	< 0.001	***

*** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ n.s. $p \geq 0.05$; MWU on installation-level means ($n_d = 237$, $n_{nd} = 160$).

Table 3.9: Mean weather variables on disturbed vs. not-disturbed installations (installation-level aggregation).

Precipitation and mean temperature show no significant difference at installation level ($p > 0.10$), consistent with this spatial interpretation.

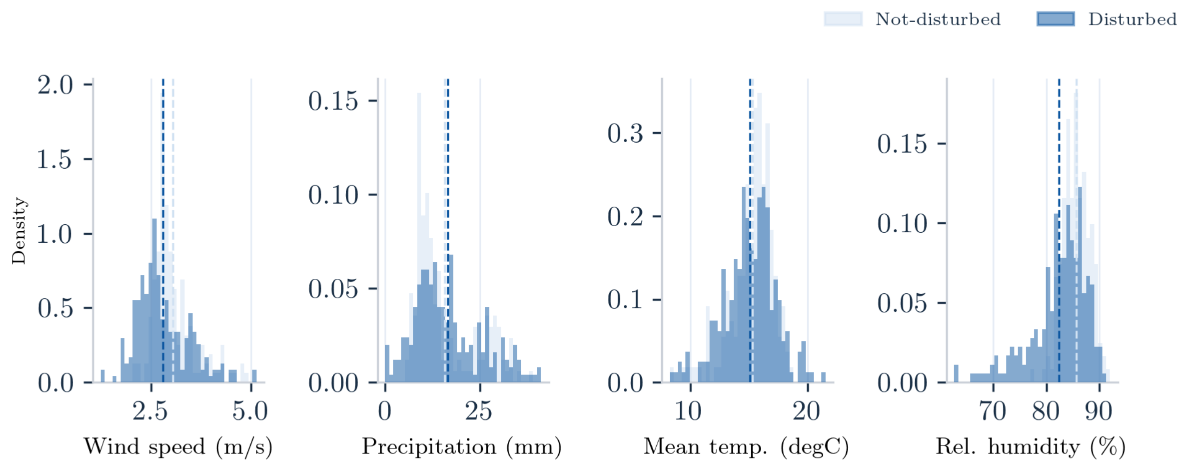


Figure 3.6: Distribution of key weather variables on disturbed vs. not-disturbed days at installation level

Binary weather flag activation rates Heavy rain ($rr > 20$ mm) is the strongest binary predictor, active on 46.6 % of disturbed versus 17.7 % of not-disturbed days ($p < 0.001$). The combined wind-rain extreme flag shows a similarly pronounced uplift (40.2 % vs. 16.9 %), confirming that compound events drive the majority of disturbances. Conversely, very hot ($t_x > 40$ °C) and freezing ($t_n < 0$ °C) days show no significant separation, as both are too rare in the study period to carry predictive power (Figure 3.7).

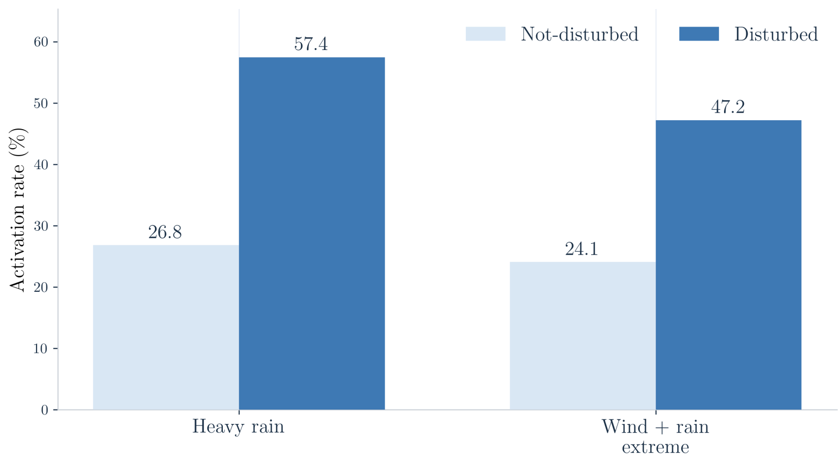


Figure 3.7: Binary flag activation rates: disturbed vs. not-disturbed

Seasonal precipitation patterns and extreme thresholds Figure 3.8 contrasts composite-year precipitation for a northern installation (GONDOMAR) and a southern one (MOURA). The north-south divergence is strong: winter thresholds reach 62.8 mm in the north versus

20.3 mm in the south, with near-zero summer precipitation in MOURA. This directly motivates the seasonal z-score normalization (Section 3.2.2). A fixed universal threshold would generate systematic false positives in the north while missing genuine extremes in the south.

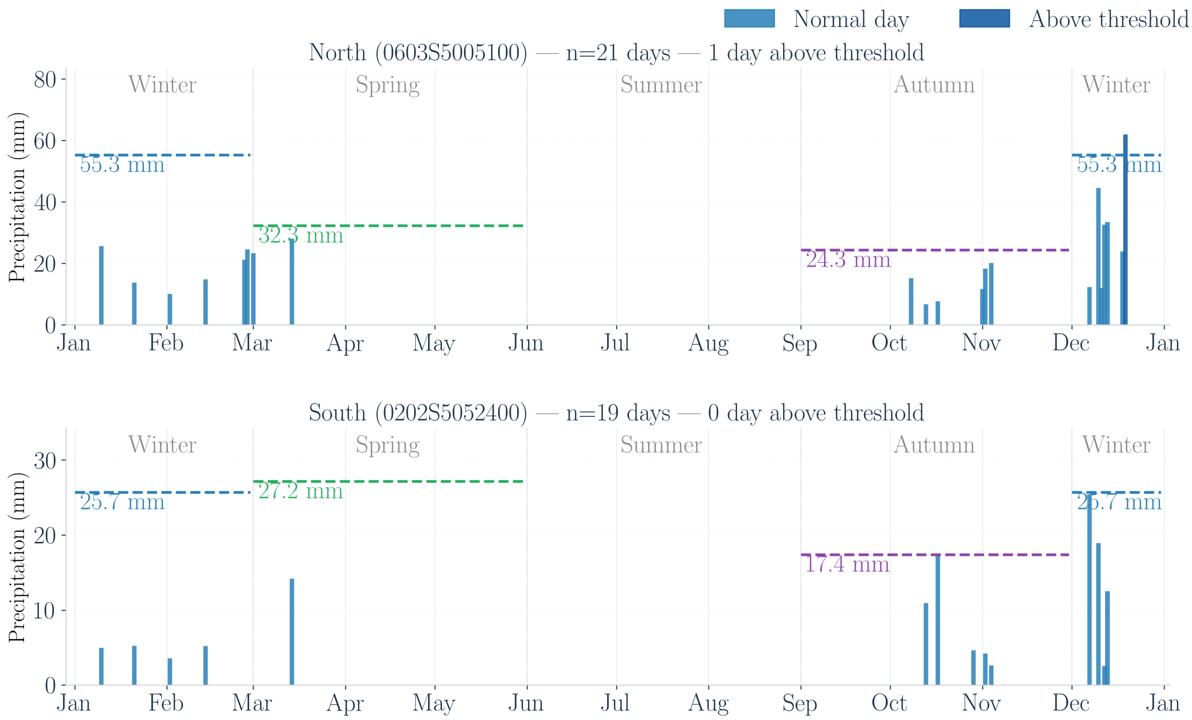


Figure 3.8: Composite-year precipitation with seasonal extreme thresholds for a northern and southern substation

Rolling and lookback features on event days Controlling for this geographic structure, consecutive rain days are significantly lower for disturbed installations than undisturbed ones on the same event days ($\bar{x}_d = 4.39$ vs. $\bar{x}_{nd} = 4.65$; MWU, $p < 0.05$), indicating that rain is not necessarily the strongest driver for disturbances on a national-scale. Maximum 3-day wind speed shows a significant reversal in the same direction as the raw wind confound ($\bar{x}_d = 2.98\ \frac{m}{s}$ vs. $\bar{x}_{nd} = 3.17\ \frac{m}{s}$; $p < 0.001$), again attributable to the geographic split rather than genuine resilience differences. Cumulative 7-day precipitation and consecutive heavy rain days show no significant difference at installation level ($p > 0.10$).

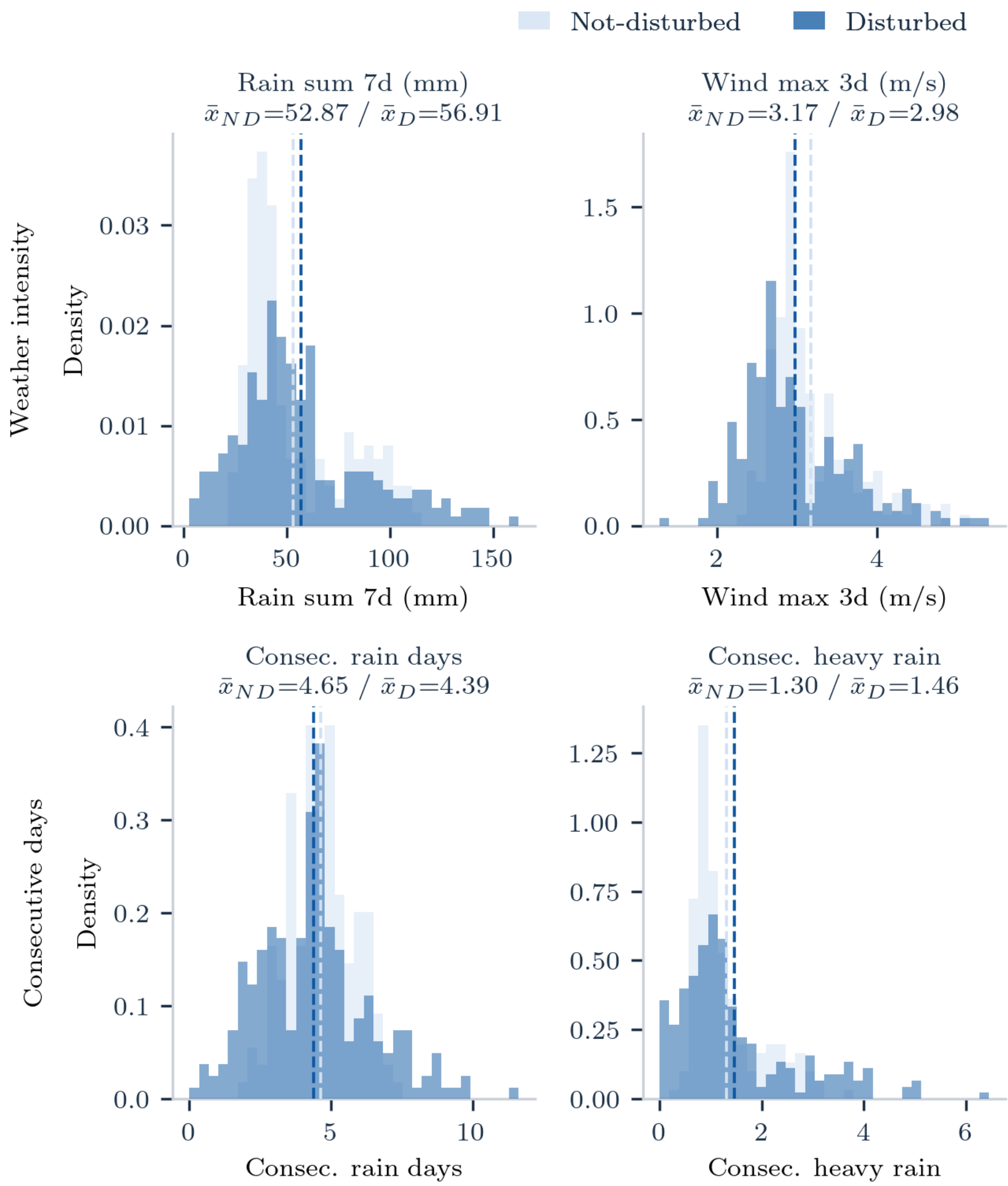


Figure 3.9: Rolling and look back features on event days and mean per group

Hazard Features

Disturbed installations carry a significantly higher mean hazard risk than undisturbed ones ($\bar{x}_d$ = 2.49, $\bar{x}_{nd}$ = 2.21; MWU, $p < 0.001$). Though modest in absolute terms, this difference is consistent across installations and supports the inclusion of static hazard features: a higher baseline exposure score predicts structural vulnerability even without an acute weather event.

Figure 3.10 shows that river flood, urban flood, coastal flood, landslide, and extreme heat dimensions all shift rightward for disturbed installations. All five dimensions are statistically significant at $p < 0.001$ (MWU, installation level), with urban flood and landslide showing the strongest effect sizes ($d = 0.77$ and $d = 0.73$ respectively), followed by river flood ($d = 0.65$), extreme heat ($\bar{x}_d = 2.63$ vs. $\bar{x}_{nd} = 2.22$; $d = 0.59$), and coastal flood. Notably, wildfire exposure shows no significant separation between groups, consistent with wildfire disturbances representing a smaller and more geographically concentrated subset of events. Earthquake and tsunami dimensions are omitted given their negligible role in the disturbance record.

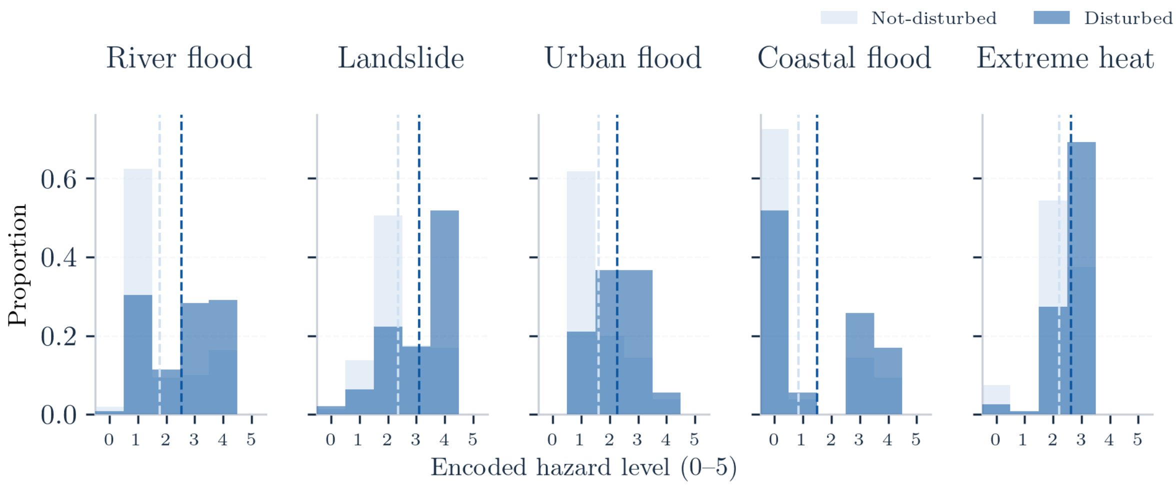


Figure 3.10: Hazard dimension distributions: disturbed vs. not-disturbed

3.3 Socio-Economic Features

Understanding the administrative structure is necessary to explain how socio-economic data is assigned to substations. The encoding of administrative regions in Portugal, as stated by (Santos, 2007) is fundamental for the use of administrative units within the National Statistical System:

1. **Level: Distrito:** district, identified by a two-digit code
2. **Level: Municipio:** municipality, identified by a four-digit code. The first two digits correspond to the district
3. **Level: Freguesia:** parish, identified by a six-digit code. The first four digits correspond to the municipality

3.3.1 Data Collection

Economic data

As an indicator for the economic activities per parish level, enterprise activity data from E-REDES, originally from INE (E-REDES – Distribuição de Eletricidade, S.A., 2025), was used. Records are identified by a numeric Parish Code corresponding to the administrative code system. The dataset displays ten categories of enterprises.

Parish code	Parish	Municipality	Admin. & support	Wholesale & retail	Agric., forestry	Geometry (truncated)
081504	Sagres	Vila do Bispo	48	42	45	`POLYGON ((-69674.5 ...))`
081003	Olhão	Olhão	384	346	232	`POLYGON ((29685.1 ...))`
080506	Montenegro	Faro	288	127	66	`POLYGON ((13917.6 ...))`

Table 3.10: Economic dataset structure

Social data

For the social activities per parish, population data was sourced from the 2021 Portuguese Census published by Statistics Portugal (INE) (Statistics Portugal – Instituto Nacional de Estatística, 2022). The dataset provides demographic indicators at census section granularity, including the primary variable resident population count `N_INDIVIDUOS`. Each census section is identified by a numeric code from which the parish code can be derived. Administrative boundary geometries were obtained from the Official Map Administration Portugal (CAOP) 2021 shapefile, published by DGT (Direção-Geral do Território (DGT), 2021), covering 2,882 parishes on the Portuguese mainland.
Given the latest official census took place in 2021, parish geometries, although adapted slightly, were used from 2021 as well, aligning data points.

Critical Infrastructure

For the critical infrastructure in each parish, relevant Point-of-Interest (POI) data was collected from Open-Street-Map (OSM) via the Overpass API (OpenStreetMap contributors, 2017), yielding 37,113 geo-referenced features across 19 facility categories. Categories span four criticality tiers: life-safety and essential services (hospitals, clinics, fire stations,. . .), social infrastructure

Parish	Municipality	Region (NUTS/District)	N_INDIVIDUOS	ORD
320101	3201	3	794	13835
320101	3201	3	1091	13836
...	...	...	...	...
320101	3201	3	843	13837

Table 3.11: Socio dataset structure

(schools, town halls, government offices, ...), transport nodes, and support services (fuel stations).

Name	Type	Geometry
Ovar	Railway Stop	POINT (-8.6166564 40.8640466)
Parada	Bus Station	POINT (-8.3729706 41.1600948)
...	...	...
Pala	Railway Stop	POINT (-8.0908687 41.1035861)

Table 3.12: OSM dataset structure

3.3.2 Data Transformation and Feature Engineering

Economic data

The dataset was first indexed on parish level, aggregated and absent combinations being filled with zero. `Total_Companies` column was computed as the row-wise sum of all numeric sector columns. Parish codes were standardized as zero-padded six-digit strings to enable a consistent join key. The resulting wide-format DataFrame was merged into the geospatial base layer via a left join on the `dtmnfr_raw` and `Parish Code` fields.

Social data

Data Collection The CAOP shapefile was loaded as a `GeoDataFrame`, column names were standardized, and duplicate `dtmnfr` entries were removed retaining the first occurrence. The census data was imported, rows with null count values were dropped, and all numeric socio-demographic indicators (excluding administrative identifiers) were aggregated to parish level.

This aggregation allows a coherent analysis and comparison with other metrics.

The aggregated census table was joined to the boundary layer via a left merge on the shared `dtmnfr` code. For unmatched parishes, a name-based fallback was applied: the census table was additionally grouped by normalized lowercase municipality and freguesia name strings (`MUNICIPIO DSG`, `FREGUESIA DSG`), and only strictly unique (1:1) name matches were accepted to fill missing population values.

Feature Engineering Further, additional indicators were derived after the merging process: Population density and companies per capita as `companies_per_capita` = `Total_Companies` / `N_INDIVIDUOS`.

Critical Infrastructure

Given the API data import with correct geometry data, no further transformation steps have been conducted.

Feature Engineering OSM-POIs were assigned a numeric tier weight according to the four-tier classification scheme, which will be explained in Section 4.2.2. Features without a recognized category mapping were excluded, retaining only POIs with a positive tier weight. The filtered POI layer was spatially joined to the parish polygon layer[8]. Tier weights were then summed per parish to receive a scalar `poi_weight_in_freg` value representing the cumulative criticality of critical infrastructure present.

[8] using a `within` predicate (`gpd.sjoin`), assigning each POI to its containing parish.

3.3.3 Data Distribution - EDA

As the DIM assesses disturbance impact based on socio-economic environment and critical infrastructure, this section describes the key regional patterns visible in Figure 3.11. All three dimensions, population, enterprises, and critical infrastructure, concentrate in the Lisbon and Porto metropolitan corridors, meaning disturbances to substations in these regions carry disproportionate societal consequence.

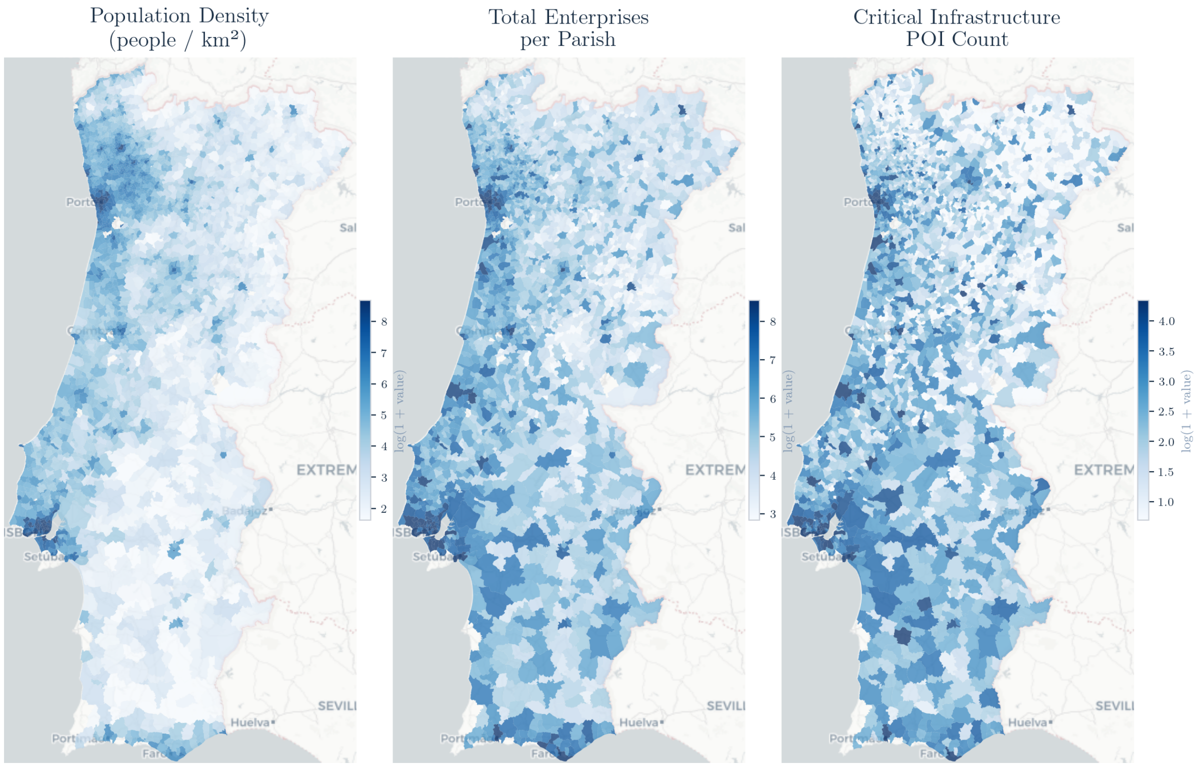


Figure 3.11: Overview: Socio-economic distribution

Enterprise Distribution

Across 1.4 million mainland enterprises, the distribution is highly uneven: the mean parish hosts 502 enterprises but with a standard deviation of 1,275, reflecting pronounced metropolitan concentration. As shown in Table 3.13, Admin and Support (235,573, 16%) and Wholesale and Retail (210,151, 15%) dominate, while energy, water, and extractive sectors, though small in count, carry high electricity dependency and therefore elevated weight in the CI scoring framework. Secondary clusters along the Algarve coastline likely reflect tourism-driven activity.

CAE Sector	Total	CAE Sector	Total
Admin & Support	235,573	Education	63,744
Wholesale & Retail	210,151	Real Estate	62,728
Consulting & Science	150,645	Transport & Storage	52,271
Hospitality	117,565	Arts & Recreation	45,029
Health & Social	113,665	ICT	32,781
Agriculture	109,126	Energy & Utilities	5,017
Construction	104,235	Water & Waste	1,244
Other Services	73,251	Extractive Industry	967
Manufacturing	67,192		

Table 3.13: Total companies across all parishes by CAE sector

Population Distribution

Of Portugal's approximately 10.3 million inhabitants, 9.8 million (95%) reside on the mainland and are considered here. Population is strongly concentrated in Greater Lisbon and Greater Porto, with secondary clusters in Coimbra, Braga, Aveiro, and Faro. The high dispersion (mean 3,419, SD 7,375 per parish) reflects a thin rural population base against a small number of dense urban parishes, of which Agualva-Mem Martins in Sintra is the largest with 66,000 residents. Relative to the enterprise layer, coastal parishes show comparatively higher population density, relevant for the coastal flood exposure assessment in the CI.

Critical Infrastructure Distribution

37,000 POIs are considered across 19 facility categories. As visible in Figure 3.12, the social tier dominates in raw count (21,100, led by schools at 7,800, while the life-safety tier, though smaller, carries the highest CI weights given direct power-continuity dependency. The transport tier is geographically sparse but concentrated at major nodes, making individual substations serving airports or railway stations disproportionately critical regardless of served population.

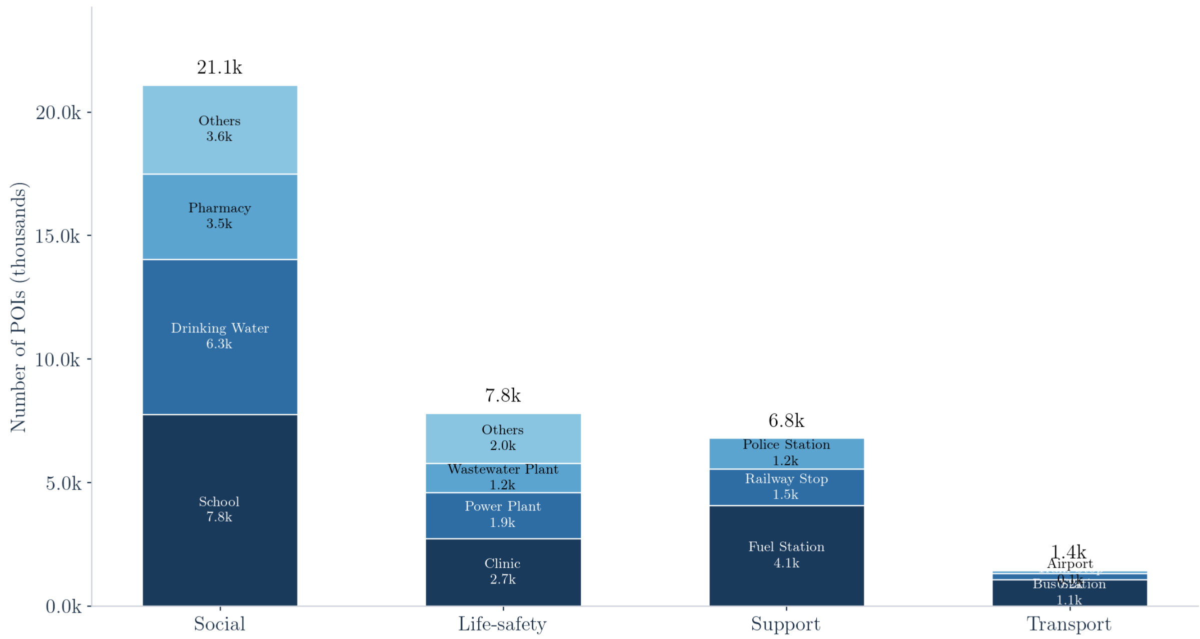


Figure 3.12: Overview: Breakdown of POI in criticality tiers

Data Summary

This thesis integrates nine datasets from seven sources across three feature categories: substation characteristics (E-REDES, Google Maps Elevation API), weather and hazard data (Copernicus E-OBS v31.0e, World Bank ThinkHazard!), and socio-economic features (E-REDES enterprise data, INE Census 2021, OpenStreetMap). In total, 114 candidate features were engineered from these inputs.

4 Results and Real-World Application

Building on the data collection and feature engineering presented in Chapter 3, this chapter develops the modelling framework and its operational deployment. Section 4.1 presents the DPM, covering the sampling approach, data preprocessing, model development, and test set evaluation. Section 4.2 develops the complementary CI, ranking substations by their socio-economic and critical-infrastructure exposure. Both models are operationalized in Section 4.3. The modelling task is decomposed into two complementary components: Model A (DPM), a binary classifier predicting whether a substation will experience a weather-driven disturbance, and Model B (DIM), a deterministic scoring framework ranking substations by their societal and economic consequence in case of disturbance. Together, they ensure that operational risk prioritization reflects both disturbance likelihood and potential impact.

4.1 A) Disturbance Prediction Model

This Section describes the DPM. After stating the sampling approach (4.1.1), the necessary data pre-processing steps (4.1.2), the model development and fine tuning (4.1.3), the model results will be stated (4.1.4, 4.1.5, 4.1.6)

4.1.1 Sampling approach

Methodological Trade-offs and Dataset Characteristics

As summarized in Table 4.1, first the installation pool was split, based on the existence of one disturbance event, distributed all 397 substations across subsets. Within each subset, the disturbed-to-non-disturbed ratio varies due to random assignment, with training containing 59.8% disturbed installations, validation 57.6%, and test 62.2%. This natural variance across subsets, arising from random partitioning of a finite installation pool, does not compromise evaluation validity as all subsets maintain representation of both installation classes.

	Training		Validation		Test	
Installation Class	**Count**	**%**	**Count**	**%**	**Count**	**%**
Disturbed installations	167	59.8	34	57.6	38	62.2
Non-disturbed installations	112	40.2	25	42.4	23	37.8
Total installations	**279**	**100.0**	**59**	**100.0**	**61**	**100.0**
Subset proportion	69.9%		14.8%		15.3%	

Table 4.1: Installation distribution across training, validation, and test subsets

Installation-Level Generalization vs. Temporal Prediction

The installation-level split changes the prediction task from temporal forecasting to a risk assessment. Traditional time-series approaches, as stated in Section 2.2, employ chronological temporal splits evaluating the model's ability to predict future disturbances of known installations based on their historical patterns. In contrast, the chosen methodology evaluates the model's ability to assess disturbance risk for stations without prior disturbance history. All three subsets contain samples spanning the entire temporal range (2015–2024), with training utilizing 498 unique dates[1], validation 128 dates, and test 134 dates from the available pool.

Critically, each subset contains samples from nearly all event dates: installations of training and validation subsets experienced disturbances on 28/29 event dates (96.6%), whereas the test subset has 23/29 dates (79.3%) coverage. Missing dates are caused by no disturbances of allocated installations. This almost complete date coverage ensures that all subsets are exposed to the full range of extreme weather conditions, preventing confounding between installation assignment and meteorological exposure.

This sampling approach aligns with the practical deployment scenario: the model must generalize to installations unseen during training (newly commissioned substations or existing ones without disturbance history) only on physical and environmental characteristics.
However, this approach sacrifices the ability to model temporal trends in disturbance rates or aging effects within individual installations. Table 4.2 presents the repeated-disturbance structure observed across subsets. Disturbed installations in the training subset experienced an average of 12.7 disturbance events, similar to the validation subset (13.0) and higher than the test subset (7.7). This structure indicates that certain installations are systematically vulnerable across multiple extreme weather events, but the installation-first split prevents the model from

[1] including disturbance events and non-disturbance

learning individual installation specific properties.

Metric	**Training**	**Validation**	**Test**
Type A samples (disturbances)	2,122	443	291
Disturbed installations	167	34	38
Avg. disturbances per installation	12.7	13.0	7.7
Event date coverage	96.6%	96.6%	79.3%
Total unique dates	498	128	134
Event dates	28	28	23
Calm dates	470	100	111
Temporal span (days)	3,641	3,641	3,641

Table 4.2: Disturbance event frequency and temporal coverage across subsets

Three-Type Sampling and Dual Learning Signals

Traditional binary classification approaches for infrastructure disturbance or failure prediction sample only from event dates, creating positive samples from disturbed installations and negative samples from non-disturbed installations. This provides a single dimension of comparison: different installations under identical (extreme) weather conditions. While informative for learning installation vulnerability patterns, this approach conflates weather effects with installation characteristics. A model trained exclusively on event dates may learn that certain substations fail without understanding that this installation only gets disturbed under specific meteorological thresholds.

Thus, to address this traditional approach, this thesis uses a dual learning signal operating as follows:

1. **Weather threshold learning (Type A vs. Type B):** Holding the installation constant, the impact of weather conditions changes on disturbance events is being assessed. The training subset contains 2,122 Type A samples (disturbed installations on event dates) compared against 501 Type B samples (identical disturbed installations on calm dates). Each of the 167 disturbed installations in training appears exactly 3 times in Type B samples (sampled from 3,509 available calm dates[2]), ensuring diverse meteorological exposure beyond disturbance events. The 470 unique calm dates utilized in training span the entire study period, capturing seasonal variability in baseline weather conditions.

   By comparing identical installations under extreme weather (Type A) versus calm weather (Type B), the model isolates the causal effect of meteorological variables. For example, if one substation gets disturbed on 15 dates when wind speeds exceed 80 km/h but will not be affected on 3 sampled calm dates when wind speeds remain below 40 km/h, the model learns the wind speed threshold rather than merely learning the vulnerability of said station. This explicit contrast prevents overfitting to installation identities.

2. **Vulnerability assessment (Type A vs. Type C):** Additionally, holding weather conditions constant (same event date), the vulnerability of installation characteristics is being assessed between labeled installations. Type C samples comprise non-disturbed installations observed on the same 29 event dates as Type A disturbances. The training subset contains 2,076 Type C samples matched to the 2,122 Type A samples, providing near-parity in extreme weather exposure across installation classes. These Type C samples are distributed across 112 non-disturbed installations, yielding an average of 18.5 samples per installation. This is substantially higher than the Type B sampling rate (3 per installation), reflecting the deliberate matching strategy to ensure equal representation of surviving installations across all event dates.

The three-type approach decomposes the classification problem into two orthogonal learning tasks with quantifiable sample allocations detailed in Table 4.3.

[2]The study period spans 3,641 calendar days (2015-01-01 to 2024-12-31). Of these, 29 constitute event dates on which at least one substation disturbance was recorded. The remaining 3,509 calm dates available for Type B sampling are not simply 3,641 − 29 = 3,612, as two additional exclusion rules apply: (1) a ±3-day buffer around each event date removes days immediately preceding or following a storm, preventing near-event weather conditions from contaminating the calm pool; and (2) days with insufficient weather data coverage at substation locations after spatial interpolation are discarded. Together these rules reduce the usable calm pool from 3,612 to 3,509 days.

	Training		Validation		Test	
Sample Type	**Count**	**%**	**Count**	**%**	**Count**	**%**
Type A (disturbed × event)	2,122	45.2	443	45.3	291	41.9
Label	1		1		1	
Installations involved	167		34		38	
Unique dates	28		28		23	
Type B (disturbed × calm)	501	10.7	102	10.4	114	16.4
Label	0		0		0	
Installations involved	167		34		38	
Calm dates per installation	3		3		3	
Unique calm dates	470		100		111	
Type C (non-disturbed × event)	2,076	44.2	432	44.2	289	41.6
Label	0		0		0	
Installations involved	112		25		23	
Avg. samples per installation	18.5		17.3		12.6	
Total samples	**4,699**	**100.0**	**977**	**100.0**	**694**	**100.0**
Positive samples (%)	45.2%		45.3%		41.9%	
Balance ratio (neg:pos)	1.21:1		1.21:1		1.38:1	
Learning signal ratios						
Type A : Type B (extreme:calm)	4.2:1		4.3:1		2.6:1	
Type A : Type C (disturbed:surviving)	1.0:1		1.0:1		1.0:1	
Type B : Type C : Type A	11:44:45		10:44:45		16:42:42	

Note: Type B samples reuse the same disturbed installations as Type A but on different (calm) dates. Type C samples are approximately matched to Type A counts to ensure equal representation across event dates.

Table 4.3: Sample type composition and learning signal distribution

This dual-signal structure mirrors the causal structure of infrastructure disturbances: $P(\text{disturbance} \mid \text{installation, weather})$ depends on both installation vulnerability and weather severity. By pairing each Type A disturbance (2,122 samples) with both calm-day observations of the same installation (501 Type B, weather signal) and surviving installations under identical storm conditions (2,076 Type C, vulnerability signal), the model learns both drivers explicitly rather than inferring them indirectly.

Temporal coverage spans both event and calm periods across all subsets. The 5–6% event-date representation, reflecting the 29 event dates out of 3,538 available, forces the model to distinguish the rare conditions that cause disturbances from the prevalent calm conditions under which even vulnerable installations remain undisturbed, directly targeting false-alarm reduction in deployment.

The within-subset reuse of the 167 disturbed installations across Type A and Type B samples is intentional: the same installation appears with label = 1 under extreme weather and label = 0 under calm conditions, preventing the model from treating installation identity as a sufficient predictor and forcing it to attend to meteorological features.

4.1.2 Data Preprocessing

Missing values in substation metadata were left empty, as imputation would introduce bias given the sparse nature of operational attributes. Missing weather data were imputed, as described in Section 3.2.1.

Categorical columns were one-hot encoded.

Feature Selection and Multicollinearity Reduction

After the initial feature engineering phase, 114 candidate features (Appendix A.4) were produced. To prevent multi-collinearity-driven overfitting, features exhibiting variance inflation factors exceeding 10 or pairwise correlations above 0.85 were systematically removed. The reduction process prioritized retention of: (1) physically interpretable features (e.g., `rr_sum_7d`, (2) features with established meteorological significance in the literature, and (3) composite indices that aggregate multiple signals (e.g., `wind_coastal_risk`). The final feature set with 26 features can be seen in Table 4.4 across seven categories: core weather measurements, temporal encodings, spatial characteristics, threshold flags, hazard interactions, temporal memory features, and voltage class indicators.

Category	Count	Features
Raw Weather	4	rr, tx, hu, fg
Temporal encoding	3	month_sin, month_cos, week_sin
Spatial	6	lat, lon, elevation_normalized, distance_to_nearest_substation, substation_density_5km, is_coastal_region
Threshold flags	2	is_heavy_rain, is_autumn
Hazard interactions	3	wind_coastal_risk, rain_landslide_risk, heat_fire_risk
Temporal memory	4	fg_max_3d, rr_sum_7d, consecutive_rain_days, tx_max_7d
Voltage / installation[3]	4	isolated_15kv; nominal_voltage (v)
Total	**26**	

Table 4.4: Feature categories in final GB model

Data Scaling

Numerical features were standardized using StandardScaler fitted exclusively on the training subset to prevent data leakage.
Binary flags, cyclical temporal encodings, and one-hot voltage columns were excluded from scaling to preserve their bounded interpretability.

4.1.3 Model Development and Model Fine Tuning

Baseline Model Comparison

As described in Section 2.1.3, the following algorithms were compared: RF, XGB, LGBM, and GB. All models were trained with default hyperparameter (Appendix A.7) on the training set and evaluated on identical validation and test splits. The GB results, as shown next to the other families in Table 4.5, with the highest accuracy, second-best precision and F2 score seems to be the best-base performing model. However, considering the ROC-AUC score, XGB and LGBM outperform the GB model.
When assessing the ROC-AUC, visualized in Figure 4.1, of the validation set, especially between the FP-Rate 0.2–0.5 the observable dip indicates heterogeneous negative samples: Type B

Model	Accuracy	Precision	Recall	F1	F2	ROC-AUC	Avg. Precision
GB	0.8084	0.8257	0.6886	0.7509	0.7122	0.8542	0.7751
Random Forest	0.7649	0.7725	0.6228	0.6897	0.6479	0.8285	0.7960
XGBoost	0.7954	0.7803	0.7128	0.7450	0.7254	0.8781	0.7978
LGBM	0.7997	0.8268	0.6609	0.7346	0.6885	0.8823	0.8219

Table 4.5: Baseline model performance on test set

(disturbed installations on calm dates) are easily separated via weather features, while Type C (non-disturbed installations on event dates) overlap with Type A positives in score space, as separation relies solely on installation vulnerability features. This pattern is amplified by the installation-level split, which forces the model to generalize vulnerability to unseen installations without exploiting installation identity. Further, the relatively small validation subset further accentuates the curvature.

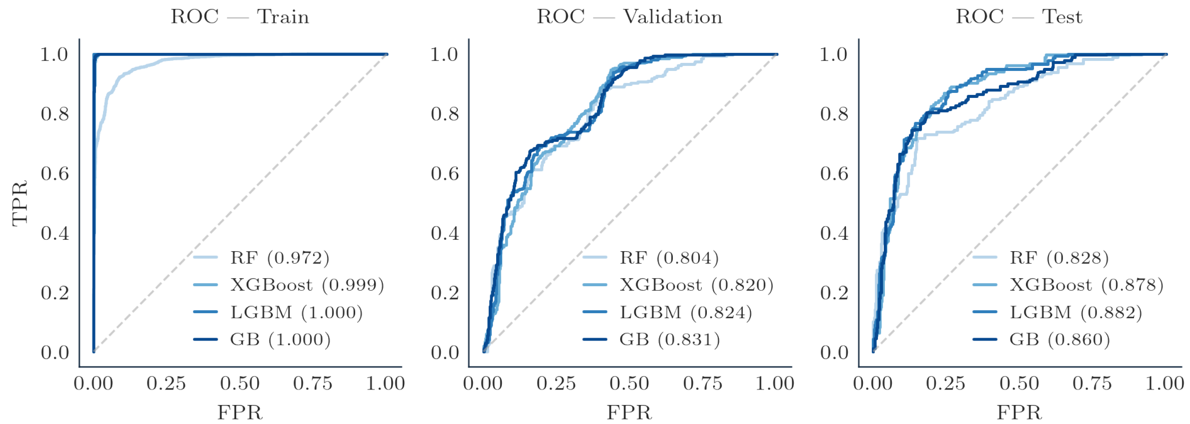


Figure 4.1: ROC Curves Base Model Comparison

At baseline, the GB delivers the strongest overall test performance (highest accuracy and second highest F2), LGBM achieves the highest precision, ROC-AUC, and average precision, and XGB provides the highest recall and F2-Score. These differences indicate model-specific trade-offs between outage detection sensitivity and false-alarm control, motivating the subsequent hyperparameter optimization phase, with primary focus on the GB model.

Hyperparameter Optimization

Hyperparameter were optimized via `RandomizedSearchCV` with 5-fold stratified cross-validation over 500 iterations per search space, using F2-score as the optimization metric. As detailed in

Table 4.6, GB Final outperforms the alternative configurations by combining the broadest yet most parsimonious search space: widening the core parameter ranges without inflating search dimensionality allowed denser sampling of the relevant hyperparameter region. In contrast, GB V2, explicitly zoomed into the top-performing region of GB V1, inherited its predecessor's suboptimal neighbourhood, illustrating that iterative narrowing around an imperfect optimum can be counterproductive when the search space is misaligned with the true optimum. The full parameter grids can be found in the Appendix (A.6)

Model	Acc.	Prec.	Recall	F1	F2	Test AUC	Avg. Prec.	Train-Val Gap	TN	FP	FN	TP
GB final	0.8012	0.8248	**0.6678**	0.7380	**0.6942**	**0.8677**	**0.8238**	**0.1686**	359	41	**96**	**193**
GB V3	**0.8055**	**0.8414**	0.6609	**0.7403**	0.6905	0.8598	0.7927	0.1691	**364**	**36**	98	191
GB V1	0.7910	0.8194	0.6436	0.7209	0.6725	0.8580	0.7908	0.2088	359	41	103	186
GB V2	0.7837	0.8125	0.6298	0.7096	0.6594	0.8467	0.7714	0.2039	358	42	107	182

Table 4.6: GB variant comparison across test performance

Given the overfitting, as visible in Figure 4.1, regularization parameters[4] were intentionally constrained to conservative ranges to combat the overfitting patterns observed in baseline runs, as visible in the parameter search space found in the Appendix A.6. The final optimized GB configuration is presented in table 4.7, although overfitting has not been eliminated fully, the gap was reduced, as shown in Table 4.6.

Hyperparameter	Optimized Value
`n_estimators`	106
`max_depth`	9
`learning_rate`	0.07884
`min_samples_split`	21
`min_samples_leaf`	13
`max_features`	sqrt
`subsample`	0.63538
`ccp_alpha`	0.00001036
`min_impurity_decrease`	0.00304
`random_state`	28

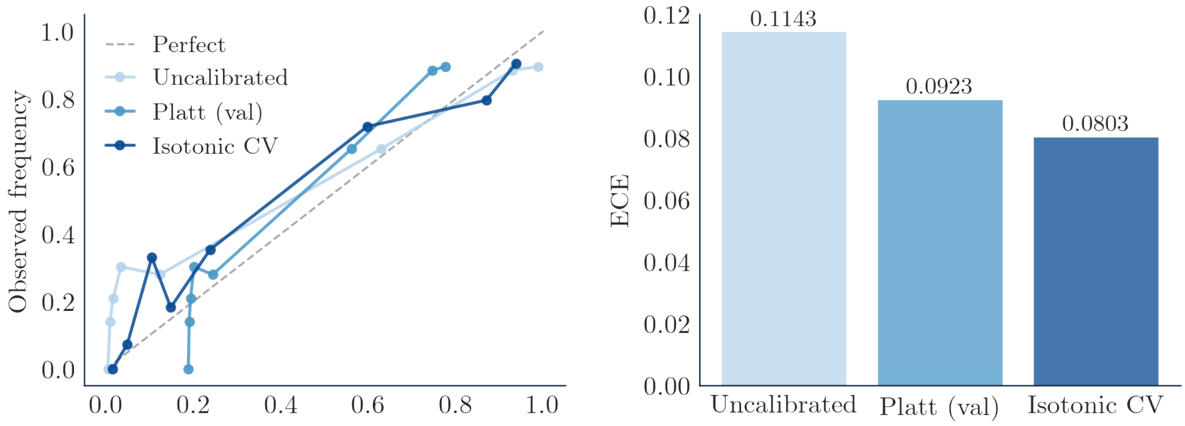
Table 4.7: Final GB model configuration after hyperparameter optimization

Inherently, the minor improvement driven by the hyperparameter tuning, showcases the difficult task of weather-related disturbance prediction, especially in combination with limited infrastructure data.

[4](`min_samples_split`, `min_samples_leaf`, `ccp_alpha`)

Probability Calibration

To produce calibrated disturbance probabilities rather than raw scores, isotonic regression calibration[5] and Platt scaling were fitted on the training set and compared via reliability diagrams. Isotonic calibration reduces the expected calibration error by 30% relative to the uncalibrated model, as visible in Figure 4.2, though residual overfitting persists across most probability bins. Nonetheless, the isotonic model was selected for all final predictions and threshold optimization.



(a) Reliability curve (b) Expected calibration error

Figure 4.2: Final GB model calibration curve

Decision Threshold Optimization

Based on the binary-classification task, the decision threshold plays an important role, as described in Section 2.1.3. Thus, the threshold was optimized to address the objective of this thesis with a function on the validation set, sweeping thresholds from 0.05 to 0.95 in 0.01 increments. Four candidate thresholds were identified for the calibrated model:

- **Best F1 threshold (0.13)**: Maximizes F1-score, achieving 60.2% precision and 95.0% recall on validation
- **Best F2 threshold (0.11)**: Prioritizes recall, favoring outage detection over false alarm minimization, achieving 58.4% precision and 96.8% recall
- **Target recall threshold (0.19)**: First threshold achieving the minimum 80.0% recall target on the validation set, yielding 62.0% precision and 87.0% recall
- **Trade off: FN & FP (0.40) - Selected approach**: Prioritizes balance in between FN and FP, with recall of 70.9% and precision of 82.0%

[5] `CalibratedClassifierCV`, method = `isotonic`, cv = 5

Based on the objective and the targeted users of the final model, a trade-off threshold was selected to avoid unnecessary concern or false alarms of potential disturbance while keeping the warning potential high.

4.1.4 Test Set Performance

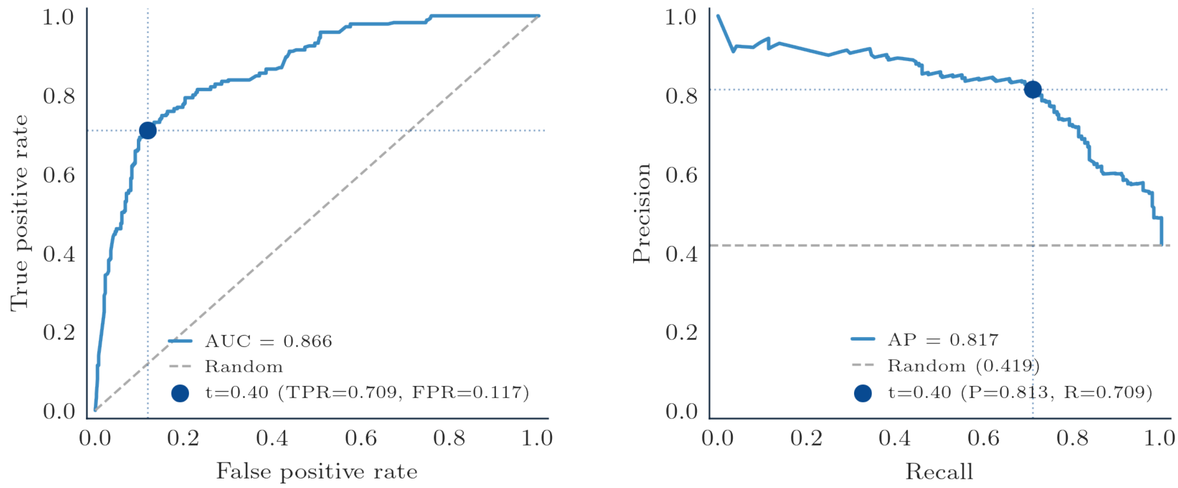
When assessing the final model's performance (Table 4.8) with a confusion matrix on the held-out test set (n = 689 samples from 61 installations). The model achieves an outage detection rate of 70.9% (205/289) with a false alarm rate of 11.3% (45/400). The 84 missed outages (FN) require further investigation, which will be discussed in Section 4.1.5.

When comparing the base GB model and the final, fully-optimized GB model, recall could be improved by 7.3%, F2-Score by 5.6% and reducing missed disturbances by 14.3%. When comparing the final model with the pre-optimized model, F2 could be further improved by 5%.

	Predicted			Metric	Value	Count
Actual	**No Dist.**	**Dist.**	**Total**			
No Disturbance	355 (TN)	45 (FP)	400	Accuracy	81.3%	560/689
Disturbance	84 (FN)	205 (TP)	289	Precision	82.0%	205/250
Total	439	250	689	Recall (Outage Detection Rate)	70.9%	205/289
				F1-Score	76.1%	—
				False Alarm Rate	11.3%	45/400
				Missed Outages	—	84

Table 4.8: Confusion matrix and performance metrics on the test set

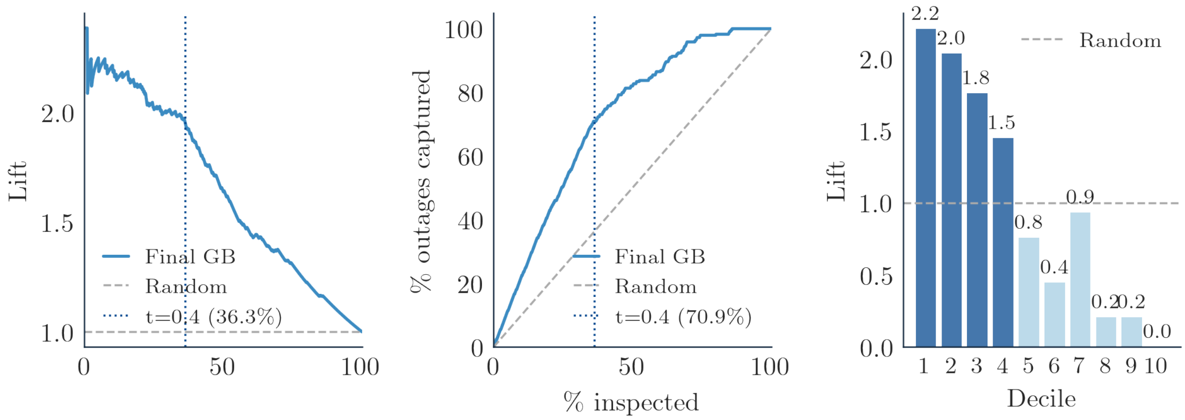
Figure 4.3 presents the ROC and precision-recall curves on the test set. Precision remains above 0.80 up to a recall of approximately 0.70, after which it drops sharply. The selected operating point at $t = 0.40$ ($P = 0.813$, $R = 0.709$) sits directly on this plateau, confirming it as a well-justified trade-off. With an average precision of 0.817 against a random baseline of 0.419, the model nearly doubles the baseline, confirming that discriminative performance is a property of the calibrated probability ranking rather than an artefact of the chosen threshold.



(a) ROC curve
(b) Precision–recall curve

Figure 4.3: ROC and PR curves on test set of final GB

The lift curve analysis (Figure 4.4) demonstrates the model's ability to prioritize high-risk installations over random selection. Inspecting only the top 36.3% of installations by predicted probability already captures 70.9% of all disturbances, yielding a cumulative lift of approximately 2.0. The highest-risk decile alone achieves a lift of 2.21, with the top three deciles all exceeding 1.45×. Lift drops below the random baseline only beyond the 59th percentile, confirming that discriminative power is concentrated in the upper risk tiers.



(a) Cumulative lift
(b) Cumulative gain
(c) Lift per decile

Figure 4.4: Lift Curve Analysis test set of final GB

4.1.5 Feature Importance Analysis

Feature importance was assessed via two complementary metrics: Gini importance (training-time) and permutation importance (mean F1 drop after random feature shuffling on the test set). Table 4.9 presents the top 15 features ranked by permutation importance alongside their Gini scores. A Spearman rank correlation of 0.703 between the two metrics indicates broad agreement in feature ordering despite differences in absolute magnitude.

Rank	Feature	Perm. Imp.	Gini Imp.	Interpretation
1	`isolated_15kv`	0.063	0.088	Isolated 15 kV substation flag
2	`rr`	0.047	0.081	Daily precipitation (raw)
3	`lat`	0.043	0.130	Geographic latitude
4	`month_cos`	0.025	0.020	Seasonal cyclical encoding
5	`nominal_voltage_15kV`	0.022	0.061	Voltage class: 15 kV (one-hot)
6	`week_sin`	0.016	0.017	Weekly cyclical encoding
7	`tx_max_7d`	0.015	0.019	7-day maximum temperature
8	`tx`	0.012	0.025	Daily maximum temperature
9	`distance_to_nearest_substation`	0.012	0.059	Inter-substation spacing
10	`consecutive_rain_days`	0.011	0.018	Consecutive days with rainfall
11	`month_sin`	0.011	0.006	Seasonal cyclical encoding
12	`heat_fire_risk`	0.010	0.021	Heat and fire risk index
13	`is_autumn`	0.010	0.003	Autumn season flag
14	`rr_sum_7d`	0.006	0.049	Cumulative 7-day precipitation
15	`elevation_normalized`	0.005	0.056	Terrain elevation (0–1 scale)

Table 4.9: Top 15 features by permutation importance on test set. Gini importance reflects each feature's impurity-based importance score from the full ranking.

A notable discrepancy exists between the two metrics for spatial features: `lat` and `lon` rank first and second by Gini importance but fall to ranks three and 19 by permutation importance, indicating that spatial splits contribute strongly to training-set discrimination but partially reflect autocorrelation rather than fully generalizable patterns.

The top-ranked feature by permutation importance, `isolated_15kv` (ΔF1 = 0.063), reflects the disproportionate vulnerability of topologically isolated 15 kV substations lacking neighbouring redundancy. The strong contributions of `rr` (rank two) and `rr_sum_7d` (rank 14) confirm that the model successfully learned both the weather-threshold and installation-vulnerability signals embedded in the sampling strategy, although not clearly visible in the EDA. The prominence of `nominal_voltage_15kV` (rank five) further suggests that grid topology and infrastructure tier are predictive of disturbance susceptibility beyond weather variables alone.

Analysis of Missed Disturbances

While analyzing the missed disturbances by the model, a key step to improve model performance, a pattern becomes visible. All missed disturbances are 15kV installations. As visible in the analysis those disturbances are in line with moderate but persistent rain rather than intense rainfall, the model is tuned toward high-rainfall events. Further, all missed disturbances have a high isolation score, indicating remote location that the model is not yet capable of capturing.

4.1.6 Model Results and Interpretation

For an operational inference run, the model can serve as a reliable first-pass screening tool, flagging high-risk substations under incoming extreme weather with a false alarm rate of only 11.3%. The cumulative lift of ≈2.0 at the operating threshold means that prioritizing the top tercile of installations ranked by predicted disturbance probability already captures 70.9% of all actual disturbances, enabling targeted awareness for inspection or potential resource allocation. However, the predicted probabilities with given residual calibration error and the missed disturbance of isolated 15kV substations exposed to prolonged moderate rainfall represent a systematic blind spot.

4.2 B) Disturbance Impact Model

Unlike DPM, which frames disturbance risk as a binary classification problem, the Disturbance Impact Model (DIM) addresses the complementary question which substation, once affected by a disturbance event, carries the highest societal and economic consequences. Therefore, the following sections describe the data preprocessing steps (4.2.1), and the scoring framework of substations (4.2.2), and the model results (4.2.3).

4.2.1 Data Preprocessing

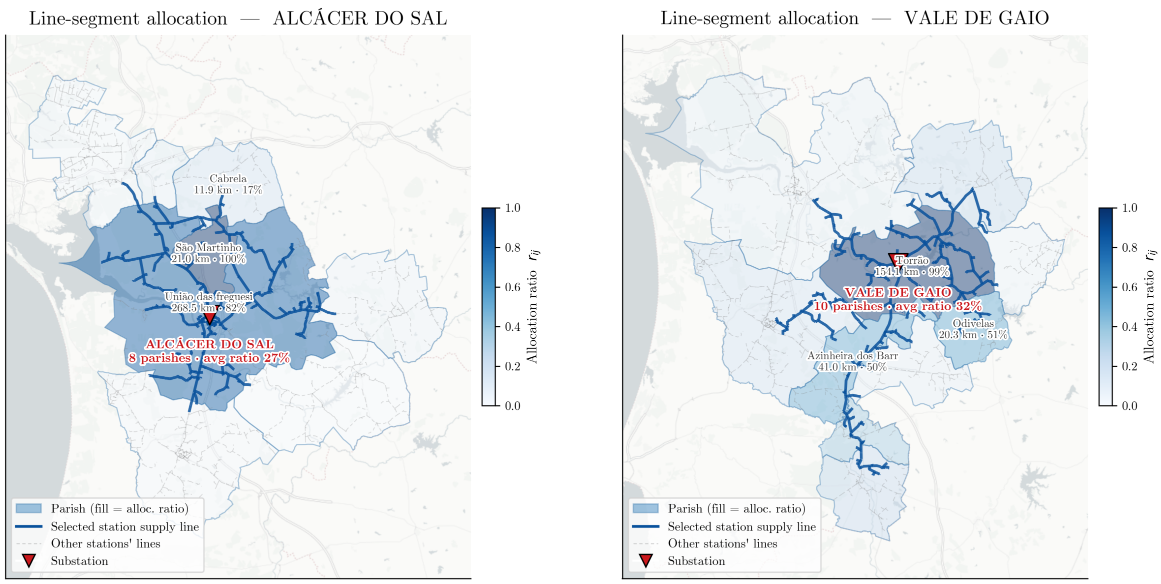
Supply-area geometry preparation Each substation is associated with one or more multi-string geometries representing supply lines, as described in section 3.1. These line features were grouped by station name and merged into a single multi-line string per station, yielding all supply-area objects[6].

[6] All layers were re-projected to the Portuguese national coordinate system (EPSG:3763) to ensure metric accuracy in subsequent length calculations.

Line-length proportional allocation Whereas socio-economic data is available at the parish level, substation impact is line-defined. Given the unavailability of street-based population and enterprise data, an allocation per substation must be done, to derive those importances. Thus, to bridge these two geometries, a concept was developed within this thesis, in which each supply-line network was intersected with the (2,882) parish polygons. For every station–parish pair the clipped segment length l_{ij} was recorded, and an allocation ratio

$$r_{ij} = \frac{l_{ij}}{\sum_k l_{kj}} \tag{4.1}$$

was computed, where the denominator sums the lengths of all supply lines crossing parish j. This ratio captures the proportional share of a parish's socio-economic exposure attributable to station i, and satisfies $\sum_i r_{ij} = 1.0$ for every parish by construction.



(a) Allocation of ALCÁCER DO SAL

(b) Allocation of VALE DE GAIO

Figure 4.5: Substation and parish attribution examples

Figure 4.5 illustrates the allocation mechanism for two neighbouring substations. ALCÁCER DO SAL supplies eight parishes, with an allocation ratio of 1.0 in São Martinho, where it is the sole supplier across 21 km of line. So, with a supply-line length of 21 km, the entire parish is supplied by ALCÁCER DO SAL, as no other lines cross the parish. VALE DE GAIO serves ten parishes, dominating in Torrão (ratio 0.99). Where both stations supply the same parish, such as Odivelas, where VALE DE GAIO holds a 51% share, exposure is split proportionally by clipped segment length. This example illustrates why a length-weighted ratio is necessary: a simple nearest-centroid assignment would allocate each parish entirely to one station, ignoring the shared service areas that characterize real distribution networks.

Fallback allocation A small number of parishes may not be crossed by any recorded supply line, either due to geometric gaps or data limitations. For these cases, a nearest-centroid fallback was applied[7]: each uncovered parish was assigned in full ($r_{ij} = 1.0$) to its geographically closest substation. After this step, full coverage was achieved across all 2,882 parishes, with zero population leakage.

Missing data handling To handle missing data, similar to DPM, no imputation was applied to substation metadata. Parish records missing census counts were filled with zero to avoid bias introduction, consistent with the sparse nature of the administrative boundary layer. Given incorrect coordinates of some stations, not all stations were successfully allocated to parishes.

4.2.2 Scoring Framework

To receive a prioritization of substations according to their supplied consumers and their importance to society, a scoring framework was introduced. The framework consists of three pillars: (I) Population, (II) Economy, and (III) Critical Infrastructure.

Pillar I — Population The resident population N_j for each parish was allocated to substations proportionally via Equation (4.1):

$$\text{pop}_i = \sum_j r_{ij} \cdot N_j \tag{4.2}$$

Across the substations with at least one intersection record, the allocated population ranges from 675 to 110,577 residents (median 21,465). The raw count is log-transformed, $\text{pop}_i^* = \ln(1+\text{pop}_i)$, to compress the pronounced right skew before normalization.

Pillar II — Economy Rather than treating all sectors equally, a three-tier criticality weighting scheme was applied prior to allocation. Sectors with direct dependency on uninterrupted

[7] using `gpd.sjoin_nearest`

electricity supply, like human health and social work, energy utilities, and water and waste management, received the highest weight of 3. Sectors with moderate dependency, e.g., manufacturing, construction, transport and storage, and information and communication, received weight 2. All remaining sectors received weight 1. The sector-weighted enterprise score per parish was then allocated using the same r_{ij} ratios and log-transformed in the same manner as population.

Pillar III — Critical infrastructure POI data was assigned tier weights reflecting their dependency on power continuity:

- **Tier 1 facilities – 10:** Hospitals, clinics, fire and police stations, water and wastewater plants, power plants
- **Tier 2 social infrastructure – 5:** Schools, pharmacies, town halls, government offices, drinking water facilities
- **Tier 3 transport nodes – 3:** Railway stations, bus stations, airports
- **Tier 4 transport nodes – 1:** Fuel stations

POIs were spatially joined to their containing parish via a *within* predicate, summed into a parish-level tier-weighted score, and allocated to substations via Equation (4.1).

Normalization All three log-transformed pillar scores were independently normalized to [0, 1] using min-max scaling, to preserve relative differences within each pillar while allowing comparison across pillars:

$$s_i^{(\cdot)} = \frac{x_i^* - \min(x^*)}{\max(x^*) - \min(x^*)} \tag{4.3}$$

Criticality Index and pillar weights The final CI is a weighted sum of the three normalized scores:

$$\mathrm{CI}_i = w_{\mathrm{pop}} \cdot s_i^{\mathrm{pop}} + w_{\mathrm{ent}} \cdot s_i^{\mathrm{ent}} + w_{\mathrm{poi}} \cdot s_i^{\mathrm{poi}} \tag{4.4}$$

with $w_{\mathrm{pop}} = 0.4$, $w_{\mathrm{ent}} = 0.2$, and $w_{\mathrm{poi}} = 0.4$. Population and critical infrastructure exposure are given equal and higher weight because they capture immediate life-safety consequences of a disturbance or outage, a substation serving many residents or a dense cluster of hospitals presents a directly human cost when disturbed. Economic activity, while significant, is spatially more diffuse and less uniquely attributable to a single upstream substation; it therefore receives a reduced weight of 0.2. All weights sum to 1, ensuring $\mathrm{CI}_i \in [0, 1]$. An overview of pillars, tiers and weights can be seen in Table 4.10.

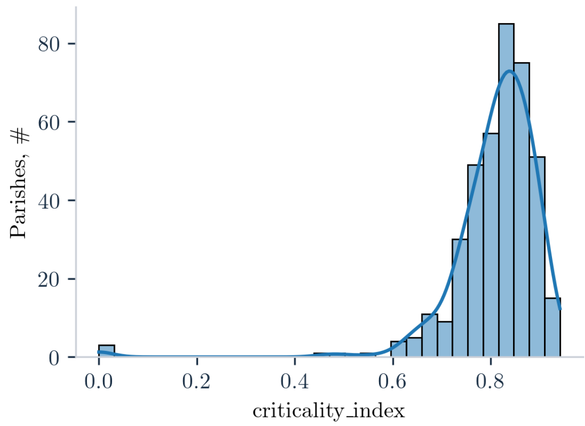
Pillar	**Component**	**Tier / Sector Weight**	**Pillar Weight**
(I) Population	Resident population (log-transformed)	—	0.4
(II) Economy	High dependency (health, energy, water & waste)	3	0.2
	Moderate dependency (manufacturing, transport, ICT)	2	
	Remaining sectors	1	
(III) Critical Infrastructure	Tier 1: Hospitals, emergency services, utilities	10	0.4
	Tier 2: Schools, pharmacies, government offices	5	
	Tier 3: Railway stations, bus stations, airports	3	
	Tier 4: Fuel stations	1	
Total			**1.0**

Table 4.10: Component weights used in the CI scoring framework

4.2.3 Model Results and Interpretation

Coverage and integrity The allocation pipeline achieves full parish coverage: all 2,882 parishes are assigned to at least one substation, with zero residual population leakage. The allocation ratio integrity check confirms that $|\sum_i r_{ij} - 1.0| < 0.01$ for every parish, validating conservation of population and enterprise counts across the spatial allocation step.

CI distribution The CI index is bounded and strongly left-skewed, visible in Figure 4.6a: the majority of substations cluster in the higher-mid range, while a small set of installations in and around the Lisbon and Porto metropolitan areas achieves scores above 0.90, reflecting their combination of large served populations and dense critical infrastructure. Table 4.6b lists the five highest-ranked substations.



(a) CI Distribution of parishes

Rank	Station name	Pop.	Ent.	POI	CI
1	VAROSA	0.919	0.887	0.992	0.942
2	TELHEIRA	0.903	0.865	1.000	0.934
3	ALHANDRA	0.946	0.889	0.935	0.930
4	SÃO SEBASTIÃO	0.948	0.901	0.908	0.923
5	CENTRAL TEJO	0.912	0.908	0.939	0.922

(b) Top 5 substations by CI

Figure 4.6: Results CI

The top-ranked substations are predominantly located in Greater Lisbon (TELHEIRA, ALHANDRA, CENTRAL TEJO), consistent with the population and enterprise concentration detailed in Section 3.3. VAROSA ranks high despite smaller served populations due to elevated POI scores, reflecting dense life-safety infrastructure relative to its service area. The CI is subsequently used in GS to contextualise predictions of DPM: substations simultaneously flagged as high-risk and carrying a high CI are surfaced as priority alerts.

4.3 A+B) Grid Sentinel PT

The following section describes the concept and architecture of GS, the final tool that serves as inference run (4.3.1), its data pipeline and calibration steps (4.3.2), and lastly the automation and deployment (4.3.3).

The live-deployment can be found here: Live Deployment of Grid Sentinel PT [8]

4.3.1 Concept, Architecture and Scoring

GS is designed to be an open-source, semi-live risk platform that operationalizes the two models developed in this thesis into a single deployable system. Its purpose is to provide grid operators, the public, and companies with a daily-updated, substation-level disturbance risk forecast for the Portuguese mainland up to three days ahead, contextualized by the societal criticality of each installation.

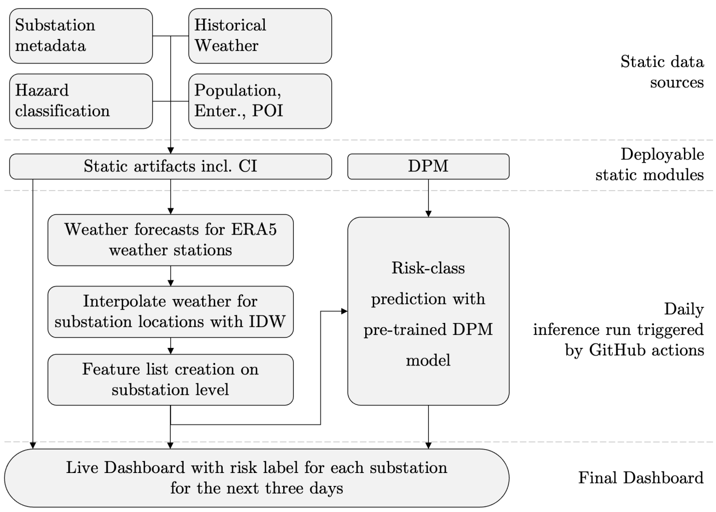


Figure 4.7: Grid Sentinel PT system architecture

GS integrates two deployable static modules and a daily inference pipeline. Offline, all static data sources, substation metadata, historical weather, hazard classifications, and socio-economic inputs, are processed once to produce the pre-trained DPM and the CI artifact, both of which are persisted and loaded at inference time without recomputation.

[8] Link: https://timbaaacodes.github.io/grid-sentinel/

The daily inference run, triggered automatically via GitHub Actions, executes three sequential steps on the dynamic side. Weather forecasts are first retrieved from the Open-Meteo API for the ERA5 grid points. The forecast values are then interpolated to each substation location using the same IDW scheme applied during training. Afterwards the 26-feature vector is recreated and scored by the pre-trained DPM to produce a per-substation disturbance probability and a `risk_class` label for each of the three forecast horizons.

Both streams, the dynamically generated risk predictions and the static CI scores, are passed independently to the live dashboard, where they are displayed as separate signals alongside each substation marker.

4.3.2 Data Pipeline and Forecast Calibration

A key methodological requirement is consistency between live forecast input and the E-OBS observations on which DPM was trained. Since Open-Meteo and E-OBS differ in ensemble construction and variable definitions, a two-step cross-calibration is applied at inference time. First, forecast values for each variable are retrieved at the five nearest grid points and aggregated via the same IDW scheme used during training (Equation 3.1). Second, the interpolated values are normalized using station- and season-specific means and standard deviations estimated from the 2023–2024[9] E-OBS baseline:

$$z_{v,s}^{\text{fcst}} = \frac{v^{\text{fcst}} - \hat{\mu}_{v,s}^{\text{E-OBS}}}{\hat{\sigma}_{v,s}^{\text{E-OBS}}} \tag{4.5}$$

This cross-calibration approximates distributional alignment by anchoring forecast values to the climatological scale of the training data, and is best understood as a bias correction rather than a direct threshold transfer. Wind speed (`fg`) and precipitation (`rr`) are additionally clipped to their empirical training maxima to guard against out-of-distribution values, and binary threshold flags are recalculated after calibration to preserve their physical meaning. The resulting feature vectors are structurally identical to those used during training.

Dynamic Risk Binning Risk class labels are assigned by comparing each station's predicted probability against a set of thresholds persisted alongside the trained model in `risk_config.json`. The base thresholds: Low ($\hat{p} < 0.15$), Medium ($0.15 \leq \hat{p} < 0.30$), Elevated ($0.30 \leq \hat{p} < 0.70$), and High ($\hat{p} \geq 0.70$) were derived from ROC analysis on the held-out test set. However, because the model was trained predominantly on extreme-weather event days, its baseline probability

[9] The window was selected as it represents the most recent climatology available in the E-OBS v31.0e release and minimizes distributional drift

output on routine forecast days can reach 0.15–0.40 even under calm meteorological conditions, driven by residual signal from rolling features such as `rr_sum_7d`. Applying static thresholds to these calm-day distributions leads to systematic over-classification: in early deployment trials, several substations were labelled Medium or Elevated on days with zero precipitation and light winds.

To address this, a dynamic calm-weather adjustment is applied at inference time. After scoring all stations, the pipeline computes the 90th percentile (p_{90}) of the network-wide probability distribution and compares it against the lowest base threshold. Two regimes are possible:

1. **Calm-day adjustment:** If $p_{90} < 0.15$, the day is classified as meteorologically calm across the network and all thresholds are scaled by a factor $\alpha = 1.5$, with each adjusted threshold capped at 0.99. This yields adjusted boundaries of [0.225, 0.45, 0.99]: only stations with near-certain predictions remain in the High tier, and the grid-wide noise floor is suppressed

2. **No adjustment:** When $p_{90} \geq 0.15$, i.e., a non-trivial share of the network already shows elevated probability,the base thresholds apply unmodified, preserving sensitivity during genuine events

This mechanism ensures that risk labels reflect relative severity given the prevailing weather regime, rather than absolute probability values that are sensitive to the distributional shift between training (event-day) and deployment (all-day) conditions.

4.3.3 Automation, Deployment and Visualization

The inference pipeline is automated via a GitHub Actions workflow executing daily at 07:00 UTC, performing four sequential steps: (i) retrieving three-day-ahead forecasts from the Open-Meteo API for each weather grid point, (ii) applying the calibration and feature construction pipeline to produce the 26 model inputs listed in Section 4.1.2, (iii) generating disturbance probabilities and dynamically-binned risk-class labels, and (iv) exporting the resulting table into per-forecast-day JSON files and commit them alongside the static station metadata to the public repository, from which a second workflow job builds the frontend and redeploys it to GitHub Pages. Next to the live-mode, a storm-mode showcases the model's output during Storm Kristin in January 2026. A preview of the deployment in the storm mode is shown in figure 4.8.

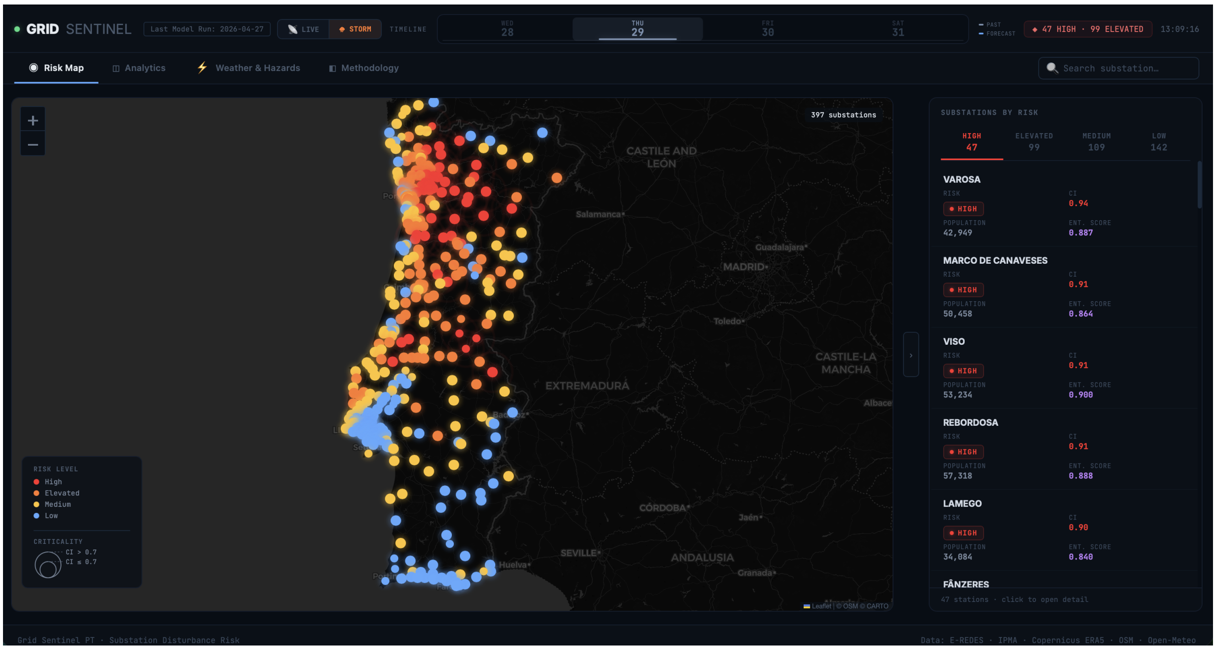


Figure 4.8: Grid Sentinel PT Preview

5 Discussion and Blueprint

This chapter discusses the methodological approach, its results and answers to the stated RQs (5.1). Further, current limitations and potential advancements for further features and usage is given (5.2).

5.1 Discussion

Extreme weather events pose a growing threat to electrical distribution infrastructure, yet substation-level, data-driven risk tools remain scarce, particularly at the high-to-medium voltage interface. This thesis addressed that gap for the Portuguese mainland by developing a modular, open-data framework that combines disturbance prediction with socio-economic impact assessment, operationalized in a publicly accessible dashboard.

The DPM demonstrates that weather-driven substation failures are predictable to a practically useful degree: the calibrated Gradient Boosting classifier achieves 81.3% accuracy and a ROC-AUC of 0.866 on a held-out set of 61 unseen installations, nearly doubling the random baseline in average precision. Daily and cumulative precipitation, temperature extremes, and geographic latitude emerge as the dominant predictors, consistent with the Atlantic-driven climate of the Iberian Peninsula. The three-type sampling strategy proved essential to decompose the classification problem into two orthogonal learning signals, weather thresholds and installation vulnerability, preventing the model from exploiting installation identity as a shortcut, while maximizing the low event number. Isolated 15 kV substations emerge as disproportionately vulnerable, accounting for the majority of recorded disturbances, yet also representing the primary blind spot of the model, which underperforms for installations disturbed by moderate but persistent rainfall rather than acute precipitation peaks.

The CI complements the prediction model by ranking every substation according to the socio-economic consequences of a disturbance. Through a self-developed line-length-proportional allocation mechanism, parish-level population, enterprise activity, and critical infrastructure

exposure are attributed to individual installations, yielding a bounded $[0, 1]$ index in which metropolitan substations in Greater Lisbon and Porto reach scores above 0.90.
Together, the two models are operationalized in GS, an open-source dashboard delivering automated, daily-updated, three-day-ahead disturbance probability forecasts for every primary substation, providing grid operators, companies, and the public with an actionable, real-time decision-support tool and a transferable blueprint for other European distribution operators.

5.2 Blueprint for Future Work

Methodological approach

Training splits are constructed at installation level without voltage-class stratification, resulting in the over-representation of 30 kV and 10 kV substations. Future work should stratify across voltage classes, ideally via cross-validation stratified simultaneously by installation, voltage class, and year to provide more reliable out-of-sample estimates. The E-OBS reanalysis product lacks coverage for coastal areas, and a more advanced IDW interpolation scheme would improve weather input quality. The CI's reliance on OSM data introduces varying regional completeness, whereas validation against official sources could improve CI robustness.

Additionally, the CI draws on the 2021 Portuguese Census, the most recent officially available source, yet parish-level population counts may have shifted meaningfully in the intervening years, particularly in fast-growing metropolitan parishes around Lisbon and Porto. Updating the population pillar as newer census releases become available would improve the accuracy of impact rankings for the highest-CI substations. Similarly, the static hazard classifications sourced from the 2021 World Bank ThinkHazard! dataset may understate current risk levels, particularly for wildfire exposure, which is expected to intensify under climate change.

The entire dual-signal sampling strategy rests on just 29 storm event dates across the ten-year study period. While the three-type sampling approach extracts substantial learning signal from this pool, the small number of distinct events imposes a structural ceiling on the model's ability to learn diverse disturbance patterns, particularly for rarer compound events such as simultaneous heat waves and drought that drive wildfire-related disturbances. Model retraining should be revisited as the disturbance record grows, particularly following major storm events in 2026.
A further source of label noise arises from grouping full outages, voltage deviations, and flicker events under a single binary label, meaning a minor flicker is treated identically to a complete

supply interruption during training and evaluation.

Disturbance Prediction Model

Despite regularization, a residual train-validation gap of 0.1686 in average precision persists, indicating incomplete generalization to unseen installations. This is compounded by the test subset covering only 79.3% of event dates compared to 96.6% in training and validation, meaning certain storm types are absent from evaluation entirely, which may present an optimistic picture of out-of-sample performance.

The spatial autocorrelation identified in feature importance, `lat` ranking substantially higher by Gini than by permutation importance, suggests the model partially exploits geographic clustering in the training partition rather than learning fully transferable physical relationships. Spatially-blocked cross-validation, holding out contiguous geographic regions rather than random installations, could produce more robust estimates of geographic transferability.

Further, the model exhibits a blind spot for isolated 15 kV substations that get disturbed under moderate but persistent rainfall, a disturbance mode structurally distinct from the high-rainfall signal the model was optimized to detect. Three mitigations merit exploration: (a) a dedicated sub-model for isolated 15 kV installations, (b) a voltage-stratified threshold rule elevating them to medium risk irrespective of predicted probability, or (c) oversampling their disturbance events during training. Priority covariate additions include real-time lightning data (currently excluded due to a lack of open-source inference-ready data), vegetation indices, and line-age or maintenance records if made available by the operator, which represent the highest-leverage additions to the feature set.

Disturbance Impact Model

The line-length-proportional allocation heuristic is tractable and reproducible but ignores actual network topology, a load-flow simulation could introduce more accurate substation importance rankings based on real supply information.

Grid Sentinel PT

The calm-day dynamic threshold adjustment is essentially a heuristic calibrated on early deployment observations rather than a held-out validation set. Its robustness across unseen meteorological regimes, including anomalous seasons or novel storm types such as those observed in early

2026, remains unvalidated, and the fixed scaling factor $\alpha = 1.5$ carries no statistical grounding beyond empirical trial. A key improvement would replace the current `calm_factor` heuristic with a percentile-based approach comparing each day's network-wide probability distribution against its historical climatological baseline. Further, the difference in weather distributions between the historical and forecast weather data can limit useability, which could be avoided by using the same source for both time horizons. GS is scoped to the Portuguese mainland but is architecturally transferable: Open-Meteo provides global forecast coverage, E-OBS can be replaced for most European countries, and the CI pipeline requires only parish-level population and POI data available through Eurostat and OSM. Replication for other operators requires retraining on operator-specific disturbance records but can reuse the full feature-engineering and calibration infrastructure.

Bibliography

Almeida, B., Soares, T., Teixeira Couto, M., Santos, T., Preto, I., Monteiro, C., Faria, G., Santos, R. J., and Ferreira Pinto, J. (2019). OUTAGE FORECAST–a real application of machine learning on grid operation management strategies. In 25th International Conference on Electricity Distribution (CIRED), pages Paper n ° 1209, Madrid, Spain. CIRED.

Arora, P. and Ceferino, L. (2023). Probabilistic and machine learning methods for uncertainty quantification in power outage prediction due to extreme events. Natural Hazards and Earth System Sciences, 23(5):1665–1683.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer, New York.

Breiman, L. (2001). Random forests. Machine Learning, 45(1):5–32.

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 785–794. ACM.

CIGRE (2020). The electrical power system portugal.

Cornes, R., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. (2018). An ensemble version of the e-obs temperature and precipitation datasets. Journal of Geophysical Research: Atmospheres, 123:9391–9409.

Direção-Geral do Território (DGT) (2021). CAOP 2021 – official administrative cartography of portugal: Continental administrative areas (polygons), PT-TM06/ETRS89. Shapefile, available at `https://www.dgterritorio.gov.pt/sites/default/files/ficheiros-cartografia/Cont_AAD_CAOP2021.zip`. Accessed: 19 March 2026.

E-REDES (2023). New E-REDES substation in Arganil starts operating. `https://www.e-redes.pt/en/news/2023/01/05/new-e-redes-substation-arganil-starts-operating`. Accessed: 2026-02-26.

E-REDES (2026a). Energy quality: Continuous phenomena [qualidade de energia: Fenómenos contínuos]. `https://e-redes.opendatasoft.com/explore/dataset/qualidade_energia_fenomenoscontinuos-final/information/`. Accessed: February 8, 2026.

E-REDES (2026b). Rnd reception capacity [capacidade de receção rnd]. `https://e-redes.opendatasoft.com/explore/dataset/capacidade-rececao-rnd/information/`. Accessed: February 8, 2026.

E-REDES – Distribuição de Eletricidade, S.A. (2025). Economic activity code (CAE) by district, municipality and parish. `https://e-redes.opendatasoft.com/explore/dataset/codigo-de-atividade-economica-por-distrito-concelho-e-freguesia/`. Data source: Instituto Nacional de Estatística (INE). License: CC BY 4.0. Dataset identifier: `codigo-de-atividade-economica-por-distrito-concelho-e-freguesia`. Last modified: 3 June 2025. Accessed: 19 March 2026.

Emmendorfer, L. R. and Dimuro, G. P. (2020). A novel formulation for inverse distance weighting from weighted linear regression. In Krzhizhanovskaya, V. V., Závodszky, G., Lees, M. H., Dongarra, J. J., Sloot, P. M. A., Brissos, S., and Teixeira, J., editors, Computational Science – ICCS 2020, pages 576–589, Cham. Springer International Publishing.

Euronews (2026). Six dead as storm kristin wreaks havoc in portugal with 202 km/h winds. `https://www.euronews.com/2026/01/29/six-dead-as-storm-kristin-wreaks-havoc-in-portugal-with-202-kmh-winds`. Accessed: 2026-04-01.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8):861–874.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5):1189–1232.

Ghomlaghi, A., Nasseri, M., and Bayat, B. (2022). How to enhance the inverse distance weighting method to detect the precipitation pattern in a large-scale watershed. Hydrological Sciences Journal, 67(13):2014–2028.

Google (2026). Google maps: Elevation data for portugal substations. `https://maps.google.com`. Accessed: February 8, 2026.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, 2nd edition.

Heuck, K., Dettmann, K.-D., and Schulz, D. (2007). Elektrische Energieversorgung: Erzeugung, Übertragung und Verteilung elektrischer Energie für Studium und Praxis.

IEEE (2014). IEEE guide for collecting, categorizing, and utilizing information related to electric power distribution interruption events. IEEE Std 1782-2014, pages 1–98.

Instituto Português do Mar e da Atmosfera (2026a). Climate normals — introduction and classification. `https://www.ipma.pt/en/oclima/normais.clima/`. Accessed: 2026-04-16.

Instituto Português do Mar e da Atmosfera (2026b). Climatological extremes portugal. `https://www.ipma.pt/en/oclima/extremos.clima/`. Accessed: 2026-04-16.

Instituto Português do Mar e da Atmosfera (2026c). Vento extremo na tempestade Kristin – Nota técnica. `https://www.ipma.pt/pt/media/noticias/news.detail.jsp?y=2026&f=Tempestade_kristin_portugal_continental.html`. Accessed: 2026-04-16.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems, volume 30, pages 3146–3154.

Louro, M. and Ferreira, L. (2022). Estimation of underground mv network failure types by applying machine learning methods to indirect observations. Energies, 15(17).

Martins, J., Rijo, T., Mar, A., César, A., Fortes, P., Calvão, T., Raiyani, K., Pereira, P., and Pires, V. F. (2024). Framework for risk assessment of the electrical power grid under extreme weather conditions. In 2024 IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), pages 1–6.

Minnegalieva, C. and Gainullina, A. (2023). Forecasting the probability of a power outage using binary classification. In 2023 17th International Conference on Electronics Computer and Computation (ICECCO), pages 1–4.

Oliveira, A. V. M., Raimundo, A. M., Gaspar, A. R., and Quintela, D. A. (2025). Heat waves in portugal during the 2001–2024 period: An overview. Climate, 13(6).

OpenStreetMap contributors (2017). Critical Public Infrastructure from https://planet.osm.org . `https://www.openstreetmap.org`.

Redondo Gonçalves, A. C. (2024). Variability of extreme weather events and influence on the renewable energy sector in the Iberian Peninsula: wind energy. PhD thesis.

Santos, A. (2007). Development and use of geocoding: Portugal approach. Expert Group Meeting on Contemporary Practices in Census Mapping and Use of Geographical Information Systems ESA/STAT/AC.115/8, United Nations, Department of Economic and Social Affairs, Statistics Division, New York. Prepared by Instituto Nacional de Estatística, Portugal; Accessed 19.04.2026.

Santos, M., Fragoso, M., and Santos, J. A. (2017). Regionalization and susceptibility assessment to daily precipitation extremes in mainland portugal. Applied Geography, 86:128–138.

Sharma, V., Hong, T., Cecchi, V., Hofmann, A., and Lee, J. Y. (2023). Forecasting weather-related power outages using weighted logistic regression. IET Smart Grid, 6(5):470–479.

Soares, P. M. M., Cardoso, R. M., Lima, D. C. A., and Miranda, P. M. A. (2017). Future precipitation in Portugal: high-resolution projections using WRF model and EURO-CORDEX multi-model ensembles. Climate Dynamics, 49(7–8):2503–2530.

Soares, T., Gonçalves, R., Louro, M., Pinto, J. F., Santos, T., Preto, I., and Monteiro, C. (2021). Forecasting the number of outage affected clients in extreme weather conditions. In CIRED 2021 - The 26th International Conference and Exhibition on Electricity Distribution, volume 2021, pages 1630–1633.

Statistics Portugal – Instituto Nacional de Estatística (2022). Census 2021: XVI General Population Census, VI General Housing Census – Final Results. INE – Instituto Nacional de Estatística, Lisbon, Portugal. Accessed: 19 March 2026.

Tan, J., Xie, X., Zuo, J., Xing, X., Liu, B., Xia, Q., and Zhang, Y. (2021). Coupling random forest and inverse distance weighting to generate climate surfaces of precipitation and temperature with multiple-covariates. Journal of Hydrology, 598:126270.

Tervo, R., Láng, I., Jung, A., and Mäkelä, A. (2021). Predicting power outages caused by extratropical storms. Natural Hazards and Earth System Sciences, 21(2):607–627.

The Guardian (2026). Portugal faces climate emergency after being battered by storms and extreme weather. `https://www.theguardian.com/world/2026/feb/12/portugal-climate-emergency-battered-storms-extreme-weather`. Accessed: 2026-04-16.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. Economic Geography, 46(sup1):234–240.

UNDP (2000). World Energy Assessment: Energy and the challenge of sustainability. United Nations Development Programme.

van Rijsbergen, C. J. (1979). Information Retrieval. Butterworths, London, 2nd edition.

World Bank (2021). Thinkhazard! natural hazard assessment tool. ID: WB_THINK_HAZARD. Licensed under GNU GPL v3.

Yang, F., Koukoula, M., Emmanouil, S., Cerrai, D., and Anagnostou, E. N. (2023). Assessing the power grid vulnerability to extreme weather events based on long-term atmospheric reanalysis. Stochastic Environmental Research and Risk Assessment, 37(11):4291–4306.

Appendix

List of Tables in Appendix

Table A.1: Portuguese Grid Overview; Source:(CIGRE, 2020)

Voltage Level	**Type**	**Operator**	**Grid Length**	**Overhead Lines**	**Comment**
400 kV	Transmission	TSO	2,712 km	–	Interconnection with Spanish grid
220 kV	Transmission	TSO	3,746 km	>50%	Direct connection with bulk industry
150 kV	Transmission	TSO	2,544 km	–	Direct connection with bulk industry
60 kV	Distribution-HV	DSO	9,543 km	>50%	Direct connection of industrial consumers
6–30 kV	Distribution-MV	DSO	73,547 km	>80%	Direct connection of industrial consumers and distributed energy producers
0.23–0.4 kV	Distribution-LV	DSO	143,440 km	~65%	Residential and SME consumers

Table A.2: Portuguese Grid Overview; Source:(CIGRE, 2020)

Table A.3: Weather related Disturbance events in the Portuguese power grid; Source: (E-REDES, 2026a)

Event	**Year**	**Month**	**Affected Substations**	**Cluster**
Windstorm	2015	10	16	Storm
Exceptional event of 10 January 2016	2016	1	24	Storm
Exceptional event of 14 February 2016	2016	2	64	Storm
Storm Dóris	2017	2	37	Storm
June wildfires	2017	6	15	Wildfire
October wildfires	2017	10	35	Wildfire
October wildfires	2017	10	23	Wildfire
Storm Ana	2017	12	133	Storm
Storm Emma	2018	2	57	Storm
Algarve tornado	2018	3	3	Other
Storm Gisele	2018	3	66	Storm
Storm Leslie	2018	10	105	Cyclone
Storm Helena	2019	2	30	Storm
Storm Elsa+Fabien	2019	12	102	Storm
Depression Glória	2020	1	28	Storm
Depression of 1–2 March	2020	3	55	Storm
Depression Hortense	2021	1	77	Storm
Atmospheric river	2021	10	73	Other
Depression Efrain	2022	12	145	Storm
Southern region storm	2022	12	85	Storm
Depression Aline	2023	10	16	Storm
Depression Ciarán	2023	11	62	Storm
Depression Domingos	2023	11	69	Storm
Cyclone Kirk	2024	10	120	Cyclone

Note: Cluster assignment groups events by meteorological type. Storm Leslie (2018) and Cyclone Kirk (2024) are classified as cyclones due to their tropical origin. Counts refer to affected primary substations and include events up to 2024 only.

Table A.4: Column inventory for training GeoDataFrame (datatype, spread/levels, source, provenance).

Column	Datatype	Data distribution / levels	Data source	Provenance
date	datetime	min=2015-01-20 00:00:00 max=2024-12-24 00:00:00	E-REDES outage data	origin
fg	float64	min=0.2831 max=7.1543	weather observations	origin
qq	float64	min=8.0507 max=340.8333	weather observations	origin
hu	float64	min=28.2733 max=94.8643	weather observations	origin
rr	float64	min=0 max=95.4757	weather observations	origin
tg	float64	min=1.4762 max=33.57	weather observations	origin
tn	float64	min=-1.9772 max=24.7695	weather observations	origin
tx	float64	min=4.8874 max=42.641	weather observations	origin
tg_extreme_indicator	float64	min=-3.0583 max=3.0307	Copernicus Dataset	Feature engineered
tg_is_extreme	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset	Feature engineered
tn_extreme_indicator	float64	min=-3.1035 max=3.0077	Copernicus Dataset	Feature engineered
tn_is_extreme	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset	Feature engineered
tx_extreme_indicator	float64	min=-2.8535 max=2.8798	Copernicus Dataset	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
tx_is_extreme	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset	Feature engineered
rr_extreme_ indicator	float64	min=-0.5652 max=10.0553	Copernicus Dataset	Feature engineered
rr_is_extreme	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset	Feature engineered
fg_extreme_ indicator	float64	min=-3.2742 max=7.3322	Copernicus Dataset	Feature engineered
fg_is_extreme	binary/ categorical	n=2 (0, 1)	Copernicus Dataset	Feature engineered
hu_extreme_ indicator	float64	min=-3.8243 max=1.9099	Copernicus Dataset	Feature engineered
hu_is_extreme	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset	Feature engineered
qq_extreme_ indicator	float64	min=-3.5451 max=3.7618	Copernicus Dataset	Feature engineered
qq_is_extreme	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset	Feature engineered
month	int32	n=12 ex: 1, 2	E-REDES outage data	engineered
month_name	object	n=12 ex: October, January	E-REDES outage data	engineered
iso_week	int32	min=1 max=53	E-REDES outage data	engineered
wind_severity	float64	min=-3.2742 max=7.3322	Copernicus Dataset	Feature engineered
rain_severity	float64	min=0 max=10.0553	Copernicus Dataset	Feature engineered
temp_severity	float64	min=0 max=3.0583	Copernicus Dataset	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
humidity_severity	float64	min=0 max=3.8243	Copernicus Dataset	Feature engineered
weather_stress_index	float64	min=2.353e-03 max=5.8703	Copernicus Dataset	Feature engineered
temp_range_severity	float64	min=4.845e-05 max=3.3004	Copernicus Dataset	Feature engineered
extreme_hazard_count	int32	n=5 ex: 0, 1	ThinkHazard, Copernicus Dataset	Feature engineered
is_multi_hazard	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered
is_severe_multi_hazard	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered
is_freezing	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engineered
is_severe_cold	binary/ categorical	n=1 (0)	E-REDES outage data	Feature engineered
is_very_hot	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engineered
is_extreme_heat	binary/ categorical	n=1 (0)	E-REDES outage data	Feature engineered
is_heavy_rain	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engineered
is_extreme_rain	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engineered
is_torrential_rain	binary/ categorical	n=1 (0)	E-REDES outage data	Feature engineered
is_strong_wind	binary/ categorical	n=1 (0)	E-REDES outage data	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
is_storm_wind	binary/ categorical	n=1 (0)	E-REDES outage data	Feature engineered
is_gale_force	binary/ categorical	n=1 (0)	E-REDES outage data	Feature engineered
is_severe_gale	binary/ categorical	n=1 (0)	E-REDES outage data	Feature engineered
is_very_dry	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engineered
is_very_humid	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engineered
temp_range	float64	min=0.3006 max=23.0238	Copernicus Dataset	Feature engineered
temp_std	float64	min=0.1761 max=11.5254	Copernicus Dataset	Feature engineered
temp_normalized	float64	min=0.2511 max=1.0282	Copernicus Dataset	Feature engineered
wind_energy	float64	min=0.0801 max=51.1843	Copernicus Dataset	Feature engineered
wind_power	float64	min=0.0227 max=366.1886	Copernicus Dataset	Feature engineered
rain_intensity	float64	min=0 max=95.4757	Copernicus Dataset	Feature engineered
ice_wind_hazard	float64	min=0 max=4.9771	Copernicus Dataset	Feature engineered
severe_ice_wind	binary/ categorical	n=1 (0)	Copernicus Dataset	Feature engineered
storm_index	float64	min=0 max=0.4734	Copernicus Dataset	Feature engineered
severe_storm	binary/ categorical	n=1 (0)	Copernicus Dataset	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
heat_humidity_ stress	float64	min=0 max=376.0207	Copernicus Dataset	Feature engi-neered
wind_chill_effect	float64	min=0 max=31.532	Copernicus Dataset	Feature engi-neered
rain_with_sun	binary/ categorical	n=2 (0, 1)	Copernicus Dataset	Feature engi-neered
wind_rain_extreme	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset	Feature engi-neered
temp_wind_ extreme	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset	Feature engi-neered
is_winter	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engi-neered
is_spring	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engi-neered
is_summer	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engi-neered
is_autumn	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engi-neered
is_peak_demand	binary/ categorical	n=2 (0, 1)	E-REDES outage data	Feature engi-neered
month_sin	float64	min=-1 max=1	E-REDES outage data	Feature engi-neered
month_cos	float64	min=-1 max=1	E-REDES outage data	Feature engi-neered
week_sin	float64	min=-1 max=1	E-REDES outage data	Feature engi-neered
week_cos	float64	min=-1 max=1	E-REDES outage data	Feature engi-neered
weather_risk_score	float64	min=0.084 max=3.3246	Copernicus Dataset	Feature engi-neered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
risk_level	binary/ categorical	n=2 (0, 1)	Copernicus Dataset	Feature engineered
is_critical_ conditions	binary/ categorical	n=1 (0)	E-REDES outage data	Feature engineered
year	object	n=10 ex: 2015, 2016	E-REDES outage data	engineered
installation code	object	n=278 ex: 1015S5002300, 1107S5903700	E-REDES substation data	origin
type of installation	object	n=1 (Subestação)	E-REDES substation data	origin
designation	object	n=278 ex: Pombal, Camarate	E-REDES substation data	origin
busbar	object	n=30 ex: 30 kV - II, 10 kV	substation asset registry	origin
municipality	object	n=123 ex: Pombal, Loures	E-REDES outage data	origin
nuts iii	object	n=36 ex: Região de Leiria, Área Metropolitana de Lisboa	E-REDES outage data	origin
nominal voltage (v)	float64	n=3 (10000, 15000, 30000)	E-REDES outage data	origin
label	binary/ categorical	n=2 (0, 1)	outage/event labeling	origin
river flood_encoded	float64	n=5 ex: 0, 1	ThinkHazard	origin
urban flood_ encoded	float64	n=4 ex: 1, 2	ThinkHazard	origin
coastal flood_ encoded	float64	n=4 ex: 0, 1	ThinkHazard	origin

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
landslide_encoded	float64	n=5 ex: 0, 1	ThinkHazard	origin
wildfire_encoded	float64	n=2 (1, 4)	ThinkHazard	origin
extreme heat_ encoded	float64	n=4 ex: 0, 1	ThinkHazard	origin
earthquake_ encoded	float64	n=3 (1, 2, 3)	ThinkHazard	origin
tsunami_encoded	float64	n=3 (0, 2, 3)	ThinkHazard	origin
water scarcity_ encoded	float64	n=3 (1, 2, 3)	ThinkHazard	origin
elevation_m	float64	min=3.2929 max=996.6923	Copernicus Dataset,Google Places API	Feature engi- neered
total_hazard_risk	float64	n=18 ex: 11, 12	ThinkHazard, Copernicus Dataset	Feature engi- neered
avg_hazard_risk	float64	min=1.2222 max=3.2222	ThinkHazard, Copernicus Dataset	Feature engi- neered
max_hazard_risk	float64	n=2 (2, 4)	ThinkHazard, Copernicus Dataset	Feature engi- neered
high_risk_hazards	int32	n=6 ex: 0, 1	ThinkHazard, Copernicus Dataset	Feature engi- neered
is_multi_hazard_ site	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engi- neered
elevation_category	float64	n=3 (0, 1, 2)	Copernicus Dataset,Google Places API	Feature engi- neered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
is_lowland	binary/ categorical	n=2 (0, 1)	Copernicus Dataset,Google Places API	Feature engineered
is_mid_elevation	binary/ categorical	n=2 (0, 1)	Copernicus Dataset,Google Places API	Feature engineered
is_highland	binary/ categorical	n=2 (0, 1)	Copernicus Dataset,Google Places API	Feature engineered
is_mountain	binary/ categorical	n=1 (0)	Copernicus Dataset,Google Places API	Feature engineered
elevation_ normalized	float64	min=3.293e-03 max=0.9967	Copernicus Dataset,Google Places API	Feature engineered
flood_exposure	float64	n=12 ex: 1, 2	ThinkHazard, Copernicus Dataset	Feature engineered
flood_prone	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered
rain_flood_risk	float64	min=0 max=969.662	ThinkHazard, Copernicus Dataset	Feature engineered
extreme_rain_flood	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered
heavy_rain_flood	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
lowland_river_flood	float64	min=0 max=35.2604	ThinkHazard, Copernicus Dataset	Feature engineered
urban_flood_active	float64	min=0 max=13.2227	ThinkHazard, Copernicus Dataset	Feature engineered
wind_coastal_risk	float64	min=0 max=5.5521	ThinkHazard, Copernicus Dataset	Feature engineered
storm_coastal	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered
rain_landslide_risk	int32	min=0 max=318	ThinkHazard	origin
highland_rain_hazard	int32	n=4 ex: 0, 2	ThinkHazard, Copernicus Dataset	Feature engineered
heat_fire_risk	float64	min=1.2044 max=17.0564	ThinkHazard, Copernicus Dataset	Feature engineered
fire_weather	float64	n=2 (0, 4)	ThinkHazard, Copernicus Dataset	Feature engineered
extreme_fire_weather	binary/ categorical	n=1 (0)	ThinkHazard, Copernicus Dataset	Feature engineered
wind_fire_risk	float64	min=0 max=0.9802	ThinkHazard, Copernicus Dataset	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
heat_exposure_stress	float64	min=0 max=12.7923	ThinkHazard, Copernicus Dataset	Feature engineered
heat_water_stress	float64	min=0.3566 max=10.071	ThinkHazard, Copernicus Dataset	Feature engineered
extreme_heat_active	binary/ categorical	n=1 (0.0)	ThinkHazard, Copernicus Dataset	Feature engineered
elevation_temp_gradient	float64	min=0.0333 max=24.1269	Copernicus Dataset,Google Places API	Feature engineered
highland_cold	binary/ categorical	n=1 (0)	Copernicus Dataset,Google Places API	Feature engineered
highland_wind	float64	min=0 max=1.3188	Copernicus Dataset,Google Places API	Feature engineered
lowland_humidity	float64	min=0 max=0.9478	Copernicus Dataset,Google Places API	Feature engineered
lowland_flood	float64	min=0 max=8.8151	Copernicus Dataset,Google Places API	Feature engineered
compound_weather_hazard	float64	n=5 ex: 0, 1	ThinkHazard, Copernicus Dataset	Feature engineered
is_compound_critical	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
stress_hazard_product	float64	min=4.967e-03 max=16.9585	ThinkHazard, Copernicus Dataset	Feature engineered
high_stress_high_risk	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered
winter_flood_risk	float64	min=0 max=8.8151	ThinkHazard, Copernicus Dataset	Feature engineered
summer_fire_risk	float64	min=0 max=5.6855	ThinkHazard, Copernicus Dataset	Feature engineered
summer_drought_stress	binary/ categorical	n=1 (0.0)	ThinkHazard, Copernicus Dataset	Feature engineered
hazard_activation_score	float64	min=0 max=10	ThinkHazard, Copernicus Dataset	Feature engineered
is_hazard_activated	binary/ categorical	n=2 (0, 1)	ThinkHazard, Copernicus Dataset	Feature engineered
distance_to_coast	float64	min=0.966 max=185.6912	Copernicus Dataset,Google Places API	Feature engineered
distance_to_center	float64	min=2.073 max=285.5474	Copernicus Dataset,Google Places API	Feature engineered
distance_to_urban_center	float64	min=0.6716 max=172.1831	Copernicus Dataset,Google Places API	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
distance_to_ nearest_substation	float64	min=0.686 max=39.8085	Copernicus Dataset,Google Places API	Feature engineered
substation_density_ 5km	float64	n=16 ex: 0, 1	Copernicus Dataset,Google Places API	Feature engineered
is_coastal_region	binary/ categorical	n=2 (0.0, 1.0)	Copernicus Dataset,Google Places API	Feature engineered
rr_sum_3d	float64	min=0 max=175.1315	Copernicus Dataset	Feature engineered
rr_sum_7d	float64	min=0 max=294.1228	Copernicus Dataset	Feature engineered
rr_sum_14d	float64	min=0 max=452.0152	Copernicus Dataset	Feature engineered
fg_max_3d	float64	min=0.5017 max=7.1543	Copernicus Dataset	Feature engineered
fg_max_7d	float64	min=0.741 max=7.1543	Copernicus Dataset	Feature engineered
fg_max_14d	float64	min=0.8998 max=7.2843	Copernicus Dataset	Feature engineered
tx_max_7d	float64	min=9.5784 max=43.2405	Copernicus Dataset	Feature engineered
tn_min_7d	float64	min=-5.1455 max=17.9577	Copernicus Dataset	Feature engineered
temp_range_max_ 7d	float64	min=4.7341 max=26.0712	Copernicus Dataset	Feature engineered
consecutive_rain_ days	float64	n=15 ex: 0, 1	Copernicus Dataset	Feature engineered
consecutive_ heavy_rain_days	float64	n=11 ex: 0, 1	Copernicus Dataset	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
consecutive_extreme_days	float64	min=0 max=35	Copernicus Dataset	Feature engineered
consecutive_freeze_days	float64	n=14 ex: 0, 1	Copernicus Dataset	Feature engineered
consecutive_wind_extreme_days	float64	n=14 ex: 0, 1	Copernicus Dataset	Feature engineered
weather_stress_mean_7d	float64	min=0.1334 max=2.3699	Copernicus Dataset	Feature engineered
weather_stress_max_7d	float64	min=0.2464 max=5.8703	Copernicus Dataset	Feature engineered
weather_risk_mean_7d	float64	min=0.1095 max=1.7859	Copernicus Dataset	Feature engineered
hazard_activation_mean_7d	float64	min=0 max=7.2598	Copernicus Dataset	Feature engineered
hazard_activation_max_7d	float64	min=0 max=10	Copernicus Dataset	Feature engineered
multi_hazard_days_7d	float64	n=8 ex: 0, 1	Copernicus Dataset	Feature engineered
severe_multi_hazard_days_7d	float64	n=5 ex: 0, 1	Copernicus Dataset	Feature engineered
rain_flood_risk_sum_7d	float64	min=0 max=2.793e+03	Copernicus Dataset	Feature engineered
flood_days_7d	float64	n=8 ex: 0, 1	Copernicus Dataset	Feature engineered
temp_std_7d	float64	min=0.0949 max=5.0974	Copernicus Dataset	Feature engineered
wind_std_7d	float64	min=0.0342 max=2.2151	Copernicus Dataset	Feature engineered
weather_stress_std_7d	float64	min=0.0571 max=1.9171	Copernicus Dataset	Feature engineered

Continued on next page

Column	Datatype	Data distribution / levels	Data source	Provenance
rr_trend_7d	float64	min=-6.3242 max=12.593	Copernicus Dataset	Feature engineered
fg_trend_7d	float64	min=-0.4257 max=0.9209	Copernicus Dataset	Feature engineered
temp_trend_7d	float64	min=-1.3333 max=1.6136	Copernicus Dataset	Feature engineered
wind_rain_product_7d	float64	min=0 max=1.209e+03	Merged training dataset	Feature engineered
storm_days_7d	binary/ categorical	n=1 (0.0)	Copernicus Dataset	Feature engineered
is_event_day	binary/ categorical	n=1 (1)	E-REDES outage data	engineered
event_number	int32	min=1 max=67	E-REDES outage data	engineered
geometry	geometry	geom_types=Point	E-REDES substation data	origin
lon	float64	min=-9.443 max=-6.5706	E-REDES substation data	engineered
lat	float64	min=37.0383 max=42.0258	E-REDES substation data	engineered

Table A.5: Final GB hyperparameter search space

Hyperparameter	**Search Distribution**
n_estimators	randint(10, 300)
max_depth	[2, 3, 4, 5, 6, 7, 8, 9]
learning_rate	uniform(0.001, 0.1)
min_samples_split	randint(10, 60)
min_samples_leaf	randint(10, 50)
max_features	['sqrt', 'log2']
subsample	uniform(0.6, 0.3) → [0.6, 0.9]
ccp_alpha	uniform(0.0, 0.001)
min_impurity_decrease	uniform(0.0, 0.005)

Table A.6: Evolution of Gradient Boosting hyperparameter search spaces across optimization stages

Variant	**Search Distribution**
Baseline Refined	
	n_estimators – randint(100, 300)
	max_depth – [2, 3, 4, 5]
	learning_rate – uniform(0.01, 0.1)
	min_samples_split – randint(30, 60)
	min_samples_leaf – randint(20, 50)
	max_features – ['sqrt', 'log2']
	subsample – uniform(0.6, 0.3) → [0.6, 0.9]
	ccp_alpha – uniform(0.0, 0.001)
	min_impurity_decrease – uniform(0.0, 0.005)
Extended Baseline	
	n_estimators – randint(10, 300)
	max_depth – [2, 3, 4, 5, 6, 7, 8, 9]

Continued on next page

Continued from previous page

Variant	Search Distribution
	`learning_rate` – `uniform(0.001, 0.1)`
	`min_samples_split` – `randint(10, 60)`
	`min_samples_leaf` – `randint(10, 50)`
	`max_features` – `['sqrt', 'log2']`
	`subsample` – `uniform(0.6, 0.3)` → [0.6, 0.9]
	`ccp_alpha` – `uniform(0.0, 0.001)`
	`min_impurity_decrease` – `uniform(0.0, 0.005)`
Regularized + Early Stopping	
	`n_estimators` – `randint(200, 800)`
	`n_iter_no_change` – `[5, 10, 20]`
	`validation_fraction` – `[0.1, 0.15]`
	`max_depth` – `[2, 3, 4, 5, 6]`
	`learning_rate` – `loguniform(0.005, 0.2)`
	`min_samples_split` – `randint(30, 120)`
	`min_samples_leaf` – `randint(20, 80)`
	`max_features` – `['sqrt', 'log2', 0.5, 0.7]`
	`subsample` – `uniform(0.5, 0.45)` → [0.5, 0.95]
	`max_leaf_nodes` – `[None, 16, 32, 64]`
	`ccp_alpha` – `loguniform(1e-5, 0.01)`
	`min_impurity_decrease` – `uniform(0.0, 0.01)`
	`loss` – `['log_loss']`

Continued on next page

Continued from previous page

Variant	Search Distribution
Final Tuned	
	n_estimators – [600, 800, 1000]
	max_depth – [3, 4]
	max_leaf_nodes – [None, 32, 64]
	learning_rate – loguniform(0.05, 0.20)
	min_samples_split – randint(40, 100)
	min_samples_leaf – randint(25, 80)
	max_features – [0.5, 0.7, 'sqrt']
	subsample – uniform(0.55, 0.40) → [0.55, 0.95]
	min_impurity_decrease – uniform(0.002, 0.008) → [0.002, 0.010]
	ccp_alpha – loguniform(1e-5, 1e-3)
	loss – ['log_loss']

Table A.7: Hyperparameter search spaces for all classifiers.

Model	Hyperparameter	Search space
Random Forest	n_estimators	Integer [200, 800)
	max_depth	2, 3, 4, 5, 6, 7, 8, None
	min_samples_split	Integer [10, 80)
	min_samples_leaf	Integer [5, 60)
	max_features	sqrt, log2, 0.3, 0.5
	max_leaf_nodes	20, 30, 50, 80, 100, None
	class_weight	balanced, balanced_subsample, None
	bootstrap	True
	max_samples	0.5, 0.6, 0.7, 0.8, 0.9, None
	min_impurity_decrease	Uniform [0.0, 0.01]
	ccp_alpha	Uniform [0.0, 0.005]
XGBoost	n_estimators	Integer [100, 300)
	max_depth	2, 3, 4, 5
	learning_rate	Uniform [0.01, 0.10]
	min_child_weight	Integer [10, 50)
	subsample	Uniform [0.6, 0.9]
	colsample_bytree	Uniform [0.5, 0.9]
	reg_alpha	Uniform [0.0, 2.0]
	reg_lambda	Uniform [1.0, 5.0]
	gamma	Uniform [0.0, 1.0]
	scale_pos_weight	1.0, 2.0, 5.0
LightGBM	n_estimators	Integer [100, 300)
	max_depth	5, 6, 7, −1
	learning_rate	Uniform [0.01, 0.10]
	num_leaves	Integer [10, 50)
	min_child_samples	Integer [20, 60)

Continued on next page

Model	Hyperparameter	Search space
	`subsample`	Uniform [0.6, 0.9]
	`colsample_bytree`	Uniform [0.5, 0.9]
	`reg_alpha`	Uniform [0.0, 2.0]
	`reg_lambda`	Uniform [1.0, 5.0]
	`min_split_gain`	Uniform [0.0, 0.5]
	`class_weight`	`balanced`, `None`
Gradient Boosting	`n_estimators`	Integer [100, 300)
	`max_depth`	2, 3, 4, 5
	`learning_rate`	Uniform [0.01, 0.10]
	`min_samples_split`	Integer [30, 60)
	`min_samples_leaf`	Integer [20, 50)
	`max_features`	`sqrt`, `log2`
	`subsample`	Uniform [0.6, 0.9]
	`ccp_alpha`	Uniform [0.0, 0.001]
	`min_impurity_decrease`	Uniform [0.0, 0.005]