POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

The Impact of Weather on Power Grid Failures in Piedmont

TESI DI LAUREA MAGISTRALE IN
AUTOMATION AND CONTROL ENGINEERING
INGEGNERIA DELL'AUTOMAZIONE

Author: Andoni Juantorena Polo

Student ID: 10989707
Advisor: Prof. Loredana Cristaldi
Co-advisors: Luca Martiri, Andrea Moschetti
Academic Year: 2024-25

Abstract

Weather is a major driver of power-grid disruptions, yet quantifying its role is difficult because outages are recorded at specific grid assets while weather is observed at stations with uneven coverage, especially in complex terrain. As a result, the nearest station is often not representative of the conditions experienced at the affected site.

Building on an existing national relational database, this work updates and extends a regional case-study dataset and develops a reproducible workflow that consolidates outage records, grid-site information, daily meteorological observations, and lightning detections into a single structure for analysis. Daily site-level weather exposure is reconstructed through spatial interpolation strategies designed to be transparent and robust under missing data and heterogeneous station availability.

On top of these reconstructed series, the work builds higher-level indicators that better reflect how weather impacts infrastructure in practice. Beyond single-day extremes, the feature layer captures multi-day persistence and accumulation, abrupt transitions, and compound situations in which multiple drivers co-occur (e.g., wet conditions together with strong winds, or precipitation interacting with snow processes). These indicators support interpretable analyses at multiple scales, from seasonal patterns to daily and location-specific comparisons.

Finally, the evidence is synthesised into an interpretable daily probability-based risk score, designed to help prioritise attention to periods and locations of highest risk. The score is evaluated with forward-in-time validation and probability calibration. Overall, the analyses suggest that disruption risk is more consistently linked to persistent and compound conditions than to isolated extremes, and the framework provides a scalable basis for future extensions with additional environmental or infrastructure information.

Key-words: Power Grid; Weather-related Outages; Spatial Interpolation; Fault Prediction.

Abstract in italiano

Il meteo è un importante fattore di disservizi della rete elettrica, ma quantificarne il ruolo è difficile perché i guasti sono registrati su specifici asset, mentre le osservazioni meteo provengono da stazioni con copertura irregolare, in particolare in aree a orografia complessa. Di conseguenza, la stazione più vicina può non rappresentare le condizioni del sito interessato.

Partendo da un database nazionale già esistente, questo lavoro aggiorna un caso di studio regionale e sviluppa un flusso di lavoro riproducibile che integra dati di guasto, informazioni sui siti, osservazioni meteorologiche giornaliere e rilevamenti di fulmini in una struttura unica per l'analisi. L'esposizione meteo giornaliera per sito è ricostruita con interpolazione spaziale trasparente e robusta rispetto a dati mancanti e disponibilità eterogenea delle stazioni.

Sulle serie ricostruite, il lavoro deriva indicatori che descrivono meglio lo stress meteorologico sulle infrastrutture. Oltre agli estremi giornalieri, le feature catturano persistenza e accumulo su più giorni, transizioni brusche e condizioni composte con più driver (ad es. bagnato con vento forte, o precipitazioni che interagiscono con la neve). Gli indicatori supportano analisi interpretabili, dai pattern stagionali ai confronti giornalieri per localizzazione.

Infine, le evidenze sono sintetizzate in un punteggio di rischio giornaliero interpretabile e basato su probabilità, per prioritizzare periodi e località a rischio più elevato. Il punteggio è valutato con validazione forward-in-time e calibrazione probabilistica. Nel complesso, le analisi suggeriscono un legame più stabile con condizioni persistenti e composte rispetto a estremi isolati e che il framework è scalabile per future estensioni con ulteriori informazioni ambientali o infrastrutturali.

Parole chiave: Rete elettrica; Guasti Meteo-correlati; Interpolazione Spaziale; Predizione dei Guasti.

Contents

List of Figures

List of Tables

1 Introduction

1.1. Context: Weather and Power Grid Reliability

Weather is one of the main drivers of disruptions to electrical systems and is therefore a key issue in resilience planning. This is particularly relevant in Italy, where service continuity performance is routinely monitored and discussed in relation to the impact of extreme weather events, including floods, windstorms, and heat waves, which can have a measurable influence on annual interruption indicators [1].

In this context, weather-related stressors affect the grid through several well-known mechanisms: mechanical load (wind, snow, and ice), insulation degradation and surface discharges (flashovers) in wet conditions, interactions between vegetation and overhead lines during storms, thermal stress during heat waves, and failures associated with atmospheric electrical discharges. It is important to note that these impacts are rarely "local" in the strict sense. Severe conditions tend to affect many components within the same spatial footprint and can result in correlated outages and longer restoration times, making weather reliability analysis directly relevant to the operation and management of user sites, and not just for post-event reporting [2], [3]. In line with this operational perspective, Terna, which is the transmission system operator that manages the Italian transmission grid, has explicitly framed resilience against extreme weather as an engineering priority, proposing asset-focused methodologies aimed at outage prevention and improved robustness under meteorological stress [4].

At the same time, the hazard environment is not stationary. In Italy, recent analyses of precipitation at high temporal resolution show that the structure of intense precipitation and short-duration extremes is a key element of risk and can evolve over time, which is consistent with a broader concern in Europe regarding changes in heavy rainfall and the likelihood of compound or concurrent extremes affecting critical infrastructures [5]. For grid operators, this implies that the conditions leading to weather-related outages cannot be treated as a fixed baseline. Risk can arise from brief, intense events, from sequences of several days that progressively increase vulnerability, or from combinations of stressors that test protection and restoration processes. This perspective also motivates the growing emphasis on "resilience planning" under extreme weather conditions. Recent work proposes risk-based methodologies for assessing and prioritising resilience measures in transmission and distribution grids under extreme events, explicitly reflecting the type of regulatory and operational framework found in Italy [6].

In regions with strong gradients and frequent regime transitions, such as the Italian Alpine arc and its transition to the Po Valley, the coupling between precipitation, temperature, and snow processes provides a clear example of how "meteorological exposure" can change abruptly over short distances. Rain-on-snow (ROS) events can accelerate melt and runoff, and their frequency and characteristics are sensitive to temperature-driven changes in precipitation phase [7], [8], [9]. More generally, persistence matters. "Wetness memory" indicators such as the Antecedent Precipitation (AP) index offer a physically motivated way to represent accumulated conditions that can increase vulnerability even when daily precipitation is only moderate [10], [11]. ROS and AP index are used here as illustrative cases within a broader set of accumulated and compound weather indicators considered in the analysis.

1.2. Problem Statement and Motivation

Despite the operational relevance of weather-driven disruptions, quantifying the relationship between weather and grid failures is not trivial. Grid failure records are associated with specific assets or user sites, while meteorological information is measured at weather stations with limited and uneven spatial density, especially in complex terrain. This creates a spatial mismatch, as the weather measured at the nearest station is not necessarily representative of the weather experienced at the failure location. This limitation is well documented in the literature on interpolation for complex terrain and for extremes, where precipitation and wind fields can vary sharply over short distances [12].

In Piedmont, this spatial mismatch is particularly relevant. The region has steep orographic gradients and frequent regime transitions between plains and mountainous areas, which can produce large differences in precipitation intensity, snowfall and temperature over relatively small spatial scales. Orographic enhancement is a known driver of these gradients in the Alpine region [13] and helps explain why site-level exposure can be underrepresented by station-level measurements if not carefully reconstructed.

A second challenge concerns the reliability of available labels. Terna provides a classification that identifies a subset of failures as weather-related, but the label does not specify the type of hazard or the physical mechanism. The baseline work of Moschetti et al. [14] explicitly highlights the limitation of Terna's data, providing an initial analysis. More generally, recent work on outage data points out that reporting formats and causal information are often deficient, making transparent data integration and careful validation necessary parts of any weather-outage study [15].

For these reasons, the first requirement of this work is to build a robust and reproducible pipeline that links each failure and each user site with daily meteorological exposure at the site level, and then enriches those exposures with interpretable features. Specifically, the pipeline produces a unified site-day analytical table, where each record represents one site on a given day, with reconstructed and aligned weather variables, derived indicators that capture intensity and persistence, and categorical bins designed for risk estimation and communication. This choice follows the same general direction as the database approach of the reference work [14], concentrating the workflow in Piedmont and on site-level exposures ready for analysis.

Crucially, before using monthly correlations or site-day risk estimates as evidence, the work introduces an explainability audit as an early validation step. The audit tests whether the failures that Terna labels as weather-related are consistent, in a substantial fraction of cases, with objectively reconstructed moderating conditions at the same sites and dates, including rainfall intensity, wetness and persistence indicators, snow and temperature regimes, wind, and lightning proximity where available. It also isolates the residual share of ambiguous cases, such as in 2024 where lightning data was unavailable, by separating indeterminate cases from genuinely unexplained ones. By acting as a certification layer for site-level attribution, the audit provides a more robust basis for subsequent analyses and for the construction of an operational Failure Probability Index (FPI).

1.3. Scope and Contributions

This work focuses on the Piedmont region (shown in Figure 1.1) and on failures classified by Terna as weather-related in the period 2008-2024. The main analytical unit is the site-day, i.e., each observation links a specific user site with the reconstructed meteorological exposure for that location on that date. The work is designed to be consistent with the data resolution available throughout the study period. Therefore, the methodological emphasis is placed on moderate, cumulative, and compound meteorological indicators, such as multi-day precipitation accumulations, persistence proxies such as the AP index, ROS conditions, and other interaction or transition flags. This choice provides features that remain meaningful and comparable even when sub-daily information is incomplete or unavailable, and aligns with the goal of producing interpretable evidence and operationally usable outputs.

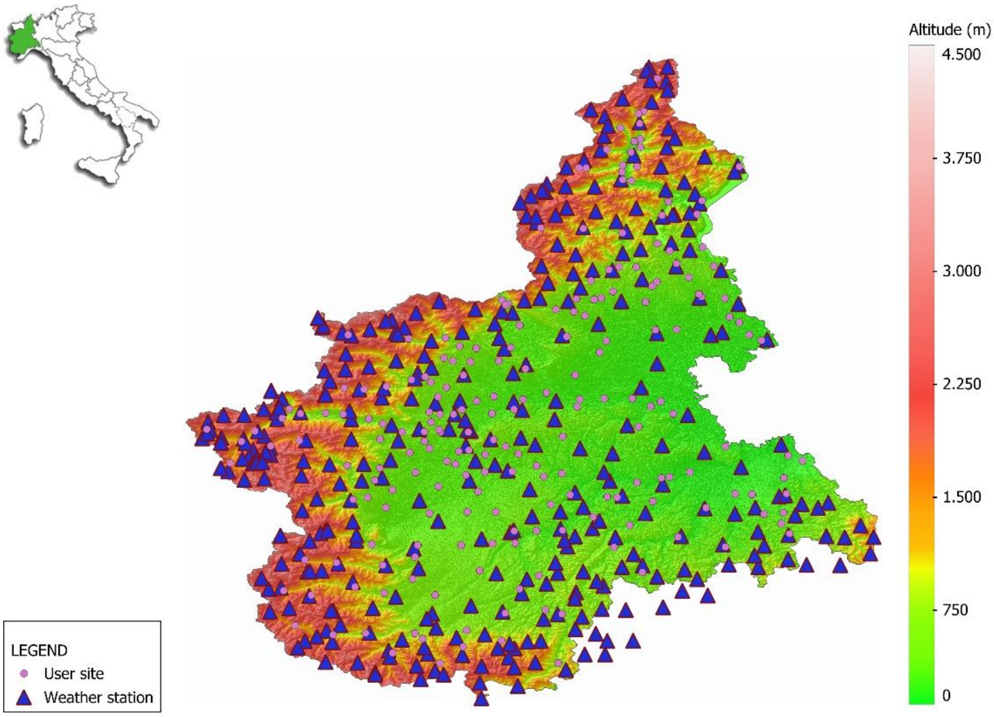


Figure 1.1 Topography of Piedmont and Location of Analysed Grid Sites and Weather Stations

Main contributions. In relation to the baseline work by Moschetti et al. [14], which provided the national database foundation and an initial framework for analysis, the contributions of this work are:

1. **Power grid / weather database and pipeline.** A reproducible data pipeline and relational database derived from Terna's national system, designed to support joins between user sites, failures, and meteorological sources. The workflow is developed and demonstrated in the Piedmont use case, but the same architecture can be extended to other Italian regions (and potentially nationwide) provided that site and meteorological data with comparable coverage and quality are available.
2. **Site level meteorology reconstruction**. A reconstruction layer that produces daily time series of meteorological exposure at each user site using a baseline inverse distance approach and variable-specific methods that better reflect the physics and characteristics of the data, including kriging for temperature and spline-based reconstruction for precipitation and snow, along with practical handling choices for wind.
3. **Explainability audit as validation**. An explainability audit introduced as an early validation step to quantify how often Terna-labelled meteorological failures align with objective site-level meteorological moderators defined by

transparent thresholds, and to separate ambiguous or undetermined cases, including those affected by the absence of lightning information in 2024.

4. **Integrated evidence across scales and space**. A compact set of results that combines a monthly macro view with deseasonalized association analysis, site-day risk estimates based on bins and interactions, and spatial structure tests to assess whether anomalous periods are geographically clustered or diffuse.
5. **Operational synthesis through the FPI.** A FPI that translates validated empirical signals into an interpretable site-day risk score formulated in odds space, with robustness controls such as shrinkage and hierarchical fallback, and with calibration evaluated using walk-forward backtesting.

1.4. Outline

The work is structured as follows:

- Chapter 1 sets the context and motivation for studying weather-driven disruptions in the power grid, defines the research scope and objectives, and summarizes the main contributions of the work.
- Chapter 2 provides the background needed to interpret the problem, combining an overview of how adverse weather can affect grid components with the methodological concepts used later (from handling environmental observations to analysing rare events).
- Chapter 3 describes the datasets used and how they are turned into a consistent analytical resource. It explains how outage information, grid-site records, meteorological measurements, and lightning observations are cleaned, aligned in space and time, and organized into a relational database suitable for downstream analyses.
- Chapter 4 presents the full analytical methodology, explaining how daily site-level weather conditions are reconstructed from station networks, how multi-day and multi-driver indicators are derived to better represent real exposure, and how the study design supports interpretable comparisons across locations and time.
- Chapter 5 reports and interprets the results obtained with this framework, moving from aggregated descriptive evidence to finer daily analyses, and showing how the proposed indicators and risk scoring behave when evaluated in a forward-in-time setting.
- Chapter 6 concludes the work by summarizing the main takeaways, discussing limitations and practical implications, and outlining how the framework can be extended in future work with additional environmental and infrastructure-related information.

2 Background and Related Work

2.1. Weather-related Outage Drivers and Physical Mechanisms

Power grid user sites are directly exposed to environmental forces, especially overhead lines, towers, insulators and outdoor substation equipment. At the component level, weather-related failures can be grouped into a small number of physical stress pathways, i.e., mechanical, electrical, or thermal, each associated with typical failure modes and weather variables that best represent exposure.

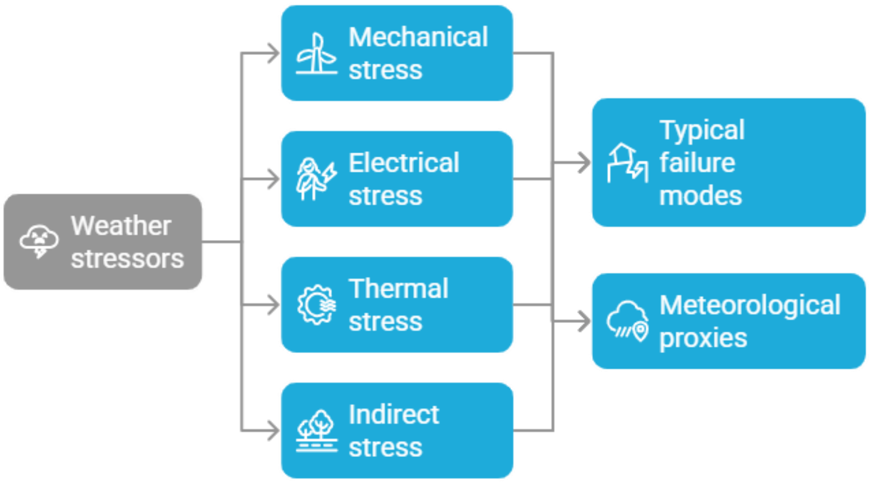


Figure 2.1 Weather Stress Pathways and Typical Outage Mechanisms in Power-Grid Assets

- **Mechanical stress** is primarily associated with **wind**, **snow**, and **ice loads** [16], [17]. Strong winds can cause conductor clashing, excessive stress on towers and damage from falling trees or branches, while snow and ice accumulation increase mechanical loads on conductors and support structures. These mechanisms tend to produce spatially clustered failures within the affected footprint, because the same storm system simultaneously exposes many spans and structures. In practice, this helps explain why outages during severe storms can involve multiple sites and strain restoration capacity.
- **Electrical stress** includes processes associated with **lightning** and **wet contamination**. Lightning can cause direct faults or protection trips [18], while persistent wet conditions can increase the probability of flashovers, especially

when combined with pollution, salt, or particle deposition on insulating surfaces [19], [20], [21]. In this category, it is not only the peak intensity that matters, the **duration and persistence** of wet conditions can modify surface conductivity and make the system more vulnerable even in the absence of an extreme daily maximum.

- **Thermal stress** is driven by sustained **high temperatures** and high loads, which increase conductor deflection and reduce thermal margins [22], [23]. Even when the triggering mechanism is not "damage" in the structural sense, temperature can degrade operating conditions and reduce robustness, particularly during prolonged heat waves. This pathway is relevant when failures are associated with persistent stress rather than to a single storm event.
- **Indirect mechanisms** play an important role in the overall impact. **Vegetation interaction** is often the main mediator between meteorology and line faults, and is influenced both by severe windstorms [24] and prolonged wet periods that increase soil saturation and tree instability [25]. Restoration results are also conditioned by accessibility restrictions, road conditions, and the simultaneity of failures in the territory, which can extend interruption durations during major events [26].

Across countries, empirical evidence supports two consistent patterns. First, severe weather events account for a disproportionate share of high-impact interruptions [3], including long-lasting ones, and days with major disruptions tend to reflect multiple stressors rather than a single driver. Second, various studies show that risk is best described by weather regimes and their temporal evolution rather than by isolated single-day extremes [27].

This motivates the use of cumulative and persistence indicators, as they represent preconditioning and multi-day sequences that can increase vulnerability prior to the triggering event.

2.2. Site-level Meteorology Reconstruction

To relate failures recorded at user sites to meteorological drivers, observations at stations must be translated into representative exposures at the coordinates of each site. This step is common in infrastructure studies because station networks are irregular and several variables exhibit intense spatial gradients, particularly in mountainous regions. There are numerous methods of spatial interpolation and prediction [28]; in this work, the reconstruction focuses on three families because they offered the best empirical performance and the most interpretable behaviour for our case study.

- **Deterministic interpolation**: among deterministic methods, **Inverse Distance Weighting (IDW)** is widely used in operational pipelines because it is simple, fast, and easy to interpret [29]. IDW estimates the value at a target location as a weighted average of nearby stations, with weights decreasing with distance. Its behaviour is controlled by practical choices such as distance power, number of neighbours and search radii. For this reason, IDW is often used as a transparent baseline and as a robust backup when the assumptions required by more complex methods are not sufficiently supported by the data.

- **Geostatistical prediction**: geostatistical methods from the **kriging** family provide a formal framework when spatial autocorrelation can be modelled [30]. Kriging treats the variable as a spatial process and predicts values at unobserved locations using a covariance structure fitted by a variogram. This allows spatial predictions based on observed spatial dependence and, in principle, provides a measure of prediction uncertainty through kriging variance. In environmental applications, **Regression Kriging (RK)** is often adopted when some of the variability can be explained by covariates such as elevation [31]. In this case, a deterministic trend is first fitted and then kriging is applied to the residual field. This separation is attractive in complex terrain, where large-scale gradients can be intense while local deviations still exhibit spatial structure.
- **Spline-based surfaces**: spline approaches, including thin-plate smoothing splines, are common in climatological interpolation because they can represent smooth spatial variation and incorporate topographic predictors flexibly [32]. In practice, splines are often effective when the spatial structure is not well captured by a stationary variogram or when the dominant signal is a smooth gradient driven by geography and elevation. This makes them a practical choice for precipitation and snow variables in mountainous contexts, where gradients can be steep and non-linearities are common. At the same time, spline-based reconstructions can smooth out very local extremes, so their performance is typically evaluated using cross-validation and diagnostics aligned with the intended use of the reconstructed fields.

In these families, a recurring theme in the literature is that no single method dominates for all variables. Temperature often benefits from approaches that combine altitude-driven trends with spatial modelling of the residual, while precipitation and snowfall typically require methods that handle strong intermittency and non-Gaussian behaviour [33]. Thus, many applied studies adopt a variable-specific strategy that balances data availability, physical plausibility and interpretability. The same principle motivates the reconstruction choices adopted in this work, presented in detail in Chapter 1.

2.3. Statistical and Spatial Tools

The literature employs a combination of descriptive, statistical and spatial tools to characterise the relationship between meteorology and failures at multiple scales.

- **Macro-scale association checks**. Correlation measures are used as a first diagnosis. **Pearson's correlation** captures linear association, while **Spearman's rank correlation** focuses on monotonic relationships and is less sensitive to outliers. Reporting both helps verify that conclusions about possible drivers are not artefacts of scale or a small number of extreme months [34].
- **Deseasonalization and anomaly extraction**. Given that outage series and weather variables exhibit strong seasonality, many studies remove the mean seasonal cycle before testing associations. A common approach is to work with monthly anomalies, calculated by subtracting the long-term mean for each month of the year. When a more flexible decomposition is required, **STL (Seasonal-Trend decomposition using Locally Estimated Scatter plot Smoothing (LOESS))** can separate seasonal and residual components through iterative smoothing [35].
- **Site-day effect sizes**. At the site-day resolution, effect sizes are more naturally expressed through risk measures rather than correlations. **Risk Ratios (RR)** and **comparisons based on odds-ratio** quantify how the probability of a failure changes under an exposure condition relative to a baseline condition. **Confidence Intervals (CI)** are essential for communicating sampling uncertainty, especially when stratifying by season, altitude bands, or compound bins where counts become small [36].
- **Spatial concentration**. Spatial structure matters because storms and multi-day wet regimes can produce geographically concentrated outages. Density-Based Spatial Clustering of Applications with Noise **(DBSCAN)** is used to detect clusters without fixing the number of clusters in advance, but the results depend on the neighbourhood radius and should be interpreted through sensitivity analysis [37]. In parallel, the **Nearest-Neighbour Index (NNI)** provides a compact summary by comparing the observed distances to the nearest neighbour with those expected under spatial randomness, offering complementary verification even when clusters are not clearly separated [38].

2.4. Related Work

A substantial body of research has investigated the relationship between weather and power interruptions with the aim of improving reliability assessment and operations. One influential line of thinking frames weather as a system-level risk that links physical exposure with restoration constraints and motivates resilience-oriented planning [39]. Panteli & Mancarella provide a widely cited synthesis of this perspective, emphasizing that impacts are often driven by combinations of hazards and operational conditions rather than by a single variable in isolation.

Many empirical studies operationalise this idea by constructing storm-focused outage models, where outages are predicted or explained using meteorological descriptors and exposure proxies at the service territory level. In distribution network contexts, Wanik et al. [40] present an example of storm outage modelling that combines event characterisation with infrastructure exposure, illustrating the type of end-to-end link between meteorological fields and outage observations required to quantify risk. Increasingly, these storm-event pipelines adopt machine learning approaches to improve predictive performance and handle nonlinearities between weather variables and contextual features [41].

A complementary direction, closely aligned with the "regime/transition" approach, is to focus on weather patterns rather than on isolated extremes. Souto et al. [42] explicitly identify weather patterns and transitions most likely to coincide with outages in the UK, supporting the idea that the probability of outage can be interpreted through discrete meteorological states and their evolution. This line of work is particularly relevant when the goal is interpretability through communicable categories as operational "conditions", rather than abstract continuous predictors.

Spatial structure is another recurring theme, especially for synoptic-scale events where the storm footprint induces geographically correlated outages. Studies focusing on extratropical storms show how prediction and attribution benefit from representing the event footprint and the associated spatial dependence of impacts, rather than treating each location as an independent trial [43].

Cold season studies further highlight that model performance improves when predictors better reflect physically meaningful loading conditions rather than relying on generic seasonal markers. Work linking outages to snow-related loads and forest-infrastructure interactions [44] (e.g., proxies for crown snow load risk) provides evidence that combining weather information with contextual susceptibility can refine outage risk signals, particularly in regions where snow processes and vegetation conditions are relevant.

Beyond these representative lines, the literature contains numerous statistical modelling approaches, from spatial Generalised Linear Mixed Models (GLMM) to Generalized Additive Models (GAM) and nonparametric tree-based methods, often evaluated in severe events such as hurricanes and ice storms [45], [46], [47]. These studies are relevant here primarily as methodological precedents. They show how prediction or association can be improved by allowing for nonlinear effects, incorporating spatial dependence, and validating models against event-based ground truth or against out-of-sample periods.

Overall, the related work establishes that rigorous weather-outage analyses depend on:

1. Reliable matching between outage observations and the corresponding meteorological conditions
2. Analytical choices that balance interpretability with the ability to capture multifactorial and spatially structured impacts

In this work, these insights motivate a site-day framework for Piedmont designed to support interpretable condition-based analyses and transparent validation of weather moderators prior to subsequent risk estimation.

3 Data Pipeline and Database Engineering

This chapter describes how heterogeneous inputs (outage records, meteorological observations, lightning detections, and geographic context) are integrated into a single dataset ready for analysis. The core output is a consistent site-day panel covering 2008-2024, where each row represents one Terna user site (identified by *codice_utente*) on one calendar day. This structure underpins all subsequent results because it enables coherent linkage between failures and local exposures, feature engineering based on persistence and extremes, and robust comparisons across sites, seasons, and provinces. The pipeline is implemented as a reproducible Extract, Transform, Load (ETL) workflow, where raw sources are preserved, intermediate harmonized layers are materialised, and a single final table is exposed for analysis.

3.1. Data Sources

3.1.1. Terna (Outage Events and User Sites Registry)

Terna annually publishes the *"Elenco delle Schede Registrazione Disalimentazioni (energia non fornita ed energia non ritirata) degli Utenti connessi alla RTN"*, reporting disconnections affecting high voltage user sites connected to the national transmission grid. The document is released within the framework of Terna's Grid Code (Allegato A.54) and the Italian Regulatory Authority for Energy, Networks and Environment (ARERA) (Deliberazione 55/2024/R/eel) [48].

Each record corresponds to a disconnection affecting a specific site and includes:

- **Two event identifiers:** *N° Disalimentazione* and *Codice Disalimentazione*. *N° Disalimentazione* can group multiple site-level disconnections belonging to the same underlying malfunction/incident, while *Codice Disalimentazione* uniquely identifies the individual interruption record.
- **Territorial area and timestamps**, plus technical descriptors (originating network element, owner/operator, voltage level, disconnection type, pre-event network configuration).
- **Cause classification (two-level):** *CODICE CAUSA AEEG - 1° LIVELLO* and *2° LIVELLO* (AEEG1-AEEG2). In Terna's interruption reports, each disconnection

is assigned a **two-level AEEG cause code** following the official classification in the Grid Code. The **1st-level code** is a macro-cause (e.g., 2FM, 4AC), and the **2nd-level code** specifies the corresponding sub-category. In particular, the codes used in this work to select candidate weather-related disconnections correspond to:

   - **(2FM, 30E / 30R / 30I) → Force Majeure (Forza Maggiore)**: Terna uses 2FM when the interruption is attributed to "extraordinary external events beyond normal operational control". Within this group, 30E refers to catastrophic natural disasters of very large magnitude (e.g., landslides, floods, earthquakes, tsunamis, volcanic eruptions), while 30R and 30I refer to exceptional meteorological events, explicitly defined as events that cause exceedance of the design limits of network elements (classified respectively when the origin is on the RTN vs on other non-RTN networks, according to the codification in force).
   - **(4AC, 300) → Other causes / transmission environmental**: Terna uses 4AC (Altre Cause) as a macro-category for interruptions whose origin is classified within the transmission domain, and the 2nd-level code 300 specifies "Trasmissione RTN - Ambientali", i.e., interruptions attributed to environmental causes affecting RTN elements.

  These cause-code combinations encode Terna's operational classification of interruptions, and they provide the starting point used in this work to extract the subset of disconnections that Terna itself flags as weather-related at reporting time.

- **Impact and site-related fields:** *codice_utente* (reported as *"CODICE UNIVOCO DEL SITO UTENTE AT DISALIMENTATO"*), connection descriptors, interrupted power, duration, ENS/ENR metrics, NF/NR/SD classification, and additional MT/BT or ENS-U/ENR-U fields when available.

In addition to outage events, Terna provides a yearly registry of user sites (*"Censimento Utenti"*). Each site is uniquely identified by the same *codice_utente* and described through stable administrative and operational attributes (e.g., area, region/province/locality, owner, denomination, load vs generation, connection code, voltage).

In the adopted workflow:

- The **disconnection report** provides the event + associated *codice_utente*
- The **site registry** provides stable descriptors needed to contextualize the event geographically and operationally

Terna does **not** directly provide final geographic coordinates for each event. Therefore, the workflow links events to locations by:

1. Joining disconnections to the site registry via *codice_utente*
2. Attaching **coordinates and elevation** through external geographic information

Disconnection reports are distributed as a paginated PDF table with a fixed grid layout. To make it machine-readable, a dedicated Python ETL pipeline was implemented (Camelot lattice extraction + pandas), designed specifically for grid-like PDF tables. The pipeline:

1. Extracts tables across all pages
2. Removes repeated headers and empty fragments
3. Normalises column names into SQL-safe identifiers
4. Parses Italian date/time formats (gg/mm/aaaa; hh.mm:ss)
5. Converts numeric quantities using Italian decimal separators
6. Converts duration fields expressed as mm,ss into seconds for consistent numerical handling

The processed data are loaded into SQL through a staging structure that preserves traceability, with a **raw table** storing the extracted content and **normalized staging tables** separating core event attributes from sparse accessory fields.

A final Quality Control (QC) step checks:

1. Identifier formats (e.g., patterns consistent with SRD for *Codice Disalimentazione*)
2. Validity of area codes
3. Duplicates
4. Semantic inconsistencies from extraction artifacts (e.g., negative interrupted power due to parsing noise; duration = 0 with non-zero energy; unrealistically large durations)

Such cases are explicitly flagged and, conservatively, treated as missing values in analytical tables to avoid bias.

3.1.2. ARPA Piedmont (Weather Stations and Daily Observations)

Local meteorological exposure is reconstructed from the observational network managed by Agenzia Regionale per la Protezione Ambientale (**ARPA)** Piedmont, using:

1. **Station metadata**
2. **Daily observations** for precipitation, temperature, wind, and snow

Metadata provide identifiers and descriptive attributes (code, name, province), together with **WGS84 coordinates (lat/lon)** and **elevation**. Observations are accessed via ARPA public web services and downloaded programmatically to ensure **repeatability and traceability**.

A practical aspect of ARPA's system is that daily meteorological observations are not always keyed by a single uniform station identifier:

- For precipitation/temperature/wind, daily records are retrieved through the meteorological daily endpoint and keyed by a **measurement point identifier** (*fk_id_punto_misura_meteo*) referenced in the metadata.
- In the extraction pipeline, this identifier is parsed and used to query time series. Station metadata (*lat/lon/alt*) are attached at download time so that raw extracts preserve location attributes.

Because ARPA endpoints are paginated, the extractor iterates pages until completion and implements basic robust measures (timeouts, limited retries, short delays) to reduce the risk of partial downloads and to avoid excessive load on the service. The output is a raw daily archive that can be regenerated deterministically for a given date window.

Snow variables are part of the same observational framework but are delivered through a dedicated **nivological endpoint**, with different linkage:

- Snow records reference *fk_id_punto_misura_nivo*
- This must be resolved to the corresponding station entity by following API links from the measurement point to the nivological station resource and retrieving *codice_stazione*

To minimize the number of requests, the implementation uses **caching** so repeated point→station lookups are performed only once. This is essential to keep snow within the same station-centric model adopted for other variables and to ensure consistent integration.

The ARPA network is heterogeneous, since stations differ in start/end dates, missingness patterns, and variable availability. Moreover, even when a station is operational, daily records may not contain a complete set of measurements. For this reason, raw extracts undergo conservative harmonization before database loading:

- Restructuring into standardized per-variable tables
- Dropping rows with no informative content for the target block (e.g., snow days where both snow-on-ground and fresh snow are missing)
- Removing duplicated station-day entries

Importantly, the goal at this stage is to store a faithful and clean representation of ARPA's reporting, preserving missingness explicitly.

Some explanatory variables used later to represent antecedent conditions are implemented as station-based proxies derived from precipitation persistence, rather than being directly observed. In particular, an AP index is used along with rolling rainfall accumulations to capture the memory effect of wetness in a consistent site-day framework.

3.1.3. Lightning Data

Lightning observations are included to represent convective electrical activity in the proximity of each user site, which is a relevant hazard for both direct (e.g., flashover, insulation breakdown, protection trips) and indirect mechanisms (e.g., concurrent severe convective conditions). Following the procedure described by Moschetti et al. in their technical documentation [14], the lightning component of this work is derived from a dataset originally provided by the Italian Armed Forces Air Force branch. In practical terms, the raw lightning detections are processed and re-organized into daily summaries referenced to each site, so they can be integrated consistently with the (site, day) analytical unit used throughout the work. **Lightning covariates are available up to 2023;** for 2024, lightning-derived features are flagged as unavailable and handled explicitly in the audit and downstream analyses.

3.2. Database Design Principles and Key Entities

This study is based on an **existing national database** covering the whole of Italy (user sites, faults, weather stations and their measurements). Starting from this baseline, the **data model is reorganized and specialised** to make the analysis more efficient, transparent, and reproducible in the Piedmont case study, while maintaining a scalable structure in case the approach is extended to the national level in the future.

The ultimate goal is to build a **single analytical table at site-day resolution**, where each record represents a user site on a specific day, and integrates in the same row:

i. Information on meteorological failures associated with the site

ii. The reconstructed meteorological exposure at the site coordinates

To reach this goal efficiently and transparently, the pipeline materialises intermediate layers that can feed alternative reconstruction branches, namely a baseline IDW table and an improved hybrid table that selects the best-performing method per variable family (Spline for precipitation/snow, RK for temperature, and IDW for wind). Best-performing refers to the cross-validation performance reported in Sections 4.1.2 and 4.1.3.

This 'one row per site and day' approach is key for two reasons:

- Many meteorological moderators that are used later (e.g., moving accumulations, persistence indicators, threshold counts, and composite event flags) require a **complete daily series** in order to consistently calculate multi-day variables.
- Concentrating the relevant variables in a single table makes subsequent analysis **computationally cheap**, since statistical models, binning, event filters, and visualisations are executed with simple queries on a single dataset, avoiding repetitive and costly joins.

However, generating this final site-day table directly from the raw national data would be computationally expensive and difficult to inspect, especially if the pipeline were to be extended beyond Piedmont. For this reason, the design follows a **layered approach**, creating **intermediate tables** that progressively add information (site selection, fault aggregation, daily meteorology, lightning variables...). The final table is then obtained as a controlled materialisation that "feeds" the analytical dataset from these layers. This structure improves performance, reduces redundant computations, and leaves clear validation points before running the analyses presented in the following sections.

Some database tables and field names are reported in their original Italian naming (e.g., *codice_utente, data*) to ensure direct traceability to the SQL implementation.

3.2.1. Baseline Database

This work is based on the national database designed and implemented by Moschetti et al. [14], which integrates three blocks of information: **power grid failures**, **weather data** and **lightning**. The baseline E-R schema describing the whole database (shown in Figure 3.1) uses *località* as the common geographic "hub" that enables consistent joins across the three domains.

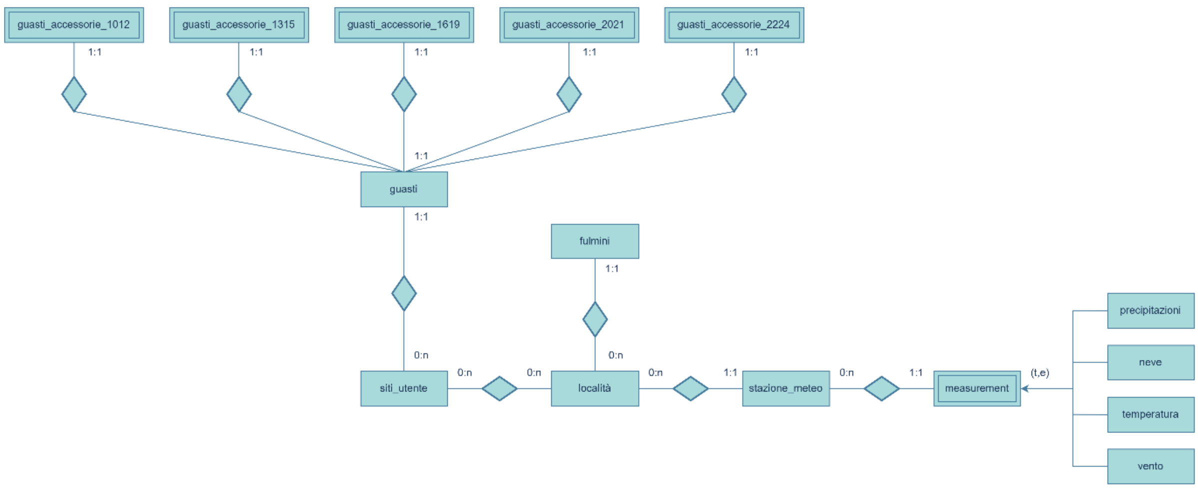


Figure 3.1 E-R Schema of the Baseline Database [14]

Starting from this database (covering Italy up to 2023), I extended the dataset to include Terna's failure records for 2024 and the corresponding ARPA Piedmont's meteorological observations for 2024. The goal of this update is to keep the case study as current as possible, and ensure that the subsequent pipeline and results are based on the most recent information available for Piedmont. Below, I summarize what each block contains and, for each entity, which columns are relevant.

A) Power grid failures

On the electrical side, the design is built around three key entities: **guasti** (failures), **siti_utente** (user sites), and **località** (geographic reference). To handle the fact that Terna's failure reporting format changes across years, Moschetti et al. keep the most stable ("primary") fields in *guasti*, and move year-dependent fields into accessory weak entities linked 1:1 to *guasti*. The baseline integrates annual Terna user site publications using a temporal validity approach (*anno_inizio*, *anno_fine*) so that changes in site attributes over time can be preserved.

Logical design:

- **guasti:** <u>id</u>, n_disalimentazione, data, ora, elemento_rete_origine, titolare_rete_origine, tensione_rete_origine, codice_causa_AEEG_1_livello, codice_causa_AEEG_2_livello, tipo_interruzione, configurazione_rete, *codice_utente*, PI, durata_disalimentazione, ENS_ENR_lorda, ENS_ENR_netta, NF_NR_SD, n_incidente_rilevante.

- **siti_utente:** codice_utente, anno_inizio, anno_fine, titolare, denominazione_impianto, codice_tipo_connessione, tensione_impianto, direttamente_connesso, altri_collegamenti, *località*, *provincia*.
- **località**: località, provincia, regione, area, lat, lon, alt.

Different accessory tables exist depending on the reporting period, e.g.: **guasti_accessorie_1012, guasti_accessorie_1315, guasti_accessorie_1619, guasti_accessorie_2021, guasti_accessorie_2224**. These typically store extra fields that are not consistently available across all years (e.g., additional ENS/ENR details, event qualifiers, or other year-specific fields).

B) Weather-related data

For meteorology, Moschetti et al. design two main entities: **stazione_meteo** and **measurement**. In implementation, measurement is represented as separate tables by variable to avoid unnecessary nulls (not all stations measure all variables).

Logical design:

- **stazione_meteo:** ID, codice_regionale, denominazione, *comune*, *provincia*, latitudine, longitudine, altitudine, frequenza.
- Measurements by variable:
    - **precipitazioni**: *id_stazione*, data, precipitazioni_totali
    - **neve**: *id_stazione*, data, neve_suolo, neve_fresca
    - **temperatura**: *id_stazione*, data, temperatura_media, temperatura_massima, temperatura_minima
    - **vento**: *id_stazione*, data, velocità_media, velocità_raffica.

C) Lightning

The lightning dataset is integrated as an independent entity, prepared for spatial queries and later aggregation at municipal/provincial level (and subsequently at site level). For integration purposes, an *id* field is added, and lightning strikes are geocoded to *comune* and *provincia*.

Logical design:

- **fulmini**: id, data, ora, lat, lon, intensità (kA), tipo (C/G), *comune, provincia.*

3.2.2. Intermediate Tables for Site-Day Reconstruction

After defining the national baseline schema and updating it with the 2024 extensions, the next step is to transform the raw relational structure into a set of **intermediate, materialised tables** that are directly usable for exposure reconstruction and for building the final site-day analytical dataset.

Rather than generating the final site-day table in a single expensive step, the pipeline is organised into **layers** where each table adds one specific piece of information (site selection, daily failure aggregation, daily station measurements, lightning aggregation). This staged approach provides three advantages:

- **Efficiency:** expensive joins and spatial/temporal aggregations are executed once and stored, avoiding repeated recomputation.
- **Verifiability:** each layer can be validated independently (coverage checks, missingness patterns, plausibility ranges).
- **Modularity:** the same intermediate layers can feed different reconstruction methods (IDW, RK, splines) and can be scaled beyond Piedmont in future work.

These tables form the "inputs" required by the interpolation and feature-engineering steps described later, and they ultimately support the controlled materialisation of the final site-day dataset.

3.2.2.1. Site-level Outage Aggregation

In this step, the raw outage event table (*guasti*) is transformed into a site-day outage summary (**siti_utente_guasti**) (Figure 3.2). Terna's outage records are stored at event level, meaning that multiple weather-related outages can occur at the same user site on the same calendar day. For the analysis, a daily representation is required in order to:

- Align outages with daily meteorological exposure
- Support risk estimation and binning on a consistent temporal grid
- Avoid repeatedly aggregating raw events during analyses

The transformation follows a simple **aggregation pipeline**, where event records are first filtered to the subset of interest, then projected to the daily grain (calendar day), and finally aggregated by (*codice_utente, data*) to compute (i) *weather_failure_count* (the number of weather-related outage events recorded for that site on that day) and (ii) *has_weather_failure* (a binary flag indicating at least one outage).

This table uses a sparse representation, since *siti_utente_guasti* **only stores days where *has_weather_failure* = 1** (i.e., days with at least one weather-related outage). When building the final analytical site-day table, *siti_utente_guasti* is **left-joined** to the complete site-day backbone, and missing matches are interpreted as "no weather-related outage recorded for that site-day" and are mapped to *has_weather_failure* = 0 and *weather_failure_count* = 0.

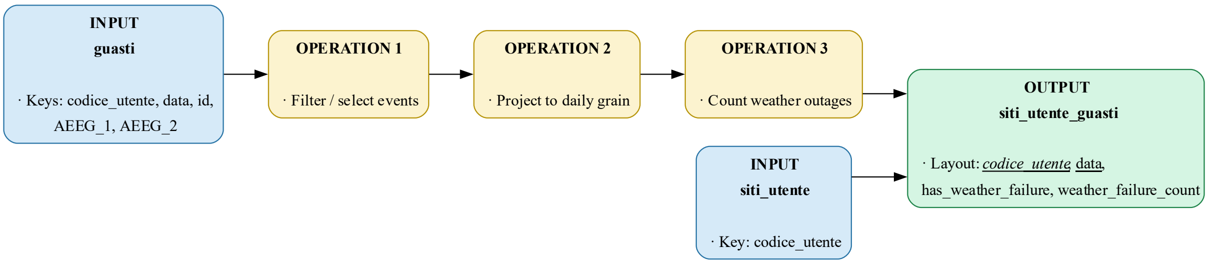


Figure 3.2 Data Lineage for the Site-Day Outage Summary Table

3.2.2.2. Weather Station Entity Improvement

To support a robust site-level meteorology reconstruction, the baseline entity *stazione_meteo* is complemented with an enriched table: **stazione_meteo_improved**. The key idea is to augment the station metadata with **operational coverage** and **variable availability flags**, which are essential for the subsequent IDW-based reconstruction pipeline.

In the baseline database, *stazione_meteo* provides the station identity and static attributes (coordinates, altitude, location, frequency). However, for interpolation at site-day resolution, two additional aspects become critical:

1. **Temporal coverage**: stations are not necessarily active for the full study period, and their measurements can be missing for long intervals. If this is not handled explicitly, interpolation may incorrectly include stations that are "nearby" but inactive on the target day (contributing only nulls).
2. **Measurement capability**: not all stations measure all meteorological variables. This is crucial for IDW because the "best" neighbours for one variable may be invalid for another; therefore, eligibility must be variable-specific.

As summarized in Figure 3.3, *stazione_meteo_improved* is built through a short lineage, where meteorological measurement tables are aggregated to derive (i) temporal coverage descriptors and (ii) capability flags, which are then joined back to the station registry:

- **date_inizio** and **date_fine**: the start and end date of each continuous operational period (i.e., consecutive days) in which the station reports at least one valid observation among the four meteorological blocks (*precipitazioni, temperatura, vento, neve*). Instead of storing a single global min/max over the full dataset, the logic identifies separate "spells" of activity; if a station stops providing data for a long gap and later becomes operative again, it generates a new period with a new (*date_inizio, date_fine)* pair. For this reason, the table uses (*id, date_inizio*) as the primary key, allowing multiple operational intervals per station and ensuring that inactive years are not mistakenly considered during interpolation (a station is eligible only if the target day falls within one of its active intervals).

- **Capability flags (*has_rain, has_temp, has_wind, has_snow*):** binary indicators (1/0) that specify whether the station provides measurements for each of the four meteorological blocks. These flags are computed by checking whether at least one record exists for the station in each corresponding measurement table (or, equivalently, whether daily aggregates can be built for that variable).

Overall, *stazione_meteo_improved* should be interpreted as a **derived station reference table** used to make neighbour selection consistent in both **time** (operativity spells) and **variable space** (capability), thereby improving the robustness of the site-day interpolation stage.

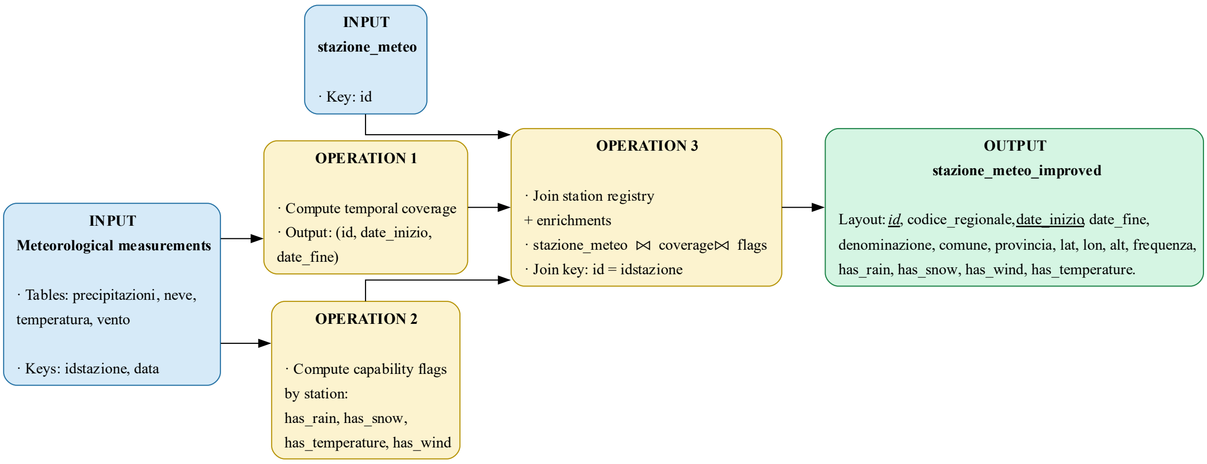


Figure 3.3 Data Lineage for Station Metadata Enrichment

3.2.2.3. Daily Weather Variables

The baseline database stores meteorological observations as station measurements, whose native **sampling frequency** depends on the specific station (e.g., 15-min, 30-min, hourly, or already daily). However, the analytical pipeline of this work is built at site-day resolution. Many downstream meteorological moderators (such as moving accumulations, persistence indicators, threshold counts, compound flags) can only be computed consistently if the input data are available as **one value per station per day**.

For this reason, the original measurement tables are materialised into four daily helper tables, one per variable family, each providing a **daily time series per station**. As shown in Figure 3.4, the process first projects timestamps to the daily grain, then performs a variable-specific daily aggregation, resulting in the four outputs. These helper tables act as the standardized input layer for the subsequent interpolation and feature-engineering phases.

In the Piedmont case study, the ARPA products used in this work are already available at daily resolution; nevertheless, these tables are still created to (i) enforce a uniform data model, (ii) make the computation of moderators independent of the original sampling frequency, and (iii) keep the pipeline portable to other regions or future national-scale implementations where heterogeneous or sub-daily sampling is more common.

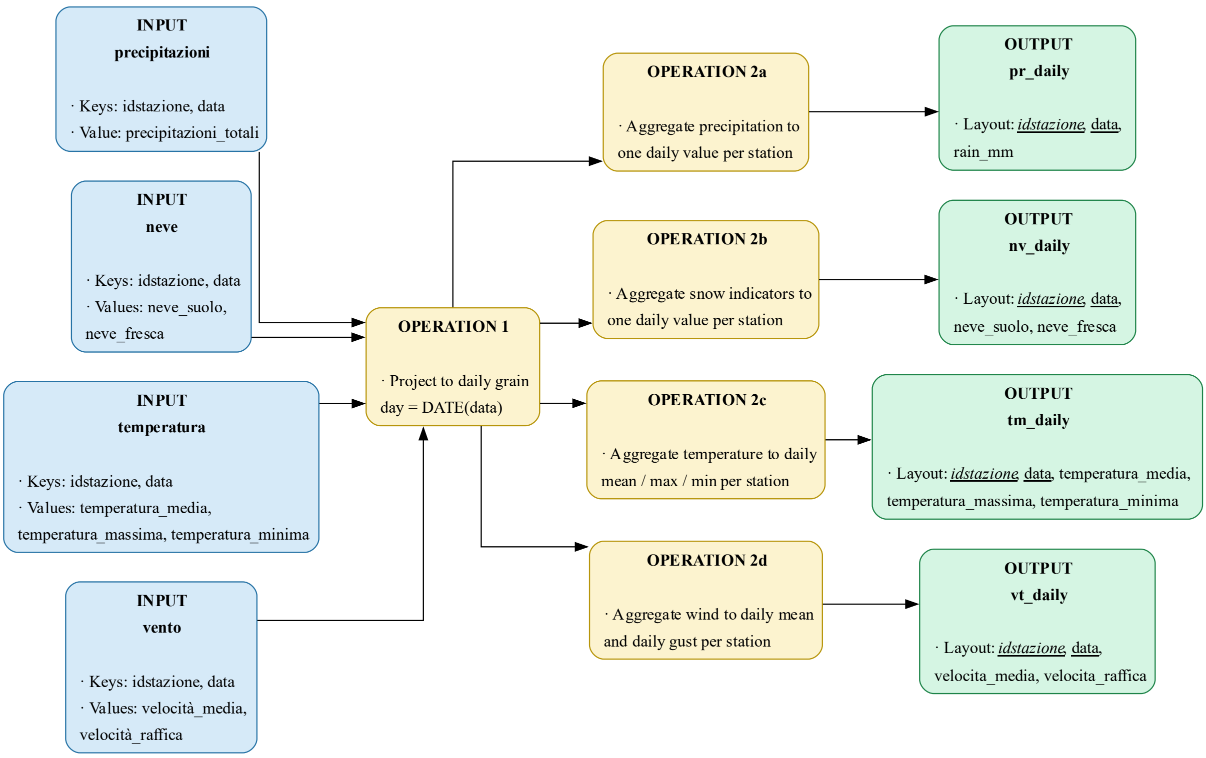


Figure 3.4 Data Lineage for Daily Meteorological Helper Tables

3.2.2.4. Site-day Lightning Exposure Features

In this step, the raw lightning strike table (*fulmini*) is transformed into a site-day lightning exposure table (**site_day_fulmini**) by aggregating individual strikes into daily indicators for each Terna user site. The objective is to convert event-level lightning information into a compact and interpretable set of features that can be merged into the final analytical dataset without repeatedly performing expensive spatial computations.

As summarized in Figure 3.5, the procedure combines the spatial location and validity interval of each site (*siti_utente*) with the georeferenced lightning strikes (*fulmini*). For each site and for each day in the analysis window (here **2008-2023**), candidate strikes are first reduced through a **coarse spatial pre-filter** (bounding box around the site) and temporal consistency checks (global period filter and compatibility with the site's validity years). For the retained candidates, the **site-strike distance** is computed using the Haversine formula, and only strikes within a **20 km radius** are kept. Finally, the retained strikes are aggregated at **(codice_utente, data)**. For each user site and each day in the analysis window, the pipeline:

- **Counts within distance bands** (≤ 20 km, ≤ 10 km, ≤ 5 km)
- **Proximity and intensity summaries** (minimum strike distance and daily maximum/average intensity)

The resulting table is **sparse by construction**, containing rows only for site-days with at least one strike within 20 km. When integrating lightning into the complete site-day analytical backbone, missing matches after a left join are interpreted as no detected lightning exposure within 20 km for that site-day (with count fields set to 0; continuous summaries can be left as null).

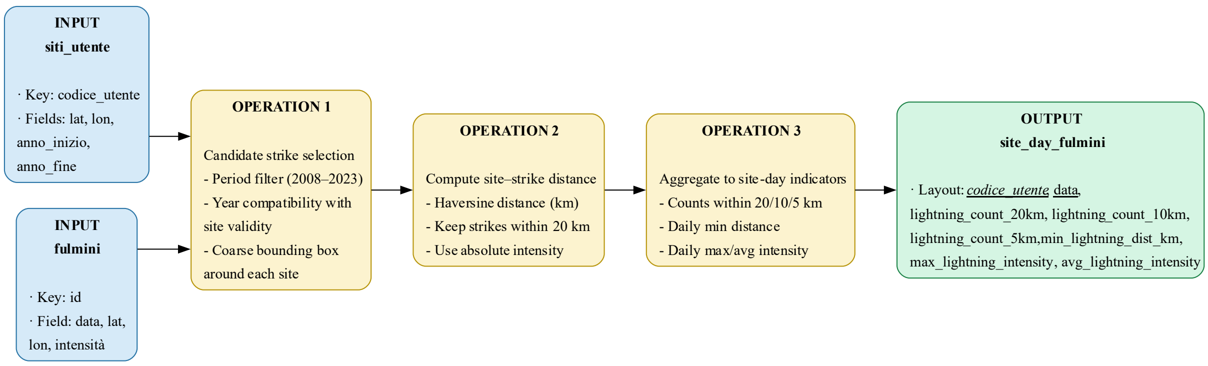


Figure 3.5 Data Lineage for Site-Day Lightning Exposure

3.2.3. IDW Analysis

3.2.3.1. Intermediate *Site-Day-Stations* Table

After generating the intermediate tables in Section 3.2.2, the next step in the IDW-based pipeline is to materialize an intermediate panel that makes the relationship between each user site and its relevant weather stations explicit at daily resolution. The resulting table, **siti_utente_stazione_daily**, provides (i) the fixed set of up to five neighbour stations selected for each user site and (ii) their corresponding daily observations, aligned on a common site-day time axis.

The construction of *siti_utente_stazione_daily* follows the four-step workflow illustrated in Figure 3.6:

1. **Site profiling and geocoding.** Site records are first consolidated at *codice_utente* level to define each site validity interval. The site is then geocoded by joining the locality registry to obtain site coordinates (latitude, longitude and altitude).
2. **Neighbour station selection (per site).** Candidate stations from **stazione_meteo_improved** are evaluated against each site using spatial and orographic criteria (distance and altitude difference), together with station availability information (coverage and capability flags). A score is computed from distance and altitude mismatch, and up to **five** stations are selected per site: three best-ranked "base" stations and up to two additional stations to improve coverage of missing variable families when needed.
3. **Daily expansion (site-day backbone).** The site-level mapping is expanded over time by joining *date_dim* within each site's validity window, producing one row per *(codice_utente, day)*.
4. **Attachment of daily station values.** For each site-day, the five selected stations are replicated as station slots (i = 1..5) and their daily observations are attached via left joins to the daily helper tables (*pr_daily, nv_daily, tm_daily, vt_daily*) using *(idstazione_i, day)*. Missing station measurements naturally appear as nulls at this stage and are handled during the interpolation and quality-control steps.

This intermediate table is primarily motivated by efficiency and auditability. IDW requires repeated access to a small set of nearby observations for each site and day; computing neighbour selection and station lookups "on the fly" would be expensive and would make the process harder to trace. By materializing the neighbour panel, the subsequent IDW reconstruction becomes a row-wise weighted aggregation using a fixed, documented set of neighbours. In addition, storing distance, altitude difference, and score makes the selection fully reproducible and supports later inspection of which stations contributed to each site-day estimate.

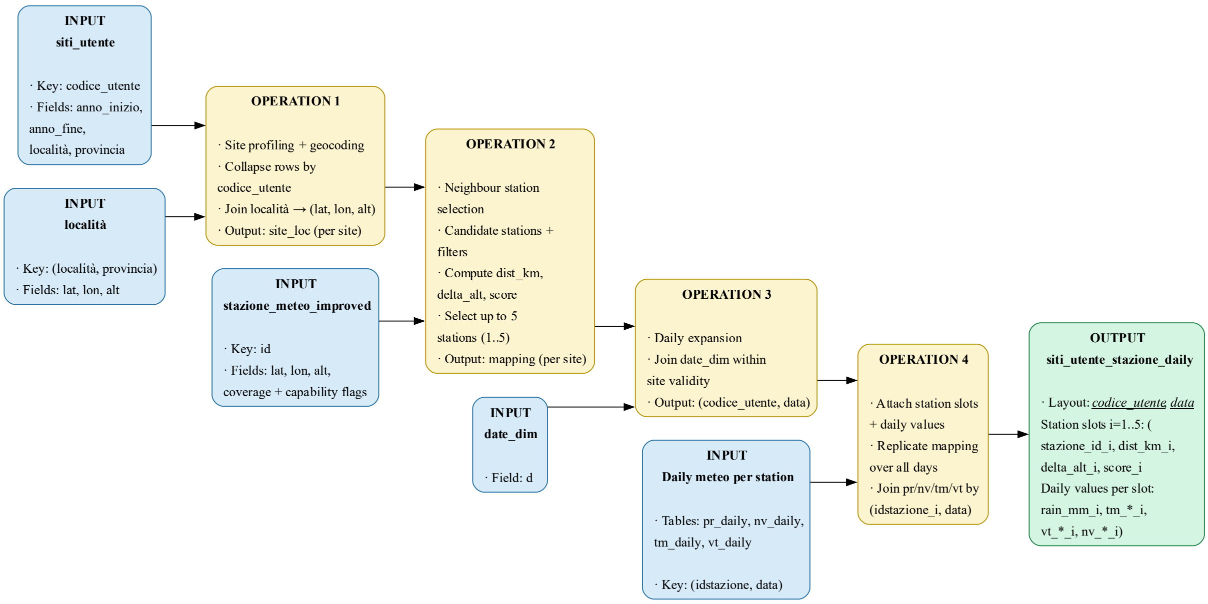


Figure 3.6 Data Lineage for the Site-Day Neighbour-Station Panel

3.2.3.2. Final IDW Analytical Table

Once the intermediate table *siti_utente_stazione_daily* has been materialised, the baseline analytical dataset for the IDW pipeline is obtained by consolidating three site-day layers into a single table, **siti_utente_weather_guasti** (Figure 3.7).

The construction follows a simple data lineage at site-day grain *(codice_utente, data)*:

i. Reconstructed daily meteorology at site level (via IDW)
ii. Daily lightning exposure features
iii. Daily weather-failure indicators

Conceptually, *siti_utente_weather_guasti* is a site-day table where each record corresponds to one *codice_utente* on one *data* and contains:

1. **IDW site-level reconstruction**: for each site-day, the daily meteorological values observed at the selected neighbour stations (up to five) are combined through an IDW scheme based on the station scores stored in

siti_utente_stazione_daily. This step produces site-level daily estimates for the four meteorological families (precipitation, temperature, wind, snow) together with a reliability indicator per variable.

2. **Merge lightning exposure**: the daily lightning exposure indicators computed in *site_day_fulmini* are merged at the same *(codice_utente, data)* resolution.
3. **Merge weather-failure outcomes**: the daily weather-related outage indicators from *siti_utente_guasti* are merged at *(codice_utente, data)*, providing the binary outcome and the daily count of weather-related events.

As a result, *siti_utente_weather_guasti* is a single, **analysis-ready site-day table** where each record corresponds to one user site on one day and contains (i) reconstructed daily meteorology at the site, (ii) lightning exposure features, and (iii) the corresponding weather-failure label/count. This table is the final output of the baseline IDW reconstruction branch and it could be directly used in downstream steps.

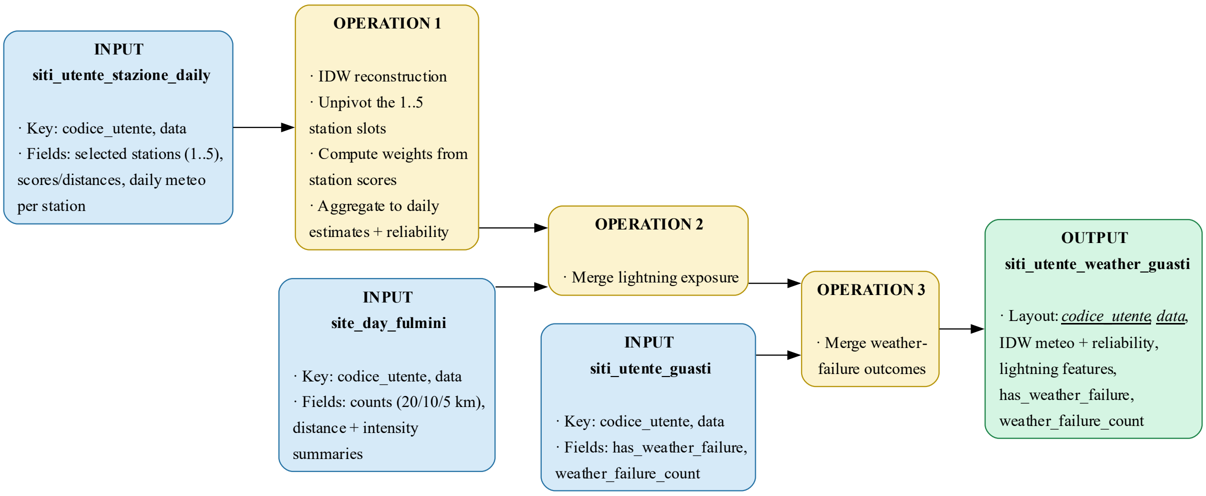


Figure 3.7 Data Lineage for the Final IDW Analytical Table

In the next step, an analogous analytical table is produced using **RK and Spline** reconstructions which is the version ultimately adopted for the main results.

3.2.4. RK and Spline Analysis

Once the baseline IDW analytical table (*siti_utente_weather_guasti*) is available, an improved reconstruction branch is built using **spline-based interpolation** (for precipitation and snow) and **RK** (for temperature). The outcome is a second site-day analytical table, **siti_utente_weather_guasti_improved**, designed to be structurally consistent with the IDW version while replacing specific meteorological fields with the best-performing reconstruction method for each variable family.

3.2.4.1. Spline Reconstruction Table

Daily Precipitation at Site Level

This table contains the daily precipitation reconstructed at each user site via spline based on Radial Basis Functions (RBF) interpolation [49]. As shown in the data lineage, the pipeline is organised into four conceptual steps:

1. **Site-day backbone:** a complete daily series is generated for each site by expanding *siti_utente* over the calendar (*date_dim*) within the site validity window. This ensures a controlled and consistent site-day grid.
2. **Valid station-day observations:** daily precipitation observations from *pr_daily* are paired with station metadata from *stazione_meteo_improved*, retaining only stations with valid geometry and active coverage on the target day.
3. **Spline interpolation to sites:** a spline (RBF) model reconstructs precipitation at site coordinates using local neighbouring stations. Altitude is included as an additional dimension and robustness controls are applied to reduce artefacts (e.g., overshoot or "phantom rain").
4. **Write results:** the reconstructed precipitation time series is written to the output table.

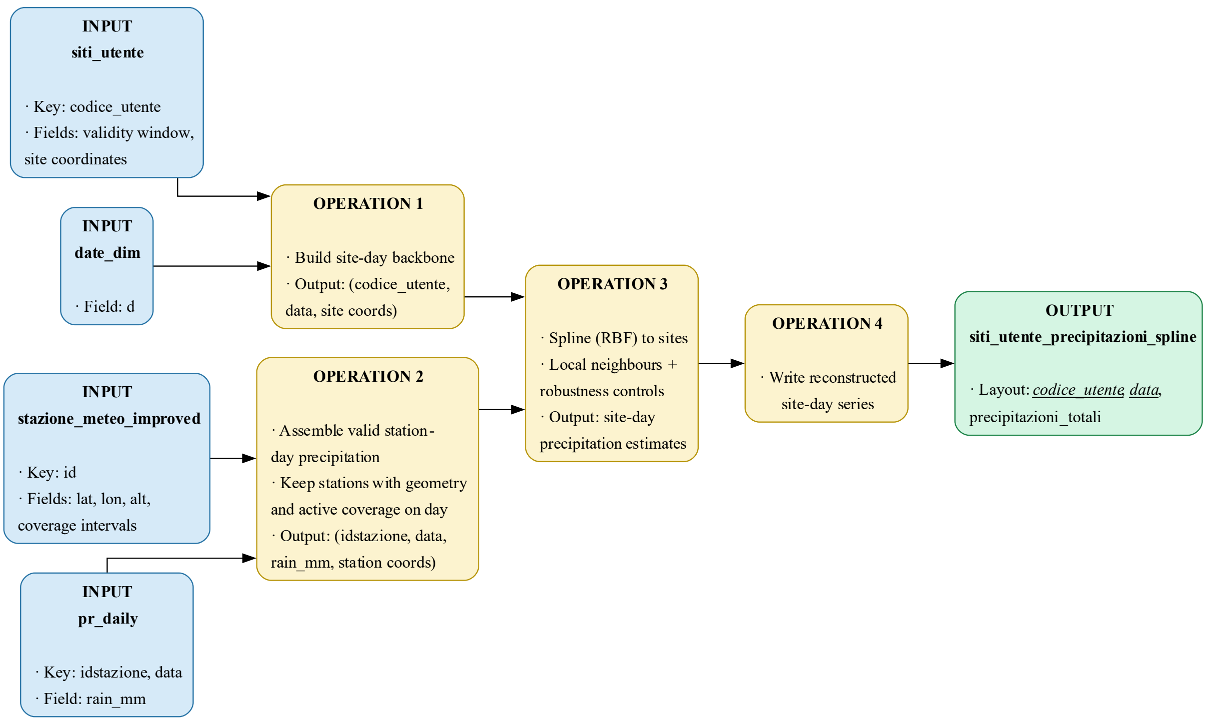


Figure 3.8 Data Lineage of Spline-based Daily Precipitation Reconstruction

Daily Snow at Site Level

This table follows the **same pipeline structure** (Figure 3.9) as the precipitation spline reconstruction (site-day backbone from *siti_utente* + *date_dim*, selection of valid station-day observations using *nv_daily* and *stazione_meteo_improved*, spline/RBF interpolation to site coordinates with altitude-aware neighbourhoods, and write-out of the reconstructed series). The only difference is that the reconstruction targets the snow variables, producing two daily site-level outputs on the same *(codice_utente, data)* grid: *neve_suolo* and *neve_fresca*.

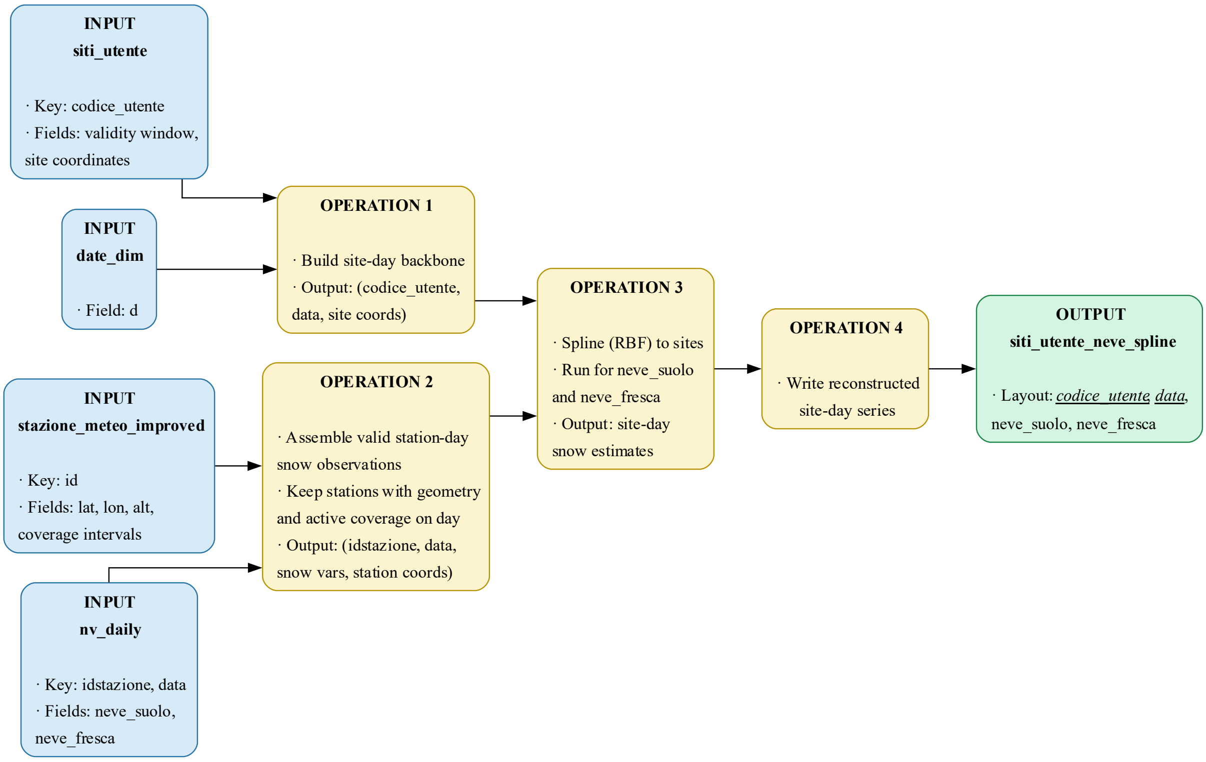


Figure 3.9 Data Lineage of Spline-based Daily Snow Reconstruction

Both spline tables are explicitly built at site-day resolution using *(codice_utente, data)* as the join key. By relying on *date_dim*, the site validity window, and station coverage constraints from *stazione_meteo_improved*, they produce reconstructions that are immediately joinable with the other daily layers (lightning exposure, failures, and the final analytical tables) without requiring repeated spatial computations.

3.2.4.2. RK Reconstruction Table

In parallel with the spline approach adopted for precipitation and snow (Section 3.2.4.1), daily temperature at site level is reconstructed through a hybrid approach that combines RK for the daily mean temperature and an IDW-based derivation for daily minima and maxima.

The process starts by building a site-day backbone. Each user site is geocoded through the locality registry (*lat, lon, alt*) and expanded over the calendar table (*date_dim*) within its validity window. In parallel, the pipeline assembles the valid station-day observations from *tm_daily*, keeping only stations with valid geometry and an active coverage interval on the target day (using *stazione_meteo_improved*). This ensures that only measurements from operative stations contribute to the reconstruction.

For each day, RK is applied to daily mean temperature. The method decomposes the station observations into:

- A **deterministic trend** explained by geographic covariates (e.g., altitude and coordinates)
- A **spatial residual component**, modelled via kriging using **monthly variogram parameters** (selected per month to capture seasonal differences in spatial structure)

This produces a site-day estimate of *Tmean*. Then, *Tmin* and *Tmax* are derived around the reconstructed mean by interpolating, via IDW, the daily deviations (amplitudes) observed at stations, and recombining them with the site-level *Tmean* estimate.

Finally, the reconstructed daily series are written into the table **siti_utente_temperatura_kriging**, which mirrors the design of the spline outputs (one row per site and day) and can therefore be joined directly with the other site-day layers (failures, lightning, reconstructed precipitation/snow, etc.) using the composite key *(codice_utente, data)*.

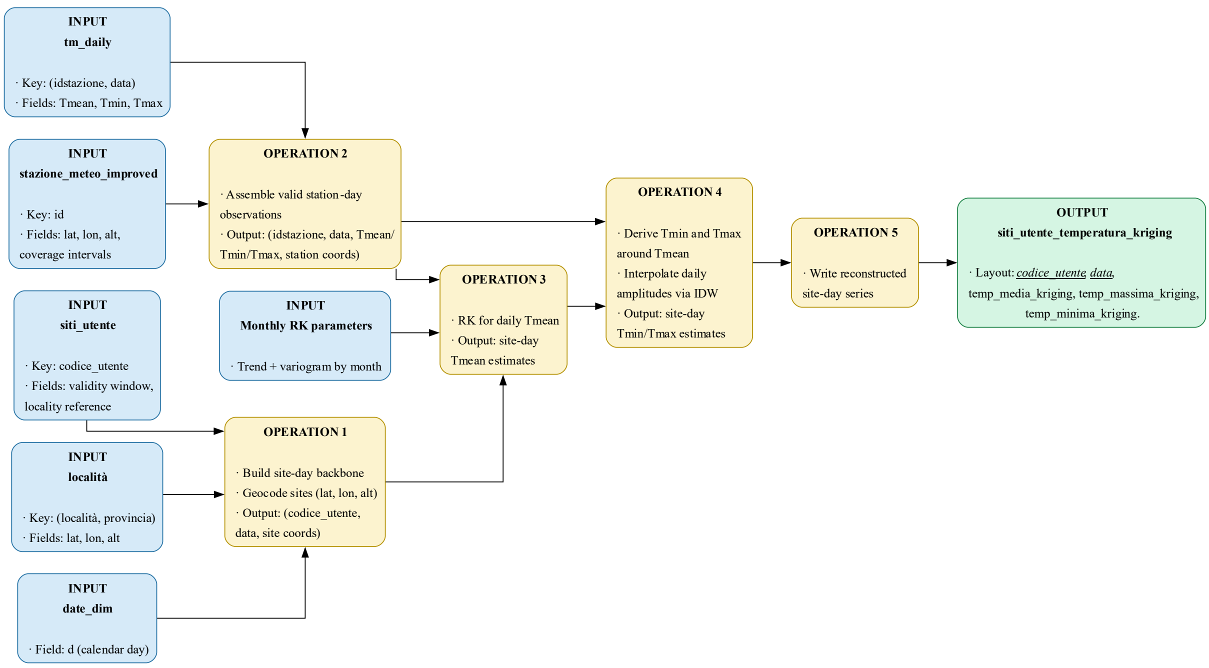


Figure 3.10 Data Lineage for RK-based Site-Day Temperature Reconstruction

3.2.4.3. Improved Analytical Table

After generating the improved site-level weather reconstructions, the pipeline consolidates all exposures and outage labels into a single denormalized site-day analytical table: **siti_utente_weather_guasti_improved**. This table is the main input for the downstream statistical analyses, because it provides one record per (*codice_utente, data*) where all explanatory variables and response variables are already aligned on the same site-day axis.

As shown in the data-lineage diagram (Figure 3.11), the build logic is organized in four conceptual steps:

1. **Initialize the site-day backbone.** The baseline analytical table *siti_utente_weather_guasti* (IDW branch) is used as the reference spine for the analytical dataset. Since it already materialises the full site-day grid, it guarantees that all (site, day) combinations are represented, and it provides the baseline fields that are kept unchanged in the improved product.
2. **Merge reconstructed meteorology.** The backbone is enriched with the best-performing weather reconstructions, all at site-day resolution:
   a. *siti_utente_precipitazioni_spline* → daily precipitation at site level (*rain_mm_spline*),
   b. *siti_utente_neve_spline* → daily snow variables (*neve_suolo_spline, neve_fresca_spline*),
   c. *siti_utente_temperatura_kriging* → daily temperature (*temp_media_kriging, temp_massima_kriging*, *temp_minima_kriging*),
   d. *siti_utente_weather_guasti* → wind variables reconstructed with IDW (*velocita_media_idw*, *velocita_raffica_idw*).
3. **Merge lightning exposure.** Finally, *site_day_fulmini* is merged at the same (site, day) key to attach lightning features (counts within 20/10/5 km and proximity/intensity summaries). The result is written as the improved analytical table.
4. **Merge outage outcomes and write the table.** Finally, *siti_utente_guasti* is merged to attach the response variables (*has_weather_failure, weather_failure_count*) on the same site-day index.

Overall, *siti_utente_weather_guasti_improved* is designed to be structurally compatible with the baseline IDW analytical table, while replacing precipitation/snow/temperature with the improved reconstruction methods and keeping wind from the IDW branch. This ensures that all later steps (binning, thresholds, persistence indicators, compound events, etc.) can operate on a single consistent site-day dataset.

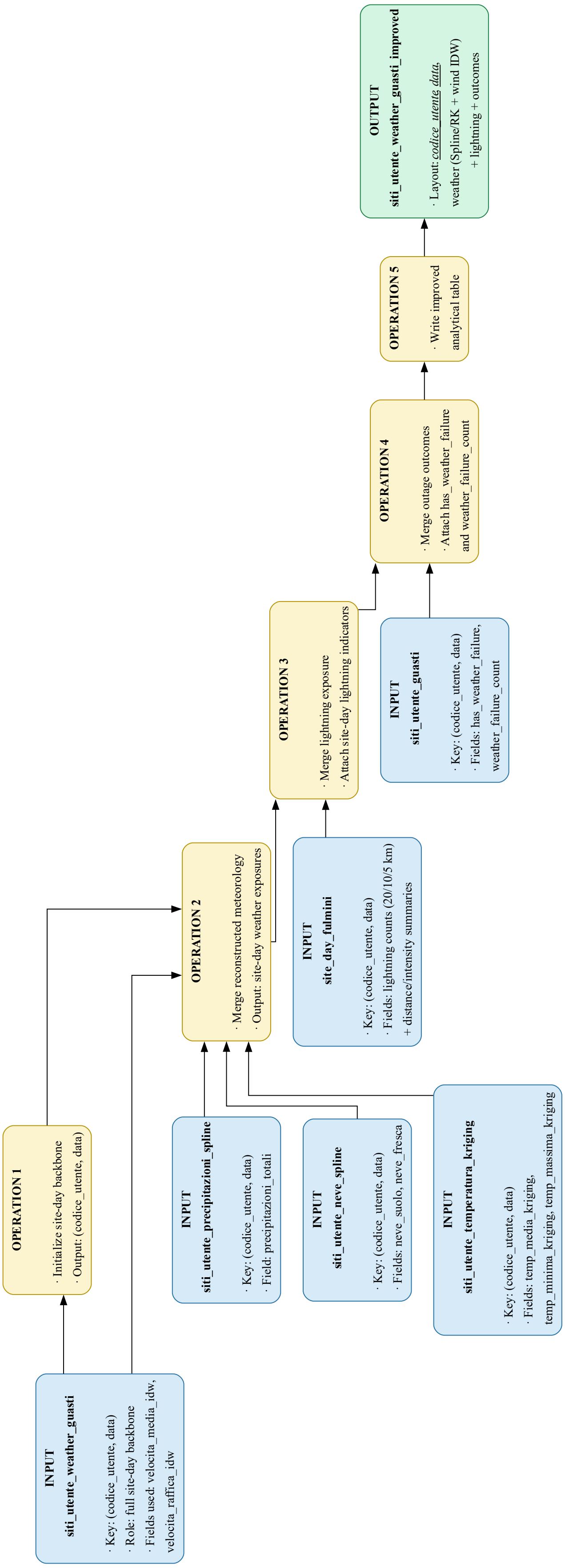


Figure 3.11 Data Lineage for the Improved Site-Day Analytical Table

3.3. Final Dataset: Site-Day Table

At this stage, two site-day weather-outage products are available for the subsequent analyses: *siti_utente_weather_guasti* (IDW-based reconstruction) and *siti_utente_weather_guasti_improved* (Spline + RK + IDW). Although both tables share the same analytical unit, the validation checks performed during the reconstruction phase indicated that *siti_utente_weather_guasti_improved* provides more reliable and physically consistent exposure estimates. For this reason, it is adopted as the reference input for the final analytical dataset.

Starting from *siti_utente_weather_guasti_improved*, an additional step is carried out to generate a fully analysis-ready table, hereafter referred to as **final_site_day_analysis**. The goal is to **enrich the site-day panel** with derived features that capture persistence, thresholds, transitions, and compound conditions, i.e., the types of mechanisms that are commonly associated with weather-driven failures and that cannot be represented by raw daily values alone.

The analytical table can be organised into six functional blocks:

1. **Index and core daily variables**

The table preserves the same primary key (**codice_utente, data**) and retains the **core daily exposures** at site level, together with the Terna-derived outcomes (*has_weather_failure, weather_failure_count*). This is the minimal structure required to relate daily weather conditions to failures on a consistent site-day grid.

2. **Lightning descriptors and data availability**

Lightning features are included both as **multi-radius counts** (e.g., within 20/10/5 km) and as **proximity/intensity summaries** (minimum distance, maximum/average intensity). A dedicated indicator (e.g., *lightning_available*) is stored to explicitly distinguish true zero lightning activity from years with missing lightning coverage, ensuring that later models do not conflate "no events" with "no data".

3. **Persistence and rolling aggregations (memory effects)**

To capture multi-day accumulation and persistence mechanisms, the table adds rolling features such as:

- **Rain accumulations and wetness proxies** (e.g., 3/7/14-day sums, wet-day counts, an AP index soil wetness indicator)
- **Temperature persistence** (weekly/monthly means, recent extremes, cooling/heating degree days)
- **Wind persistence and extremes** (7-day mean wind, 3/7-day gust maxima, moderate-wind day counts)
- **Snow persistence** (multi-day fresh snow sums and maximum ground snow in recent windows)

- **Lightning persistence** (rolling sums over recent days at multiple radii)

4. **Transitions and compound-event flags**

Beyond persistence, specific **transition indicators** are computed to represent abrupt changes and combined hazards, such as:

- Temperature jumps and freeze/thaw transitions
- Dry-to-wet switches relevant for insulation/contamination effects
- Wind ramp-up events
- Compound conditions like ROS, which is often associated with rapid melt and runoff processes

These features are designed to support analyses focused on trigger-like behaviour rather than gradual accumulation.

5. **Time and geography descriptors (stratification controls)**

The final table also includes contextual variables that support stratification and heterogeneity testing: calendar descriptors (year, month, season) and site descriptors (province, altitude). These fields make it possible to separate seasonal regimes, compare provinces, and control for altitude-related exposure gradients without requiring additional joins.

6. **Binning and regime indicators (categorical supports for risk estimation)**

Finally, continuous exposures are complemented with **categorical binning and regime flags** (for rain, wind, temperature, snow, wetness, and lightning). These derived categorical variables provide:

- A robust basis for **risk curves and RR estimation** (e.g., comparing failure frequency across bins)
- Interpretable interaction testing (e.g., wet + windy, cold + wet, lightning near + windy)
- Stable group definitions for summarization and diagnostics

Two aspects make *final_site_day_analysis* the central dataset for the work. First, it **unifies exposure and outcomes on a single site-day index**, which is a prerequisite for defensible attribution and risk estimation. Second, it supports both **macro-scale summaries** (e.g., monthly aggregations, regime-based comparisons) and the **full-resolution daily panel** needed for interaction testing, persistence effects, and spatial diagnostics, without changing dataset or redefining the analytical unit.

4 Methodology

This chapter describes the analytical methods applied to the enriched site-day database introduced in Chapter 3. The workflow has four steps:

1. Reconstruct daily meteorological data at each user site from the ARPA station network
2. Generate higher-level moderators that capture persistence, antecedent load, extremes, and compound conditions
3. Quantify associations between weather and outages using interpretable monthly and site-day analyses
4. Operationalise the evidence into a site-day probabilistic risk score (FPI) with walk-forward validation and calibration

4.1. Meteorology at User Sites

This section reconstructs daily weather at site-level for each Terna user site located in Piedmont throughout the analysis window. The ARPA station network is heterogeneous in terms of spatial distribution, elevation, temporal coverage, and availability of variables; therefore, a naïve assignment to the nearest station is not robust, especially in complex terrain.

Two complementary reconstruction layers are implemented:

- **Baseline IDW layer** (Section 4.1.1). A transparent approach based on compact, fixed mapping of up to five neighbour stations per site, used to calculate daily IDW estimates and an operational reliability indicator.
- **Advanced spatial interpolators** (Sections 4.1.2-4.1.3). RK for temperature and spline-based interpolation for precipitation and snow, each with its own neighbourhood logic (without reusing the fixed mapping of 5 stations). Wind remains treated with IDW due to the practical limitations discussed in Section 4.1.4.

4.1.1. Station Selection Logic and Baseline IDW

A score-based IDW [50] workflow is used to reconstruct daily meteorology at each user site. The goal is to (i) provide a simple and reproducible site-day estimate based on observations from nearby stations; and (ii) associate a reliability indicator that flags weak support due to geometry, sparse contributors, or poor cross-station agreement. All inputs (daily station tables) and station availability metadata are those defined in Chapter 3.

For each site, candidate stations are extracted from the enriched station registry (coverage period + capability flags). A station can contribute on day t only if it is operative on that day and meets locality constraints:

- A coarse spatial pre-filter
- **Distance** constraint $d \leq 30\text{km}$
- **Altitude mismatch** constraint $\Delta alt \leq 150\text{m}$,
- The requirement that the station provides at least one of the meteorological blocks.

Candidates are ranked using a dimensionless composite score:

$$s_i = \left(\frac{d_i}{30}\right)^2 + 2\left(\frac{\Delta alt_i}{150}\right)^2 \tag{4.1}$$

The quadratic penalties discourage candidates that are acceptable in one dimension but poor in the other, and the altitude term is weighted more strongly to reflect the relevance of orography in mountainous areas. Each site is assigned a compact and fixed neighbour set of up to five stations:

- **Base set (3 stations):** the three stations with the lowest score.
- **Complementary set (up to 2 stations):** selected to complete coverage of missing variable families when needed, or to add redundancy when coverage is already complete.

For a given variable x and a target site-day, the IDW estimate uses only stations with non-missing x_i on that day. Weights are derived from the score [50]:

$$\omega_i^{raw} = \frac{1}{s_i + \varepsilon} \quad , \quad \omega_i = \frac{\omega_i^{raw}}{\sum_{j=1}^{n} \omega_j^{raw}} \quad , \quad \hat{x} = \sum_{i=1}^{n} \omega_i \, x_i \tag{4.2}$$

where ε is a small numerical tolerance to avoid floating-point division issues when s_i becomes extremely small. In the (rare) case where a station coincides with the user site (i.e., $s_{\min} \approx 0$, within a small threshold τ), no interpolation is performed and the site-day estimate is set directly to that station's observation (weight 1 for the coincident station, 0 for the others). Otherwise, standard score-based IDW weighting is applied across all available stations.

Each IDW estimate is accompanied by a reliability **index** $R_{total} \in [0,1]$, defined as:

$$R_{total} = R_{geom} \cdot R_n \cdot R_{cons} \tag{4.3}$$

where:

1. **Geometric:** let $\bar{s} = \sum_{i=1}^{n} \omega_i \, s_i$ and map it to

$$R_{geom} = 1 - \min\left(1, \frac{\bar{s}}{3}\right) \tag{4.4}$$

2. **Contributors:** with n available station values,

$$R_n = \min\left(1, \frac{n}{3}\right) \tag{4.5}$$

3. **Consistency:** dispersion around the estimate

$$\sigma_\omega = \sqrt{\sum_{i=1}^{n} \omega_i \, (x_i - \hat{x})^2} \tag{4.6}$$

is converted into $CV = \frac{\sigma_\omega}{|\hat{x}| + \varepsilon}$ and mapped to:

$$R_{cons} = \frac{1}{1 + CV} \tag{4.7}$$

clipped to [0,1]. For $n = 1$, $R_{cons} = 0.5$.

This reliability index is a **quality flag**. It is used for diagnostics and explicit handling of low-support site-days. In the exact-match case, the reliability is set to 1, as the estimate is a direct measurement at the site location.

4.1.2. Temperature: Regression Kriging

The temperature in Piedmont has a strong deterministic structure (altitude and broad spatial gradients) as well as a spatially correlated residual. The daily mean temperature is reconstructed using RK [51], decomposing the field as:

$$T_{mean}(x,t) = m_m(x) + r_m(x,t) \tag{4.8}$$

where $m_m(\cdot)$ is a month-specific regression trend and $r_m(\cdot)$ is the residual field interpolated by Ordinary Kriging (OK).

For each calendar month m, a parsimonious regression on altitude and coordinates is fitted:

$$m_m(x) = \beta_0 + \beta_1\, alt + \beta_2 \frac{alt^2}{10^6} + \beta_3\, lat + \beta_4\, lon \tag{4.9}$$

Residuals at stations are:

$$r_m(x_i, t) = T_{\text{mean,obs}}(x_i, t) - m_m(x_i) \tag{4.10}$$

A monthly variogram is estimated from within-day station pairs to separate spatial structure from temporal variability, then a spherical model is fitted [52]. In daily production, residuals are predicted via **local OK** (distance-ranked neighbours with caps and fallback):

$$\hat{r}_m(x_0, t) = \sum_i \lambda_i\, r_m(x_i, t), \quad \hat{T}_{\text{mean,RK}}(x_0, t) = m_m(x_0) + \hat{r}_m(x_0, t) \tag{4.11}$$

T_{min}/T_{max} are derived by interpolating daily deviations around the mean (amplitudes) using standard IDW (power $p = 2$), then recombining with $\hat{T}_{mean}$ (with ordering enforced).

The RK outputs were validated with basic QC checks and with a quantitative station-based evaluation. Specifically, station observations were paired to the nearest Terna site within a 5 km radius and compared against the reconstructed site-day values using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), plus related error summaries. The resulting performance metrics and their temporal breakdowns are reported in Section 5.1.1.

4.1.3. Precipitation and Snow: Spline-based Approach

Precipitation and snow are intermittent, with heavy tails, and strongly shaped by the terrain. They are reconstructed using **local RBF spline** surfaces [53] with explicit robustness safeguards.

Interpolation is performed in an **effective 3D space** where altitude mismatch increases the separation [54]. Altitude is converted into equivalent horizontal distance:

$$z = \left(\frac{\text{alt}_m}{1000}\right) \cdot \text{ALT\_EQ}_{km/1000m} \tag{4.12}$$

where ***ALT_EQ* = 7 km per 1000 m** elevation difference.

To stabilize extremes while retaining sensitivity on light rain, precipitation is interpolated in log space [55]. Let $r_i(t)$ denote the observed daily precipitation (mm) at station i on day t, and define the transformed variable $y_i = \log\big(1 + r_i(t)\big)$, which is used for spline fitting. The spline predicts $\hat{y}(t)$ at the target site-day (in log space), which is then back-transformed to obtain the precipitation estimate in mm (hats denoting interpolated (predicted) quantities at the site):

$$\hat{r}(t) = \exp\big(\hat{y}(t)\big) - 1 \tag{4.13}$$

A local model is fitted per site-day on a neighbourhood of nearest stations (limited by availability). The production configuration uses a **linear RBF kernel** with **non-zero smoothing** to avoid unstable exact interpolations.

Snow variables are reconstructed with the same geometry and local RBF approach, enforcing non-negativity (with different smoothing settings for each variable).

Three safeguards prevent spline artefacts:

1. **Wet/dry gating** (force 0 if no neighbour exceeds an event threshold)
2. **Local caps** based on nearby observed maxima (plus a margin)
3. **Post-processing** to remove tiny positive artefacts and avoid spurious precision

Spline reconstructions were evaluated in two complementary ways: (i) station-level cross-validation, systematically withholding stations and recomputing daily interpolations to measure generalization, and (ii) a station-to-nearest-site comparison (within 5 km, using the same effective 3D distance) to assess representativeness at the grid-site resolution. Performance was summarized with MAE/RMSE (and tail-error diagnostics) consistently across variables and interpolation methods (Section 5.1.1).

4.1.4. Wind and Practical Limitations

Wind is strongly controlled by local exposure (channelling in valleys, ridges, sheltering) and has a smaller effective observing network. In this scenario, more sophisticated interpolators did not provide stable improvements. Therefore, wind is reconstructed using the conservative score-based IDW layer and is interpreted with caution in subsequent analyses.

Wind speed and gust reconstructions from the baseline IDW layer were validated through basic QC and a quantitative station-based evaluation. As for temperature, station observations were paired to the nearest Terna site within a 5 km radius and compared against the reconstructed site-day values using MAE/RMSE (Section 5.1.1).

4.2. Feature Engineering

Daily exposures are transformed into moderators representing the mechanisms highlighted in Chapters 0-2: persistence, accumulation, memory effects, extremes, and compound regimes.

For each site, daily series are processed chronologically and transformed using fixed-length rolling windows $w \in \{3, 7, 14, 30\}$ days:

- **Rolling sums** (e.g. precipitation/snow/lightning accumulation)
- **Rolling event counts** above a low threshold (persistence by frequency)
- **Rolling mean/max/min** where appropriate (regime vs peak behaviour)

To represent hydrological memory beyond simple accumulations, an AP index is computed as an exponential decay filter on daily precipitation:

$$AP_{\mathrm{t}} = P_t + k\,AP_{t-1}, \qquad 0 < k < 1 \tag{4.14}$$

with $k \approx 0.85$ (moderate persistence). The AP index is computed only over **continuous runs with valid daily precipitation**; gaps reset evaluability.

Continuous exposures are mapped to interpretable regimes through a mixed thresholding strategy:

- **Quantile-based thresholds** computed within local strata (e.g., **province × season/month**) so that “extreme” is defined relative to the local baseline
- For precipitation, quantiles are computed **conditional on rain>0**, while **no-rain** is kept as an explicit baseline bin
- **Fallback thresholds** are used when a stratum has insufficient data (replace local quantiles with pooled/global quantiles to preserve full bin coverage)
- **Domain-informed cutoffs** where they are more meaningful than quantiles:
  - **Near-freezing band** around 0°C (freeze-thaw / icing relevance)
  - **Snow load levels** using fixed depth intervals (load/accessibility meaning)
- **Lightning bins** are defined in two complementary categorical views:
  - **Proximity bands** (minimum strike distance)
  - **Activity/intensity level** with an explicit “no lightning” baseline

Once thresholds are available, values are discretized into a small set of ordered bins designed for transparent comparisons:

- **Rain intensity**: none → light → moderate → heavy → very heavy
- **Wind intensity**: calm → increasing gust regimes (quantile-based)

- **Temperature regimes**: cold / normal / near-zero / heat (tail quantiles + near-zero band)
- **Snow load**: none → increasing depth classes (domain edges)
- **Antecedent wetness**: dry / normal / wet / very wet, built from a wetness proxy

Multi-hazard and **trigger-like scenarios** are represented by:

- **Binary scenario flags** (e.g., wet + windy, near-freezing + wet, AP + wind, ROS)
- **Interaction bins** (e.g., rain×wind, temperature×wetness, snow×near-freezing, and lightning interactions)
- **Transition flags** (dry→wet onset, gust ramp-up, freeze/thaw entry/exit)

Low-support bins, categories and sparse interaction cells are explicitly flagged to avoid unstable rate estimates.

4.3. Analysis Design

The analysis quantifies how the occurrence of outages varies with weather conditions using an **interpretable descriptive framework** on two complementary scales: a **monthly macro view** (for screening with seasonal awareness and context) and a **site-day view** (for non-linearities, regime effects and compound amplification).

4.3.1. Explainability Audit

Before interpreting correlations or risk estimates as evidence, the work performs an explainability audit to quantify how often Terna-labelled weather failures coincide with objective meteorological moderators at the same site-day.

For each site-day with a failure, the audit checks whether at least one relevant moderator is active on that same day (exposure bins, compound/transition flags, and lightning moderators only when lightning data are available).

Sensitivity is evaluated with two same-day severity settings:

- **BROAD:** moderate to severe regimes plus compound/transition flags.
- **STRICT:** only the most severe regimes (keeping key compound indicators unchanged).

To enable accounting and stacked compositions, each site-day with a failure is assigned to exactly one mutually exclusive class following a hierarchical order:

1. **Explained** (same-day)
2. **Explained accumulated-only** (no same-day trigger but exceptionally high accumulated/memory metrics)
3. Residual:

1. **Unexplained (lightning available)** vs
2. **Undetermined (missing lightning coverage)**

Antecedent load thresholds are defined by upper tail quantiles computed on positive values, with sensitivity analysis on alternative cutoffs.

4.3.2. Monthly Analysis

The dataset is aggregated by (year, month) to construct a macro picture containing **exposure terms** (site-days, active sites), **failure burden**, and a **normalized failure rate** (failures per site-day). The monthly meteorological context is summarized by a compact set of descriptors derived from engineered features: **monthly means** (background regimes), **monthly maxima** (episodic peaks when relevant), and **regime shares** (fraction of site-days in selected categories or scenario flags).

Associations between monthly failures and candidate drivers are screened using **Pearson** (linear) and **Spearman** (rank/monotonic) correlations in two variants:

- **Raw monthly series** (to capture overall co-variation)
- **Deseasonalized anomalies** (obtained by subtracting the long-term mean for each calendar month from both the failure target and the driver series to reduce shared seasonality)

The monthly analysis also supports a simple anomaly extraction. Months can be ranked by standardized failure burden (with an **exposure guardrail** to avoid artefacts due to small denominators) and reported together with concise meteorological context. This macro layer acts as a **seasonality-aware screening step** that informs which drivers and compound scenarios deserve deeper site-day and spatial inspection.

4.3.3. Site-Day Analysis

At site-day resolution, the analysis estimates how failure probability changes between **exposure regimes** and **multi-hazard combinations**. For each categorical driver and flag, failure rates are computed as $\hat{p} = k/n$ where n is the number of site-days in the category and k is the number with *has_weather_failure*=1. Uncertainty is reported using **95% Wilson CIs**, stable in rare events and with uneven support.

Compound amplification is assessed using **two-way interaction tables** (and some selected three-way tables) that report support, number of failures, and rates, applying filtering by minimum support.

Effect sizes are summarized through **RR** with respect to a baseline category (typically none/calm/normal/dry), with CIs computed on the **log scale** and exponentiated back:

$$RR = \frac{a/(a+b)}{c/(c+d)} \tag{4.15}$$

where (a) and (b) are failures and non-failures within the exposed category and (c) and (d) are failures and non-failures within the baseline category.

Core summaries (rates, interactions, RR) are repeated stratified by **altitude bands**, **season**, and **province** to test robustness and avoid mixing distinct regimes. As an additional control for site-specific heterogeneity and seasonal context, a matched **case-crossover** comparison is performed within **site-month** strata (failure days vs non-failure days from the same site and month), reducing confounding by fixed site characteristics and monthly background. For selected continuous metrics, the analysis reports **additive contrasts** (fail - no-fail) and **multiplicative contrasts** (fail/no-fail, optionally in log space), retaining only strata with sufficient informative comparisons.

4.3.4. Spatial Structure Tests

Maps communicate where failures occur, but they do not quantify whether failures are clustered beyond the spatial distribution of user sites. Thus, spatial diagnostics are applied to the set of failure point (site coordinates replicated for failure days), under two scopes: **all months** and **anomalous months.**

Diagnostics include:

- Terrain and coarse geographic signatures (**altitude bands**, **province composition**)
- **DBSCAN** clustering with Haversine distance (minimum samples fixed, eps tested at 10/15/20 km)
- **NNI** as a complementary summary against Complete Spatial Randomness (CSR)

These diagnostics provide quantitative evidence of spatial concentration and its sensitivity to scale.

4.4. Operational Risk Score: FPI

The preceding steps produce interpretable evidence through many bins and compound conditions. For operational use, the work synthesizes this evidence into a single site-day probability score that supports **ranking**, **monitoring**, and **explainable inspection**.

The FPI is formulated in **odds space**, so that multiplicative effects become additive in **log-odds** [56]. Each site-day belongs to a contextual stratum g capturing background heterogeneity:

$$g = (province, season)$$

A contextual baseline probability $p_0(g)$ is estimated using a **low-stress reference exposure** within each stratum (benign categories in the main hazards, with compound/transition flags inactive). The baseline is stored as:

$$\log\ odds_0(g) = \log\left(\frac{p_0(g)}{1 - p_0(g)}\right) \tag{4.16}$$

becoming the starting point for all site-days in the same stratum.

On top of this baseline, the score combines three families of components estimated within stratum g:

- Main effects as stratum-specific **odds ratios** for each driver category [57]
- **Uplifts** due to pair interactions that capture additional risk in joint configurations beyond what main effects predict (avoiding double counting) [58]
- A limited set of residual **compound/transition flags** as additional adjustments

Rare-event instability is addressed through two mechanisms:

1. **Haldane-Anscombe correction** [59]: a small sample correction to avoid infinite or undefined ratios.
2. **Shrinkage in log-odds space**: each log-effect is **shrunk toward zero** as support decreases, with stronger shrinkage for sparser components (interactions and flags).

When parameters are weak or unavailable at (*province,season*) level, a **hierarchical fallback** is applied (province-season → season-only → pooled), and unseen categories default to **neutral** contributions. This guarantees full site-day coverage while preserving context whenever the data supports it.

4.4.1. Calibration and walk-forward validation

Because multiplicative stacking can be poorly calibrated at the upper tail, calibration is applied as a monotonic post-processing step using **Platt scaling** in logit space [60], fitted on each training window:

$$\hat{p}_{\mathrm{cal}} = \sigma\left(b + a\,\mathrm{logit}(\hat{p}_{\mathrm{raw}})\right) \tag{4.17}$$

To efficiently adjust calibration under extreme imbalance, training uses all events and a reproducible hashed subsample of non-events, with **weights** to reflect true prevalence.

Validation uses a **walk-forward backtest** aligned with deployment. For each test year Y, all parameters (baseline rates, odds ratios, uplifts, shrinkage/fallback lookups, and calibration mapping) are estimated using only data up to $Y-1$, then applied to score all site-days in year Y. Performance is evaluated along three axes:

- Discrimination: Area Under the Receiver Operating Characteristic Curve (**AUC-ROC**) and Area Under the Precision-Recall Curve (**PR-AUC**) / Average Precision (**AvgPrec**), with PR-AUC interpreted relative to the yearly event rate.
- Calibration: **reliability curves** and **Brier score**, plus a **skill score** relative to a contextual baseline using only province-season reference rates.
- Operational ranking value: **capture rate** and **lift** at fixed monitoring budgets (top 1%, 5%, 10% of site-days), quantifying how concentrated failures are in the top-ranked subset.

5 Results

After building the modelling-ready **site-day dataset** and reconstructing site-level meteorological exposure, this chapter presents the main empirical results of the work. The results are organized in four layers:

1) A **descriptive overview** of coverage and failure statistics (Section 5.1)
2) An **explainability audit** used as a validation lens for the reconstructed weather attribution at site level (Section 5.2)
3) Two complementary evidence scales: **monthly macro-patterns** (Sections 5.3-5.4) and **site-day evidence** based on bins, interactions, and relative-risk estimates (Section 5.5)
4) An out-of-sample **walk-forward backtest** of the FPI, assessing discrimination, calibration, and operational usefulness under realistic monitoring budgets (Section 5.6)

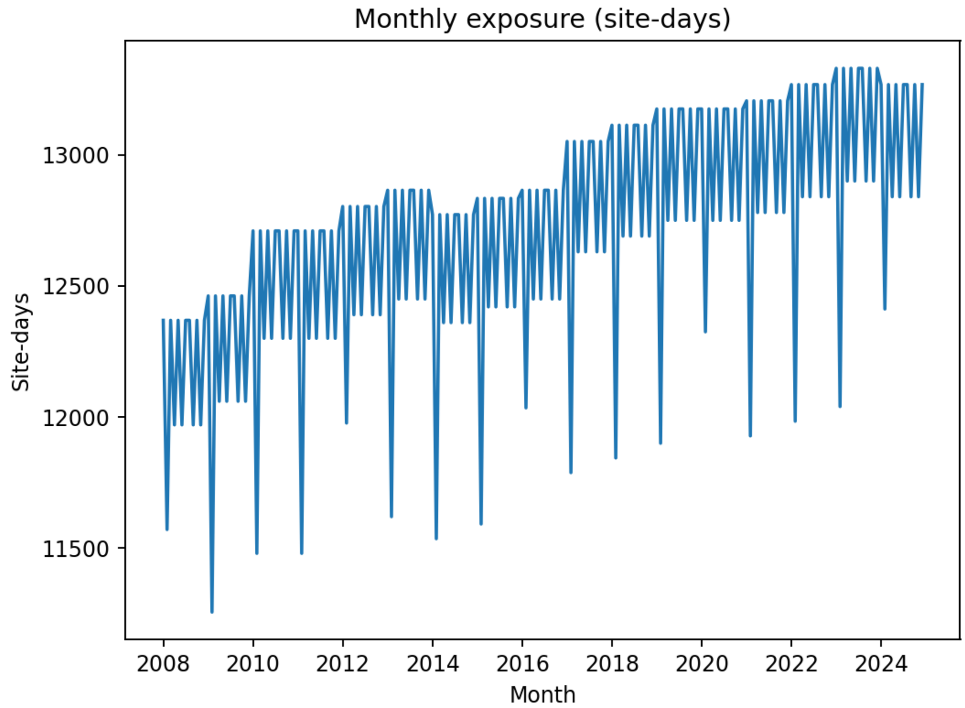
5.1. Descriptive Overview

The analysis focuses on Piedmont over the 2008-2024 window, using a daily site-day panel of Terna user sites, enriched with reconstructed meteorological covariates and derived categorical moderators.

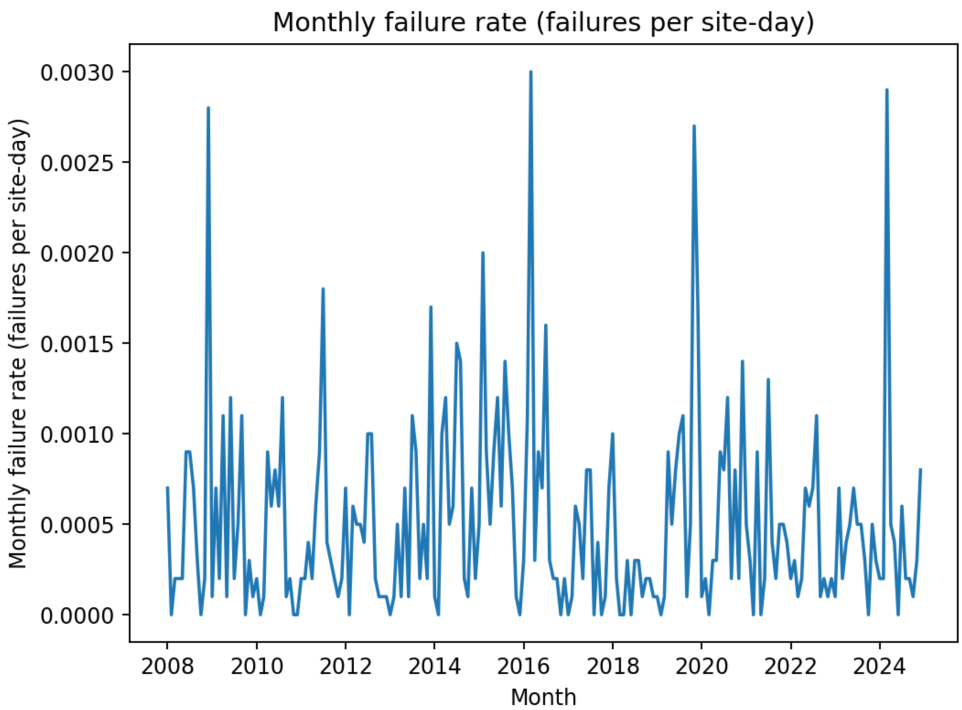
Table 5.1 Dataset Summary

Temporal coverage	Exposure site-days	Failure site-days	Weighted failures
2008-2024	**2,592,120**	**1,309**	**1,573**

At monthly scale, exposure is not perfectly constant due to data availability and quality filters. Monthly exposure spans approximately **11,256 to 13,330 site-days**, with a mean of **~12,706 site-days per month** (Figure 5.1.a). Accordingly, monthly failure burden is reported primarily as a normalized failure rate (failures per site-day; Section 4.3.2), which is strongly variable and heavy-tailed, with median 0.0003 and mean 0.0005 failures per site-day; the 90th percentile reaches 0.0011 and the maximum month reaches 0.0030 (Figure 5.1.b).



(a) Monthly Exposure Over Time (Site-Days)



(b) Monthly Failure Rate Over Time (Failures per Site-Day)

Figure 5.1 Temporal Coverage of Sites and Failures

The spatial unit of analysis is the Terna user site, geocoded within the Piedmont region and linked to province and altitude attributes. In the Terna registry, Piedmont contains **465 user sites**, making it one of the densest regions in terms of monitored grid nodes.

Failures are not uniformly distributed across the region. Later sections show that both seasonal regimes and anomalous months display clear geographic signatures, including (i) **mountain vs lowland contrasts** (altitude stratification) and (ii) **provincial concentration patterns**, which become even stronger when restricting to the most anomalous months (Sections 5.4-5.5).

5.1.1. Reconstruction Validation

To quantify the quality of the site-day meteorological reconstruction, the main interpolation methods are compared using **MAE** and **RMSE** against daily station observations (via a station-to-nearest-site pairing), using mean temperature, mean wind speed, and fresh snow as reference variables. Table 5.2 summarizes the results by variable and method.

Table 5.2 Interpolation Accuracy by Variable and Method

Interpolation Method	Temperature (ºC)	Precipitation (mm)	Snow (cm)	Wind (m/s)
IDW (MAE/RMSE)	1.19 / 2.14	0.72 / 2.87	1.16 / 3.75	0.41 / 0.81
RK (MAE/RMSE)	1.32 / 2.09	-	-	-
Spline (MAE/RMSE)	-	0.44 / 1.84	0.81 / 2.11	-

For **precipitation**, the **Spline** approach substantially outperforms **IDW** (0.44/1.84 vs 0.72/2.87), indicating lower average error and a markedly reduced error tail (lower RMSE). For **temperature**, **RK** slightly improves RMSE relative to **IDW** (2.09 vs 2.14) but shows a higher MAE (1.32 vs 1.19), suggesting more stable behaviour for larger deviations while not necessarily reducing typical absolute error. For **snow**, **Spline** performs much better than **IDW** (0.81/2.11 vs 1.16/3.75), consistent with improved spatial generalization in complex terrain. For **wind**, the **IDW** baseline is retained; its values (0.41/0.81) provide a reference level of reconstruction accuracy.

5.2. Explainability Audit

This section reports the results of the explainability audit defined in Section 4.3.1, which tests whether Terna-labelled weather failures coincide with at least one reconstructed adverse weather moderator at the same site-day (and, secondarily, with exceptional antecedent loading).

5.2.1. Overall Explained Share and Stability Over Time

Table 5.3 summarizes the pooled explainability results over 2008-2024.

Table 5.3 Overall Explainability Summary

Explained (BROAD)	Explained (Accumulated only)	Explained (STRICT)	Unexplained (lightning available)	Undetermined (missing lightning)
94.58%	0.00%	80.37%	4.66%	0.76%

Under the BROAD definition, 94.58% of failure site-days are explained, while under the STRICT definition the explained share is 80.37%. Only 4.66% remain unexplained when lightning data are available; an additional 0.76% are classified as undetermined due to missing lightning information.

The "Explained (accumulated only)" category is 0.00% in this run. This category is evaluated only for failure site-days not already explained by any BROAD same-day or compound moderator. The zero share therefore indicates that exceptionally high antecedent loading did not add additional explained cases beyond the same-day/compound set; either because high-load conditions typically co-occur with same-day moderators (especially rainfall/wetness), or because the remaining residual cases do not exceed the selected upper-tail thresholds. Moreover, the upper-bound BROAD explainability is ~95.34% once missing-lightning uncertainty is accounted for.

A key result is that explainability is not only high but also **stable** (Figure 5.2), meaning that the reconstruction does not rely on a single year or a single hazard dominating the explanation.

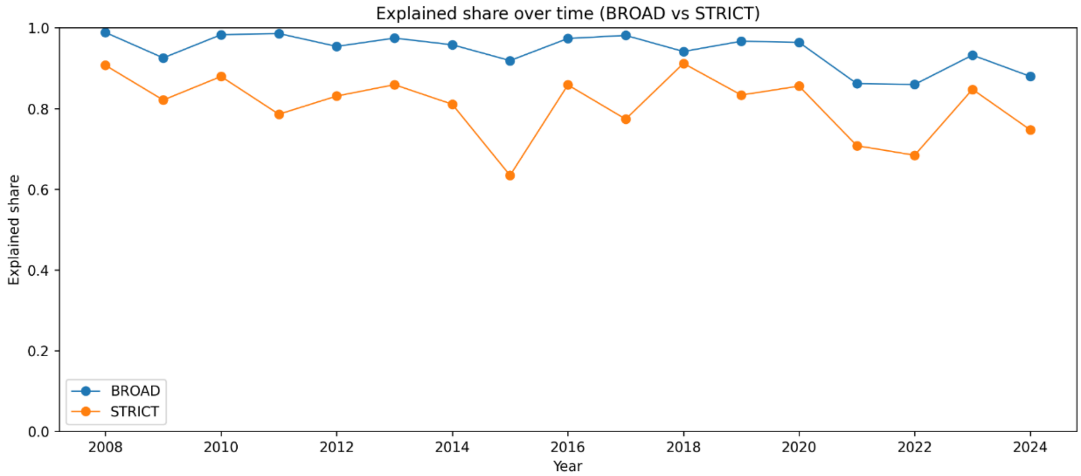


Figure 5.2 Explained Share Over Time

There are some notable deviations, such as 2021-22, which show a lower BROAD explained share (~86%) and a larger unexplained fraction. 2015 stands out for a very large BROAD-STRICT gap (91.9% vs 63.4%). This indicates that many Terna-labelled weather-related failures occurred under moderate-to-strong conditions and/or compound mechanisms, rather than under the most extreme daily bins used by the STRICT definition. In other words, the audit suggests a vulnerability-driven regime where failures can be triggered without requiring the most severe rainfall or wind categories. The only structural deviation is 2024, where missing lightning data creates a non-zero "undetermined" share by construction; nevertheless, even without lightning, **88%** of 2024 failures are still BROAD-explained by non-lightning moderators (Figure 5.3).

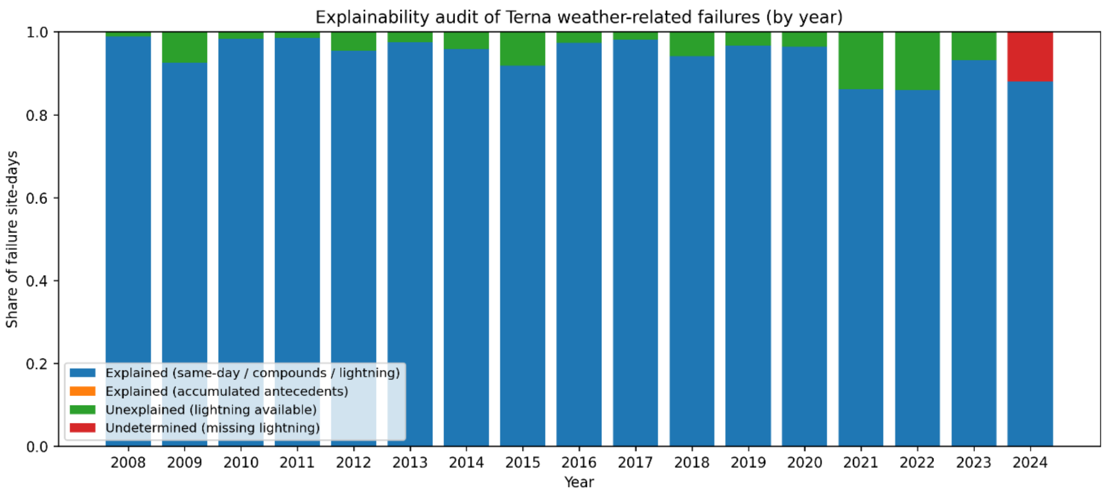


Figure 5.3 Explainability by Year

5.2.2. Explained Share by Season / Altitude / Province

By season, explained share is highest in summer (BROAD 98.8%, STRICT 83.4%) and lowest in autumn (BROAD 89.2%, STRICT 73.7%). In particular, summer shows very low unexplained share, while autumn concentrates the largest fraction of failures without any BROAD moderator, suggesting either harder-to-capture mechanisms (short-lived localized events, vegetation interactions, etc.) or potential label noise (Table 5.4).

Table 5.4 Explainability by Season

Season	N (failure site-days)	Explained (BROAD)	Explained (STRICT)	Unexplained (lgt available)	Undetermined (missing lgt)
Winter	273	91.9%	82.4%	6.2%	1.8%
Spring	347	93.7%	78.1%	5.2%	1.2%
Summer	495	98.8%	83.4%	1.0%	0.2%
Autumn	194	89.2%	73.7%	10.8%	0.0%

By altitude, explainability also varies with terrain context. Low-altitude sites (0-200 m) have lower STRICT explainability (64.3%) and a higher unexplained share (12.2%), whereas the highest-altitude band (1000+ m) is almost fully explained under BROAD (99.4%) and remains high under STRICT (89.4%) (Table 5.5). This is consistent with the idea that mountain-related weather stressors (snowpack, persistent wetness, ROS) are captured more robustly by daily indicators than some lowland mechanisms.

Table 5.5 Explainability by Altitude Band

Altitude band (m)	N (failure site-days)	Explained (BROAD)	Explained (STRICT)	Unexplained (lgt available)	Undetermined (missing lgt)
0-200	196	85.2%	64.3%	12.2%	2.6%
200-500	651	94.5%	79.7%	4.8%	0.8%
500-1000	283	98.2%	87.3%	1.8%	0.0%
1000+	179	99.4%	89.4%	0.6%	0.0%

By province, explained shares are high in the provinces with the largest number of failure site-days (e.g., Cuneo 97.9% explained total; Torino 97.0%), while provinces with smaller counts can show higher unexplained shares (e.g., Asti 18.8%, Alessandria 11.8%) (Table 5.6).

Table 5.6 Explainability by Province

Province	N (failure site-days)	Explained total	Unexplained (lgt avail)	Undetermined (missing lgt)
CUNEO	426	97.9%	1.9%	0.2%
TORINO	299	97.0%	2.3%	0.7%
VERBANO-CUSIO-OSSOLA	213	95.3%	4.7%	0.0%
VERCELLI	162	87.7%	10.5%	1.9%
NOVARA	83	95.2%	4.8%	0.0%
ALESSANDRIA	68	83.8%	11.8%	4.4%
ASTI	32	78.1%	18.8%	3.1%
BIELLA	26	96.2%	3.8%	0.0%

5.2.3. Moderators and Combinations

Table 5.7 reports prevalence of moderators among failure site-days. Rain (moderate+) is the most frequent single hazard (71.1% of failure site-days), followed by wet soil (65.9%) and lightning (57.1%). Composite regimes are also common (e.g., wet & windy 37.1%, wet soil & wind 26.1%, ROS 18.9%), supporting a multi-hazard interpretation.

Table 5.7 Moderator Occurrence within Terna-Labelled Failures

Moderator	N	Share of failure site-days
Rain (moderate+)	931	71.1%
Wet soil (wet+)	863	65.9%
Lightning	747	57.1%
Wet & windy	486	37.1%
Snow (moderate+)	403	30.8%
Accumulated (antecedent load)	383	29.3%
Wind (strong+)	375	28.6%
Cold & wet	363	27.7%
Wet soil & wind	341	26.1%
ROS	247	18.9%
Wind ramp	110	8.4%
Dry → wet transition	98	7.5%
Freeze/thaw	67	5.1%

Figure 5.4 summarizes this multi-cause nature by showing the distribution of the number of most common active moderators per failure day (rain, wind, snow, wet soil, freeze/thaw, lightning, accumulated). The distribution is clearly shifted towards **multi-driver cases**. Only 6.0% of failure site-days show 0 active moderators, meaning that in those cases none of the main conditions is detected (these days either fall into the "unexplained/undetermined" categories, or they are explained by "secondary" composite flags that are not part of the main ones). Single-driver days exist but are not dominant (13.8%). In contrast, the majority of events occur with multiple concurrent moderators: 2 moderators (16.3%), 3 moderators (26.8%), and 4 moderators (25.8%). Overall, more than half of the failures (~52.6%) happen with 3-4 active moderators, which supports the interpretation that weather-related outages are frequently associated with compound meteorological conditions.

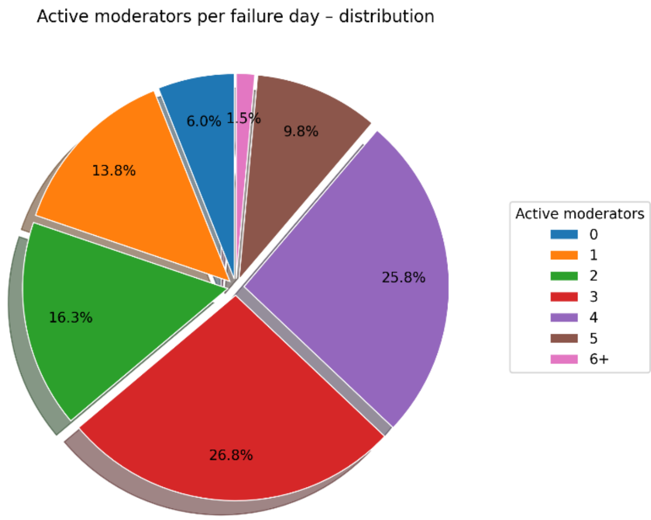


Figure 5.4 Active Moderators per Failure Day.

Multi-moderator days are dominated by a small set of recurring combinations. The most common is Rain + Wet soil + Lightning (N=131), followed by Lightning alone (N=90) and Rain + Wind + Wet soil + Lightning (N=87). Snow-related combinations (Rain + Snow + Wet soil; and variants with accumulated antecedent load) are also prominent (Table 5.8).

Table 5.8 Multi-Moderator Combinations (Top-15)

Main combination	N
Rain + Wet soil + Lightning	131
Lightning	90
Rain + Wind + Wet soil + Lightning	87
Other composites	79
Rain + Snow + Wet soil	77
Rain + Snow + Wet soil + Accumulated	75
Rain + Wet soil + Lightning + Accumulated	66
Rain + Snow + Wet soil + Lightning + Accumulated	57
Rain + Lightning	48
Rain + Wind + Lightning	47
Rain + Wind + Wet soil + Lightning + Accumulated	41
Wet soil + Lightning	40
Rain + Snow + Wet soil + Lightning	38
Rain + Wet soil	36
Rain + Wet soil + Accumulated	30

5.3. Monthly Patterns

5.3.1. Seasonality and Anomalies

After validating the site-level reconstruction via the explainability audit, the analysis shifts to the monthly scale to identify macro patterns. Monthly failures show strong seasonality and episodic spikes (Figure 5.1b); rather than being uniformly distributed, weather-related failures concentrate in specific months and periods, motivating the definition of **anomalous months** and regime archetypes. To quantify anomaly intensity, the distribution of monthly failure rate is complemented by a Top-20 list of anomalous months (Table 5.9):

Table 5.9 Top-20 Anomalous Months

Year	Month	Monthly failure rate (failures per site-day)	Failure count	Monthly exposure (site-days)	Mean 7-day cumulative rainfall	Mean AP index (k=0.85)	Share of ROS days	Mean snow-on-ground (7-day maximum)	Share of saturated-soil + wind days	Share of strong-wind days	Mean lightning within 20 km (7-day sum)
2016	3	0.0030	63	12865	26.5474	25.4242	0.047	37.6206	0.165	0.1807	5.3452
2024	3	0.0029	59	13268	60.1109	57.1508	0.1336	36.4155	0.2728	0.2268	NA
2008	12	0.0028	96	12369	48.7584	44.5100	0.1851	63.3978	0.4219	0.2342	1.5379
2019	11	0.0027	46	12750	70.2142	60.4291	0.1932	20.3786	0.3597	0.2282	2.1422
2015	2	0.0020	36	11592	25.2840	22.4404	0.1613	56.1827	0	0	0.2058
2011	7	0.0018	26	12710	23.0444	21.4591	0	0.0265	0.2033	0.2238	382.8831
2013	12	0.0017	30	12865	24.7459	20.1525	0.0751	21.2270	0	0	0.0402
2019	12	0.0017	26	13175	25.7324	34.4729	0.0865	31.1150	0.3801	0.2248	0.0612
2016	7	0.0016	21	12865	13.7788	12.8907	0.0011	0.3911	0.1288	0.255	678.7197
2014	7	0.0015	19	12772	33.0371	30.1198	0.0016	0.1110	0	0	273.7787
2014	8	0.0014	19	12772	22.7313	23.8691	0	0.0089	0	0	244.0705
2015	8	0.0014	18	12834	28.2842	25.0238	0	0.0110	0	0	569.8220
2020	12	0.0014	20	13175	15.8619	14.5626	0.0803	25.2856	0.1262	0.1248	0.0528
2021	7	0.0013	17	13206	22.1677	20.5824	0.0008	0.1442	0.1914	0.2334	432.0753
2020	8	0.0012	16	13175	12.8902	12.1179	0	0.0063	0.1009	0.2347	501.2519
2014	4	0.0012	17	12360	13.5325	12.5442	0.0324	22.2227	0	0	14.4278
2015	6	0.0012	15	12420	19.5847	19.2955	0.0073	1.0613	0	0	607.6106
2010	8	0.0012	15	12710	27.5534	25.6724	0.0107	3.0583	0.1663	0.1947	147.5728
2009	6	0.0012	14	12060	16.9387	15.3127	0.0162	4.9611	0.1678	0.2842	152.0681
2013	7	0.0011	15	12865	13.5219	12.5520	0.0038	0.6973	0	0	251.7939

Table 5.8 is not only a ranking of high monthly failure rates, it also provides a first hypothesis on *why* these months were anomalous, by contextualizing the failure rate

with aggregated hazard indicators. The majority of anomalous months exhibit outstanding values in **wetness-memory metrics** (mean 7-day accumulated rainfall and/or AP index), often accompanied by **ROS** and **wet-soil + wind** conditions, consistent with saturation and snowpack-transition mechanisms. A second recurring subset of anomalous months is characterized by **exceptionally high lightning activity** (7-day lightning sums), sometimes combined with strong-wind conditions, which is consistent with a warm-season convective regime. Finally, a smaller set of months shows high failure rates without extreme values in region-wide monthly means. This is consistent either with spatially localized storm footprints affecting only a subset of sites and therefore being diluted in monthly averaging, or short-lived extremes (gust or lightning bursts) that are better captured at site-day resolution in the subsequent analyses. These two recurring signatures anticipate the two archetypes formalized in the next subsection (Profile A and Profile B).

5.3.2. Deseasonalized Correlations

To separate "seasonal co-movement" from more informative associations, monthly features are deseasonalized and correlated with the deseasonalized monthly failure rate. The key result is that, after removing seasonality, the strongest correlations are associated with **persistent wetness and antecedent load:** share of ROS site-days (Pearson 0.451), mean 7-day accumulated rainfall (0.445), and mean AP index (0.437). Snow and compound wet-soil + wind effects are secondary but still present.

Table 5.10 Top Deseasonalized Correlations

Variable	Pearson (deseasonalized)	Spearman (deseasonalized)
Share of site-days with ROS condition	0.451	0.246
Monthly mean of 7-day accumulated rainfall	0.445	0.347
Monthly mean of AP index (AP, k=0.85)	0.437	0.331
Monthly mean of daily rainfall (interpolated)	0.413	0.358
Share of days with moderate rainfall	0.357	0.339
Share of days with light rainfall	0.312	0.292
Monthly mean of max snow-on-ground (last 7 days)	0.285	0.214
Share of site-days with wet soil & wind condition	0.283	0.129
Share of days with moderate snow	0.277	0.165
Share of days with breezy wind (%)	0.231	0.229

This supports the interpretation that "wetness memory + snowpack transition processes" represent a robust monthly driver even beyond simple winter/summer seasonality.

5.3.3. Anomalous-Month Case Profile

Anomalous months are defined as the Top-20 months ranked by monthly failure rate (with minimum exposure guardrails). They span a wide range of monthly failure rates and, despite being only a small subset of all months, account for a disproportionate share of failures: **612 weighted failures**, i.e., **~38.9% of all weighted weather failures**.

A crucial finding is that anomalous months do not share a single meteorological signature. Instead, they tend to fall into two recurring meteorological archetypes (Figure 5.5):

- **Profile A - Wetness + Snowpack regime:** a cold/shoulder-season regime driven by wetness persistence and snowpack-related processes
- **Profile B - Convective/Lightning regime:** a warm-season convective regime dominated by lightning conditions.

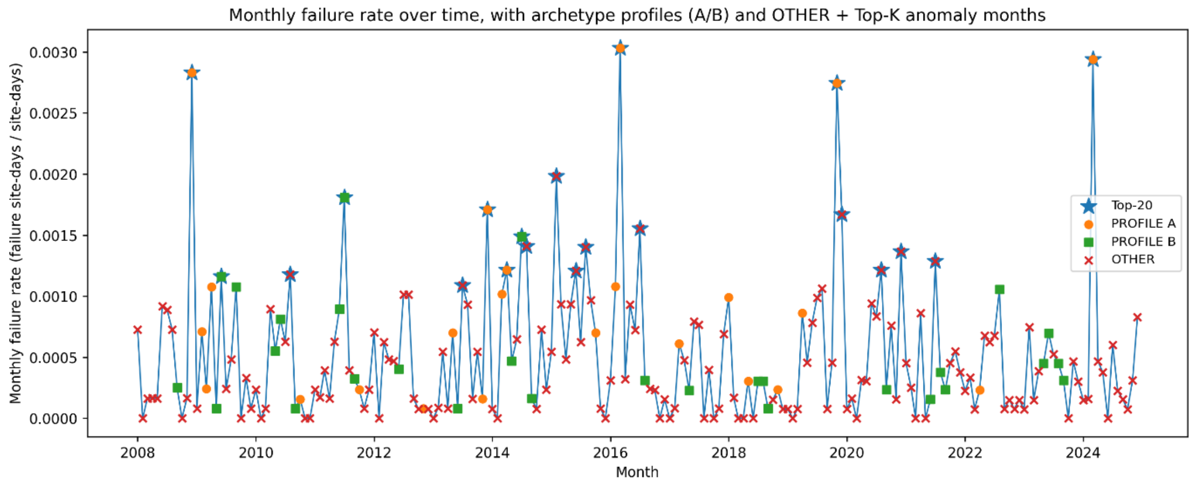


Figure 5.5 Monthly Failure Rate Over Time with Profiles

Although Profile A accounts for only 12.7% of months, it contributes 30.0% of weighted failures and has the highest mean monthly failure rate (0.001620). Profile B accounts for 16.0% of months and contributes 12.7% of weighted failures, with a mean monthly rate comparable to OTHER (Table 5.11).

Table 5.11 Month-Profile Counts and Weighted Failure Contribution by Profile

Profile	Months	Months share	Weighted failures (sum)	Failures share	Mean monthly failure rate	Mean exposure (site-days)
Profile A	23	12.7%	472.0	30.0%	0.001620	12634
Profile B	29	16.0%	199.0	12.7%	0.000542	12689
OTHER	129	71.3%	902.0	57.3%	0.000549	12742

Within Profile A, the signature is dominated by ROS processes (64.2% of weighted failures), followed by wet-soil + wind (13.2%) and wet & windy (8.2%). Within Profile B, lightning dominates (70.4%), with an additional contribution from wet-soil + wind (19.1%). These compositions provide a direct physical interpretation of the archetypes and motivate the compound analyses at site-day level.

5.4. Spatial Signature of Anomalous Months

5.4.1. Altitude and Province Concentration

The archetypes imply an explicitly spatial hypothesis, namely that if they reflect different drivers, their failures should differ geographically.

Altitude stratification strongly supports this expectation (Figure 5.6). Across all months, Profile A is predominantly high-altitude: **~75.4% of weighted failures occur above 500 m**, rising to **~82.6%** when restricting to Top-20 anomalous months. In contrast, Profile B is predominantly lowland: **~83.9% below 500 m** across all months, and **~88.9% below 500 m** in Top-20 months.

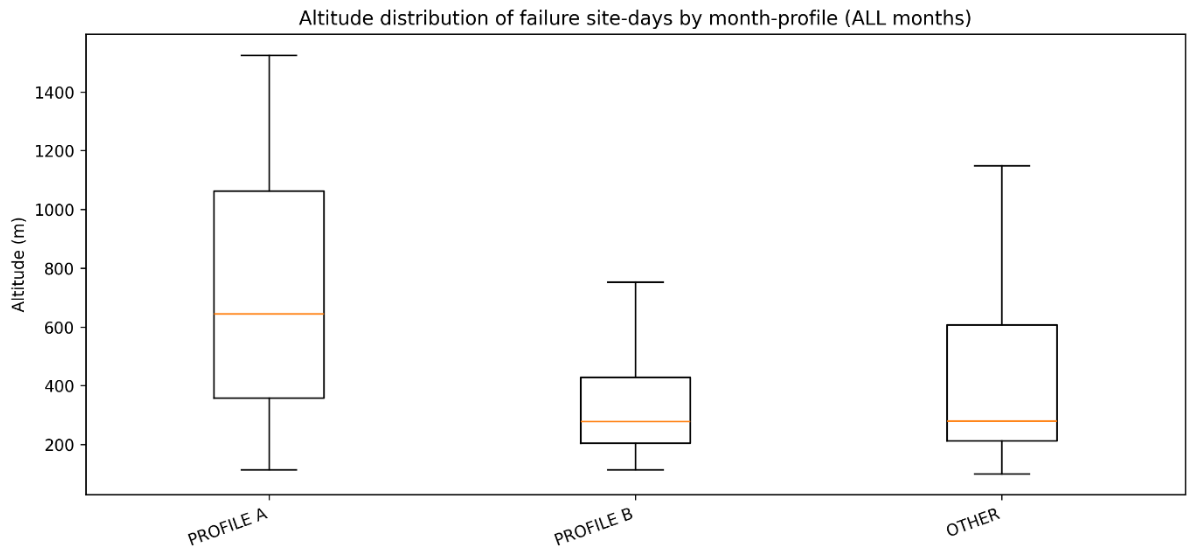


Figure 5.6 Altitude-Band Distribution by Profile

Province aggregation provides a complementary administrative-scale validation. Across all months, Profile A concentrates in **Cuneo (~47.6%)** and **Verbano-Cusio-Ossola (~14.9%)**, consistent with Alpine / pre-Alpine exposure, while Profile B concentrates in lowland corridors dominated by **Torino (~38.2%)** and **Vercelli (~21.0%)**.

5.4.2. Clustering Evidence (DBSCAN/NNI)

Spatial point-pattern tests confirm that anomalous-month failures are not spatially diffuse. Using DBSCAN and NNI, the results are these ones:

- **ALL months:**
  - Profile A: 310 points, 8 clusters, noise share 0.19, largest cluster share 0.27, **NNI = 0.197**
  - Profile B: 171 points, 2 clusters, noise share 0.16, largest cluster share 0.74, **NNI = 0.305**
- **Top-20 months:**
  - Profile A: 170 points, 3 clusters, noise share 0.33, largest cluster share 0.42, **NNI = 0.264**
  - Profile B: 42 points, 1 cluster, noise share 0.69, largest cluster share 0.31, **NNI = 0.283** (interpret with caution due to sample size)

Profile A indicates strong clustering into multiple mountain sub-regions (Figure 5.7), while Profile B indicates clustering but with one dominant hotspot at this scale (Figure 5.8).

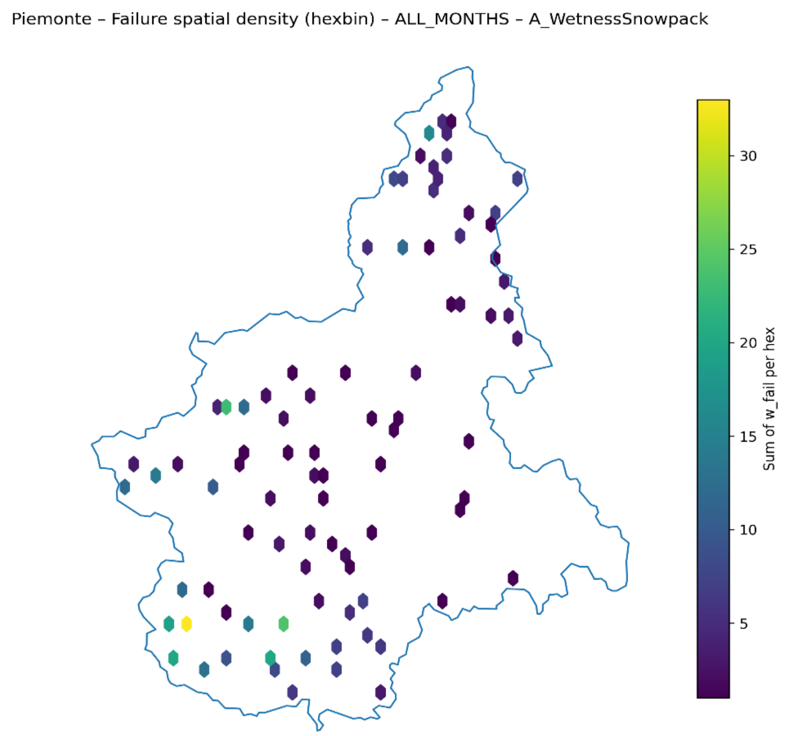


Figure 5.7 Failure Spatial Density Map for Profile A

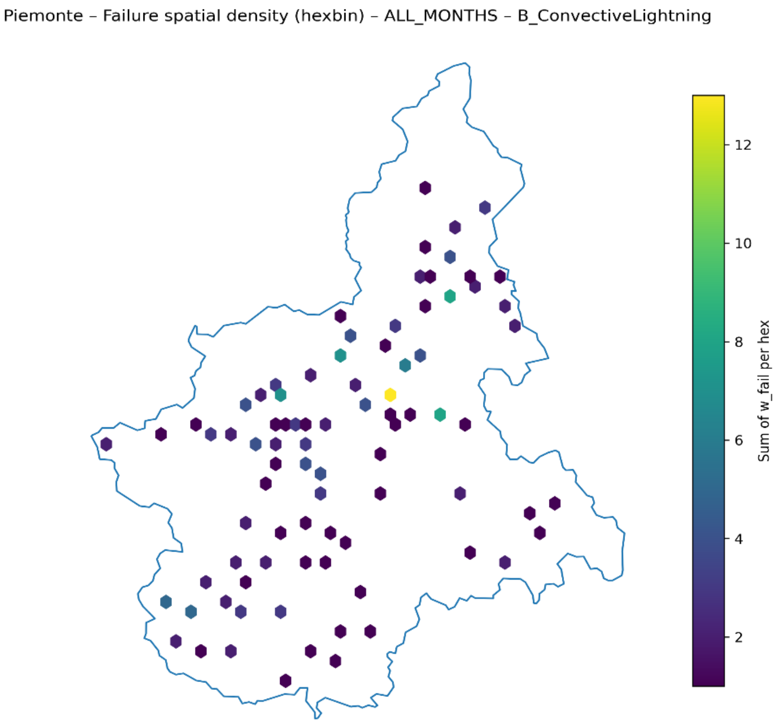


Figure 5.8 Failure Spatial Density Map for Profile B

Across eps = 10-20 km, sensitivity plots show that Profile A is **multi-nodal at local scales** (many clusters at 10 km) but becomes **regionally connected** as eps increases (cluster merging, noise collapse), consistent with failures concentrated along multiple mountain corridors. Profile B, instead, does not develop stable dense clusters under the same density criteria in Top-20 months, aligning with a more spatially spread lowland pattern rather than a single compact hotspot (Figure 5.9).

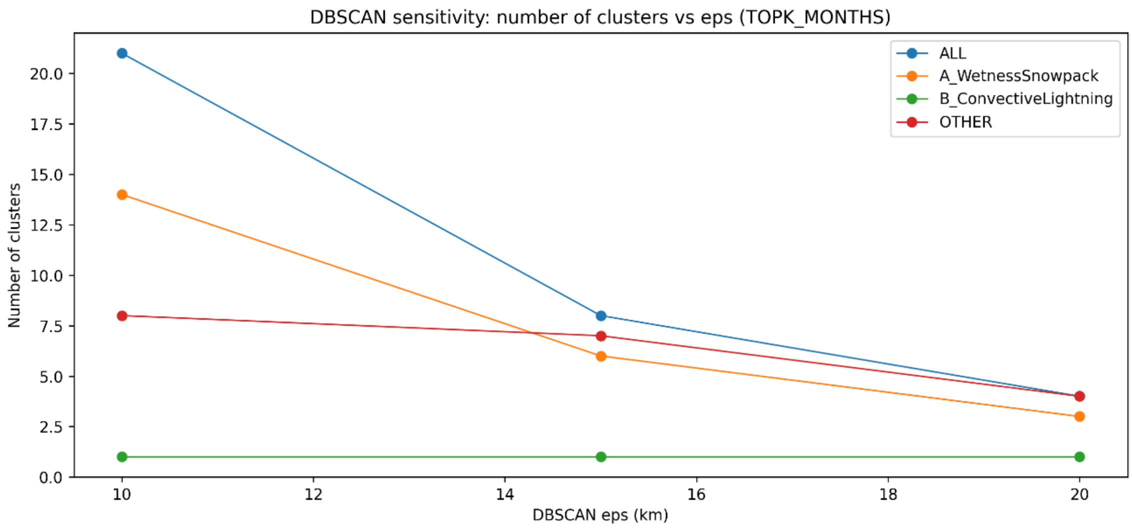


Figure 5.9 DBSCAN Sensitivity: Number of Clusters vs eps

5.5. Site-Day Results

Site-day bin, interaction and effect-size results are reported following the framework in Section 4.3.3. In particular, results are presented as: (i) bin-wise failure-rate profiles with 95% Wilson CIs; (ii) physically motivated two-way interaction summaries shown as uplift surfaces; and (iii) compact effect-size comparisons expressed as RR relative to a baseline bin.

5.5.1. Single-Variable Bin Results

Figures Figure 5.10-Figure 5.15 report single-variable bin profiles at site-day resolution. For interpretability across heterogeneous bins, the corresponding effect sizes are summarized as RR relative to a baseline category (e.g., no-rain, dry soil, calm wind). A ranked synthesis of the strongest regimes and stratified robustness checks is then provided in Section 5.5.3.

Rainfall intensity shows the strongest monotonic increase, as the RR relative to the no-rain baseline rises from 2.38 in light-rain days to 8.02 in moderate, 13.22 in heavy, and 29.94 in very heavy rainfall days (Figure 5.10).

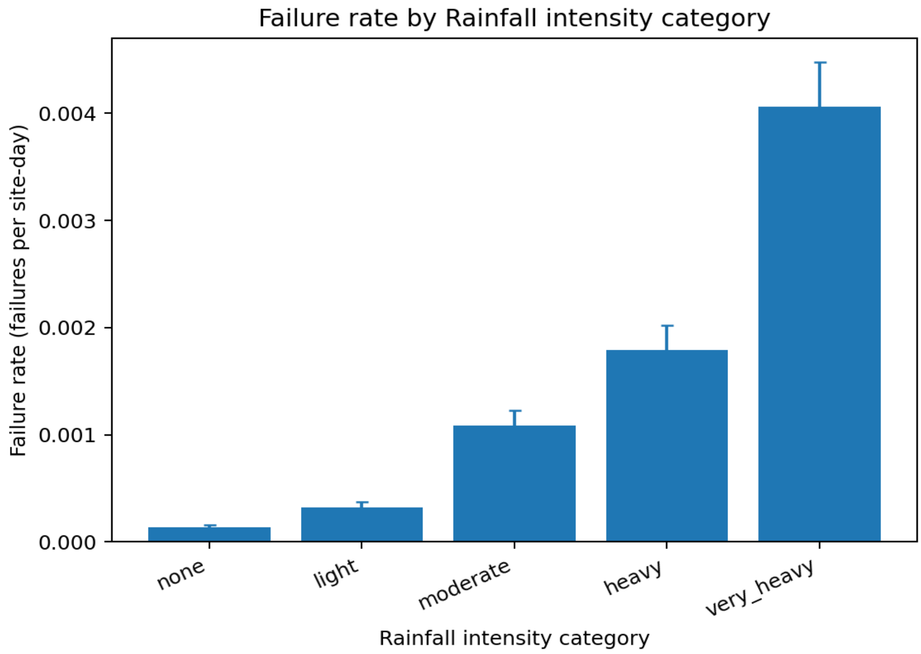


Figure 5.10 Failure Rate by Rain Bin

Soil wetness is a key persistence indicator. Compared to dry conditions, the RR increases to 1.91 (normal), 3.51 (wet), and 10.22 (very wet) (Figure 5.11).

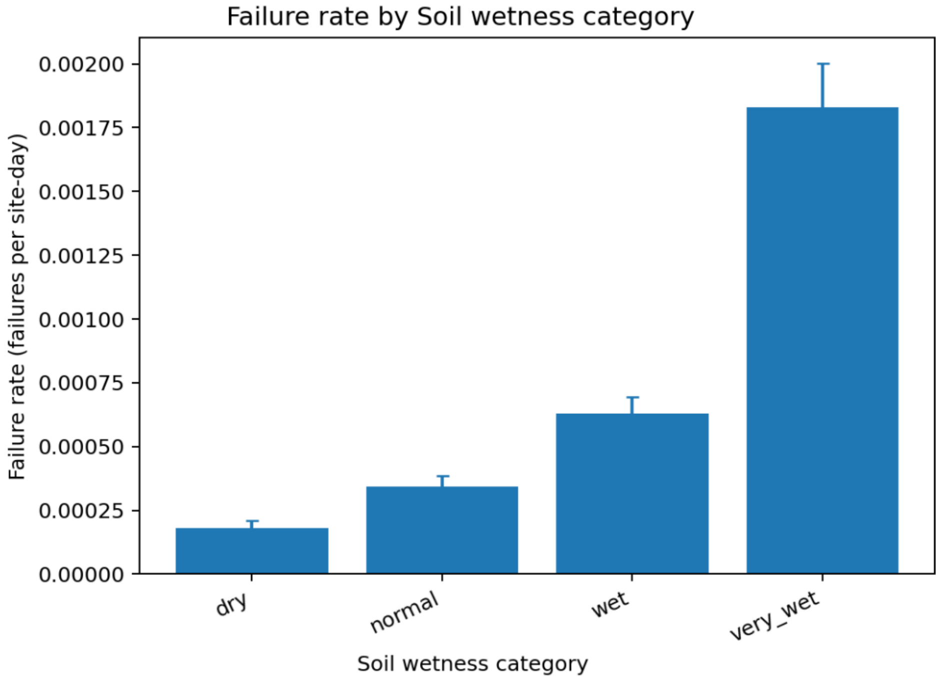


Figure 5.11 Failure Rate by Soil Wetness Bin

Wind intensity has a clearer but smaller effect than precipitation. RR reaches 4.61 in very strong wind days relative to calm (Figure 5.12).

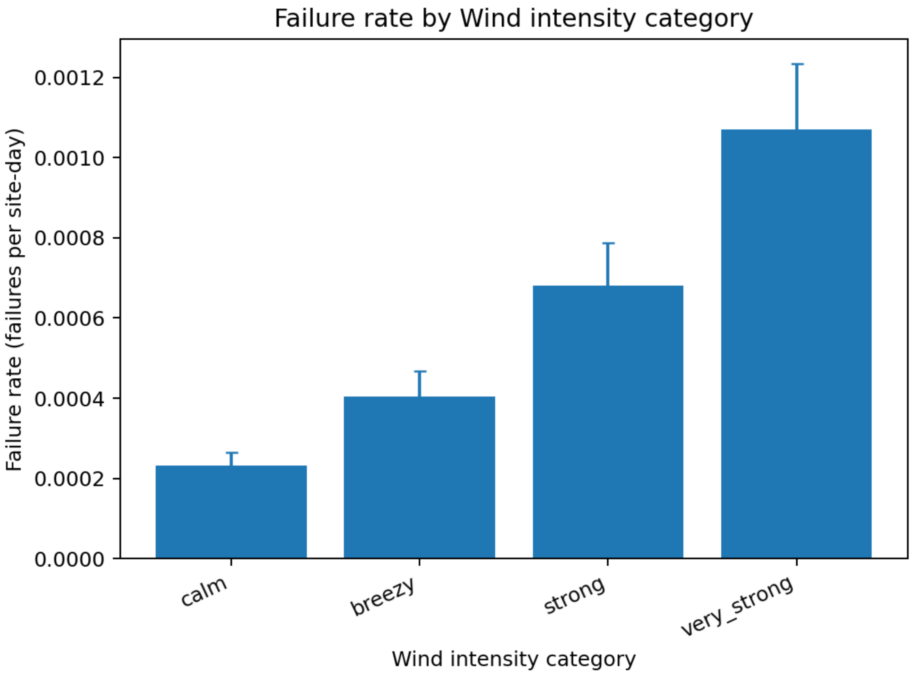


Figure 5.12 Failure Rate by Wind Bin

Temperature regimes highlight a critical near-zero band. RR is 3.15 for near-zero days and 2.30 for cold days, while heat does not show increased daily triggering risk (RR≈0.69) (Figure 5.13)

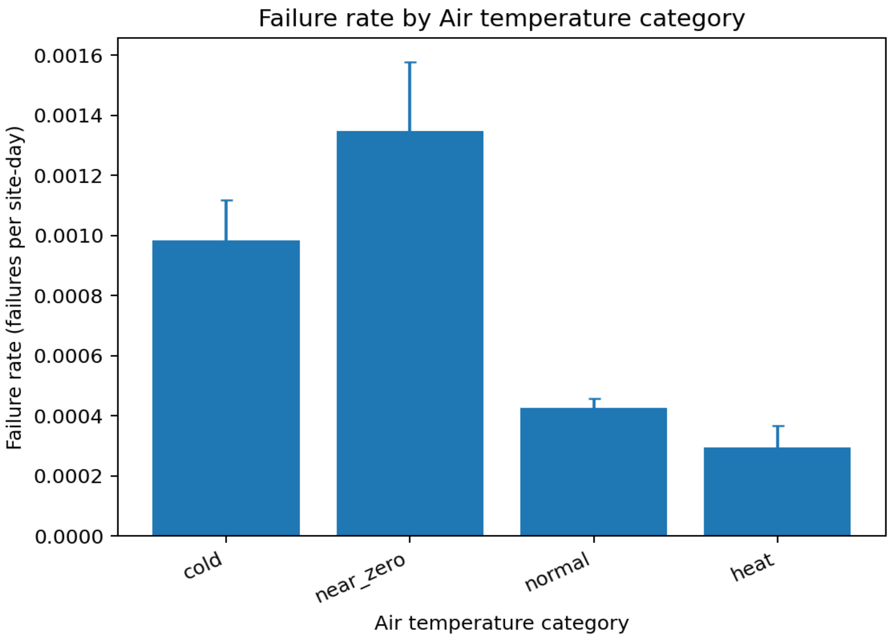


Figure 5.13 Failure Rate by Temperature Bin

Snow-only effects are weaker than rain and wetness but still relevant at extremes. The very heavy snow category reaches RR≈2.25 compared to no-snow day (Figure 5.14).

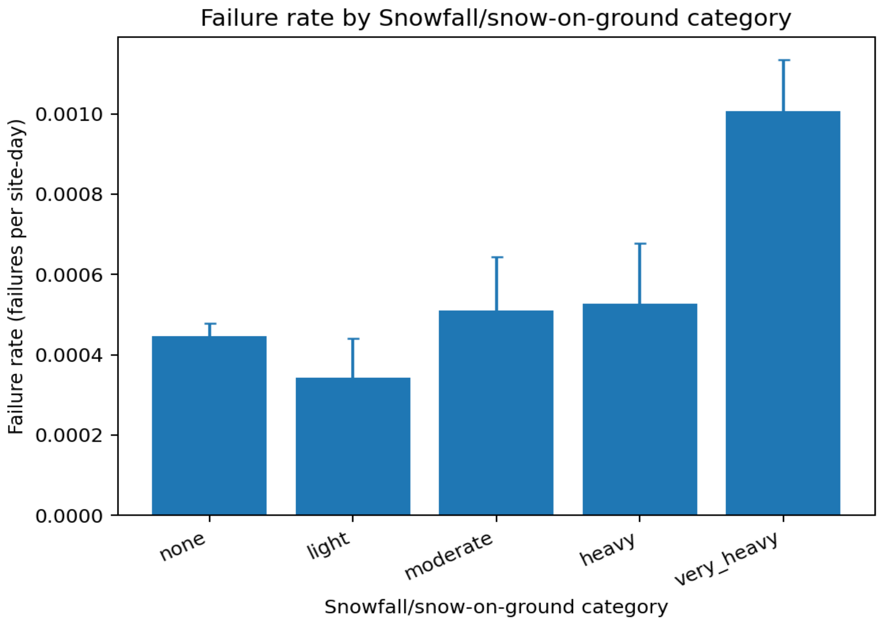


Figure 5.14 Failure Rate by Snow Bin

Lightning exposure is best communicated through proximity bins (minimum strike distance), which typically show a steep risk gradient at close distances (Figure 5.15). This complements the explainability audit, where lightning is among the most frequent moderators.

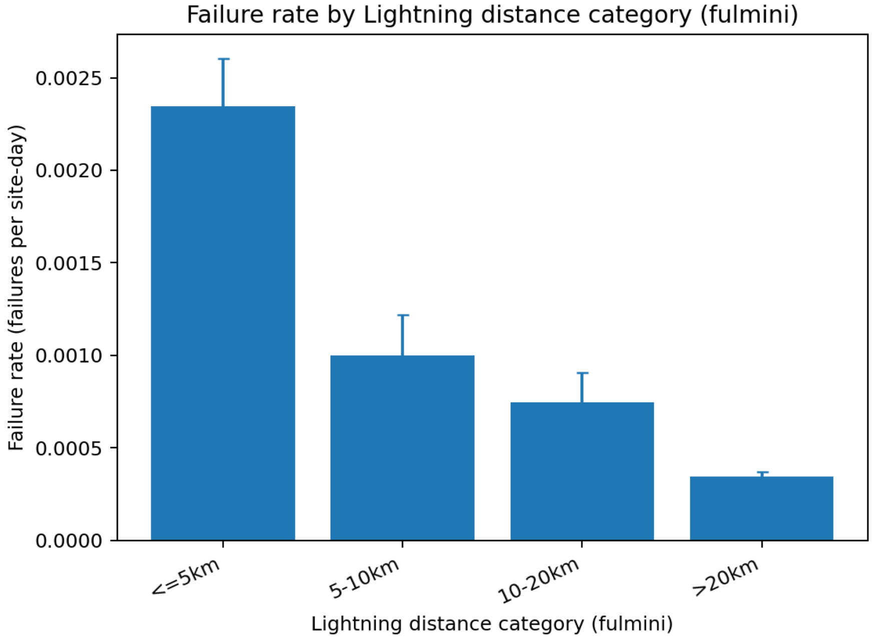


Figure 5.15 Failure Rate by Lightning Proximity

5.5.2. Compound Regimes

Single-variable bins cannot fully represent physical stress pathways when hazards interact. Therefore, the analysis includes explicit compound regimes, focusing on interaction families that are physically plausible and empirically relevant. These interaction results are presented as uplift surfaces (2D heatmaps) and compared against the baseline implied by marginal effects.

For example, the joint regime of very heavy rain and very strong wind forms the highest-risk corner in the **Rain × Wind** heatmap (Figure 5.16):

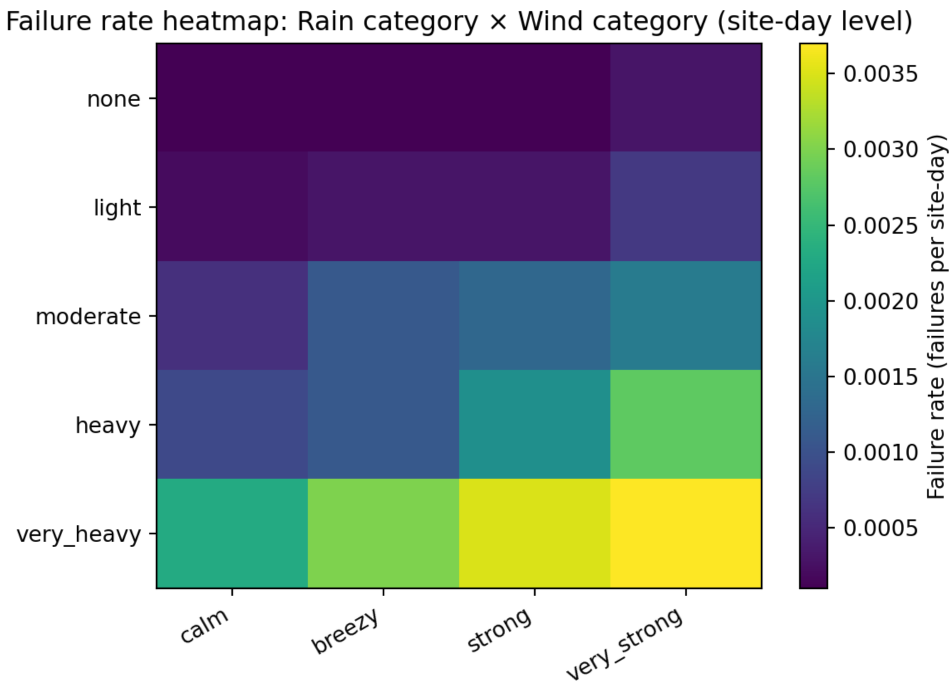


Figure 5.16 Interaction Heatmap: Rain × Wind

Temperature × Soil Wetness interactions emphasize the near-zero + wet corner, consistent with freeze-thaw, icing, and ROS processes (Figure 5.17).

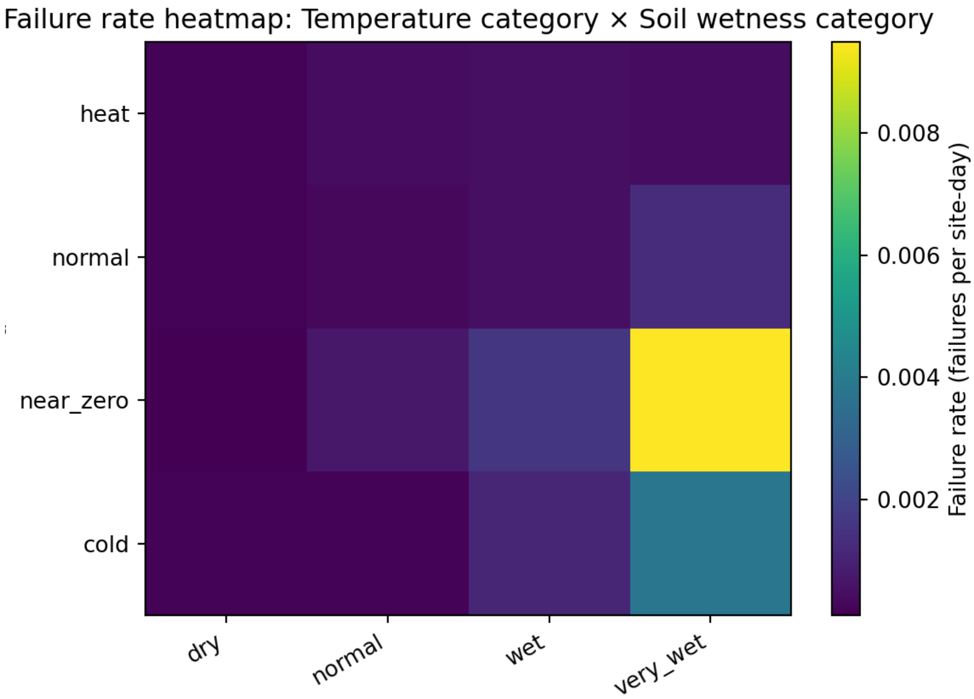


Figure 5.17 Interaction Heatmap: Temperature × Soil Wetness

Snow × freeze-thaw interactions further isolate the regimes where snow presence and freezing transitions co-occur (Figure 5.18).

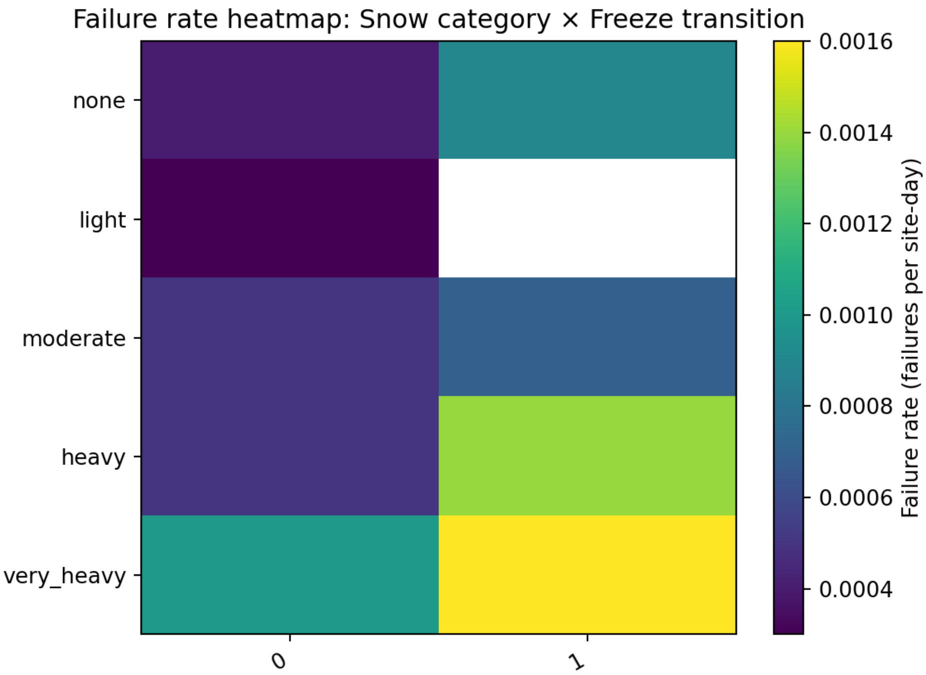


Figure 5.18 Interaction Heatmap: Snow × Freeze

Lightning interactions show that lightning-related risk is amplified under wet and/or windy conditions (Figure 5.19-Figure 5.20), consistent with the compound signals observed in the explainability audit and flag-based RR analysis.

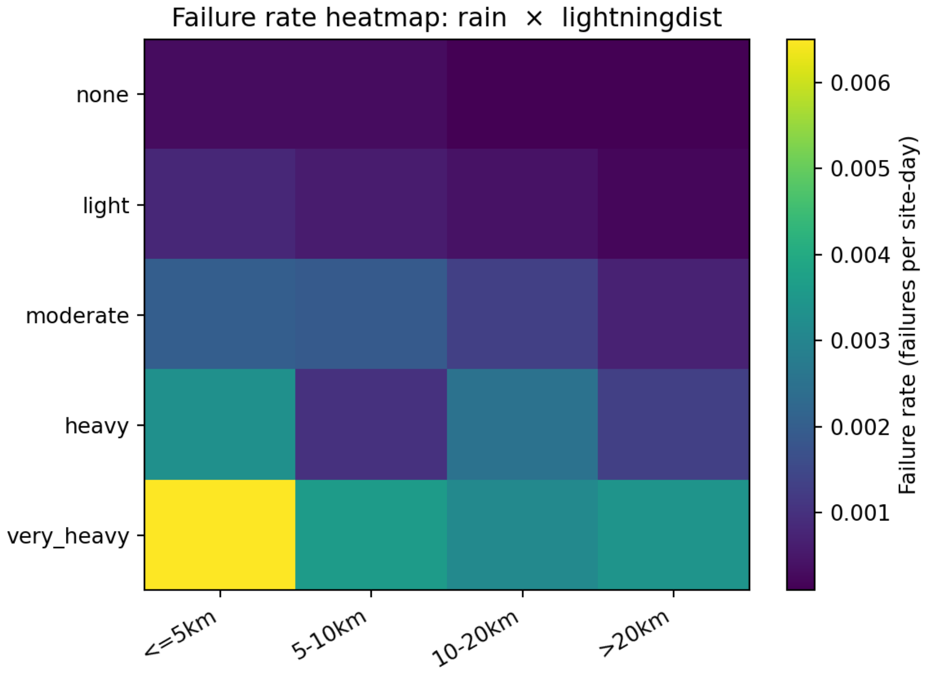


Figure 5.19 Interaction Heatmap: Rain × Lightning Proximity

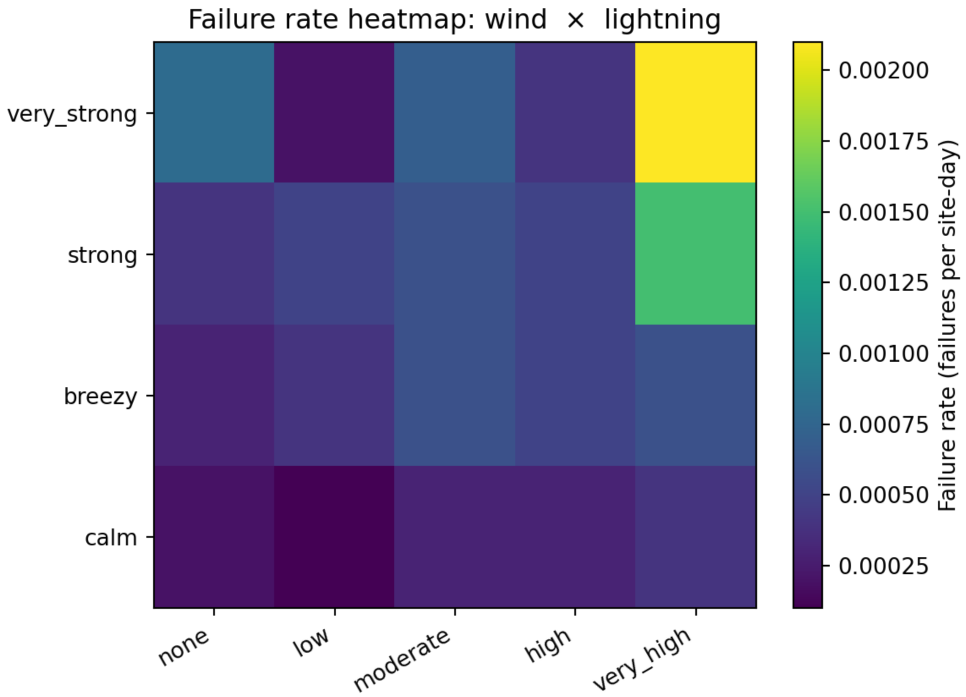


Figure 5.20 Interaction Heatmap: Wind × Lightning Intensity

5.5.3. Flag-based RR ranking and Stratified Results

For interpretability and operational comparability, the work reports a "flag-based" view of risk through RR with CIs (Figure 5.21). This allows a direct reading of which bins or regimes correspond to the largest multiplicative risk increase. To synthesize compound mechanisms, flag-based RRs (with CIs) rank the most discriminative regimes.

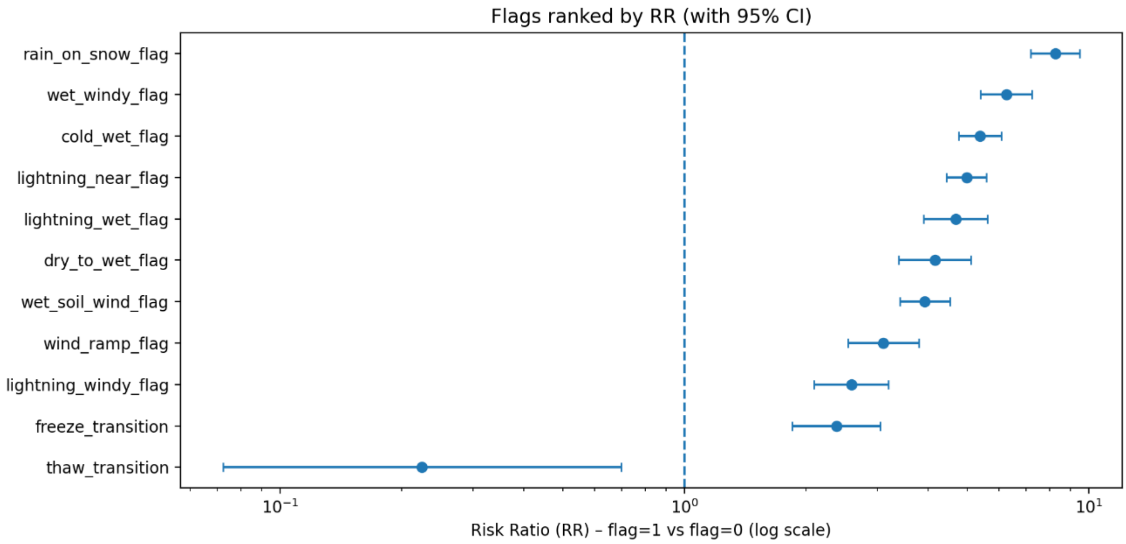


Figure 5.21 Flags Ranked by RR with 95% CI

ROS is the strongest uplift (RR≈8.25, 95% CI [4.17, 16.32]), followed by wet-windy (RR≈6.25 [3.90, 10.01]) and cold-wet (RR≈5.38 [3.88, 7.46]). Lightning-related flags are also strong, especially when combined with wet conditions (e.g., *lightning_near_flag* RR≈4.98 [2.92, 8.50]; *lightning_wet_flag* RR≈4.68 [2.65, 8.29]).

Dynamic-transition indicators matter as well: *wind_ramp_flag* is elevated (RR≈3.10 [2.15, 4.46]) and *freeze_transition* shows RR≈2.38 [1.41, 4.04], while *thaw_transition* appears negatively associated (RR≈0.22 [0.07, 0.70]) and should be interpreted cautiously as an association conditioned on the flag definition and seasonal context.

Stratified analyses show that the dominant compound drivers vary by season. In winter, ROS and cold-wet regimes dominate; in summer, wet-windy and lightning-related compound flags are most prominent; spring and autumn behave as transition seasons with mixed signatures.

Altitude stratification reinforces the mountain vs lowland split. High-altitude strata amplify snowpack and ROS signals, while low-altitude strata emphasize convective/lightning and wet-ground + wind regimes.

5.6. FPI Backtest

This section reports the **out-of-sample walk-forward backtest** of the FPI (Section 4.4). Results are shown year-by-year to evaluate: (i) discrimination / ranking, (ii) calibration realism, and (iii) operational usefulness through capture rates under fixed monitoring budgets. Backtesting starts in 2012 to ensure a minimum historical training window and stable parameter estimates in the early years.

Each yearly fold scores **145k-156k site-days**, with only **34-123 events/year**, confirming an **extreme rare-event regime**. Because the raw FPI stacks multiplicative odds contributions, it can produce overconfident upper-tail probabilities when several sparse multipliers coincide. Platt scaling is therefore applied as a monotonic calibration layer that compresses extremes while preserving ranking.

Figure 5.22 summarizes the fitted Platt parameters across folds. The slope ***a*** decreases over time and the intercept ***b*** shifts more negative, indicating progressively stronger **compression/down-shift** of raw probabilities. This behaviour is consistent with the increasing tendency of multiplicative rare-event models to generate a heavier upper tail in some folds, and shows that calibration adapts **per year** rather than acting as a one-time correction.

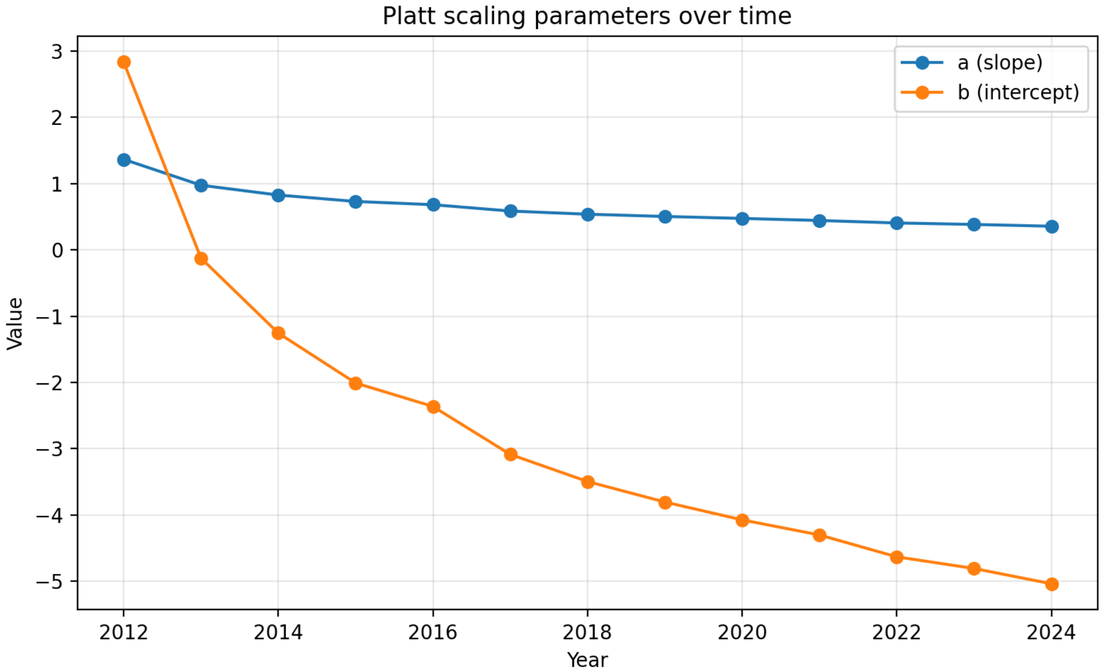


Figure 5.22 Platt Scaling Parameters Over Time

A basic realism check is whether the **mean calibrated probability** remains aligned with the observed prevalence. Figure 5.23 compares the **yearly event rate** to the **mean calibrated prediction**. The calibrated mean stays within the correct **order of magnitude** across years. The observed event rate varies more sharply year-to-year, while the mean prediction is smoother. This is expected under walk-forward evaluation because each fold is trained on past years and is not designed to anticipate year-specific prevalence shifts.

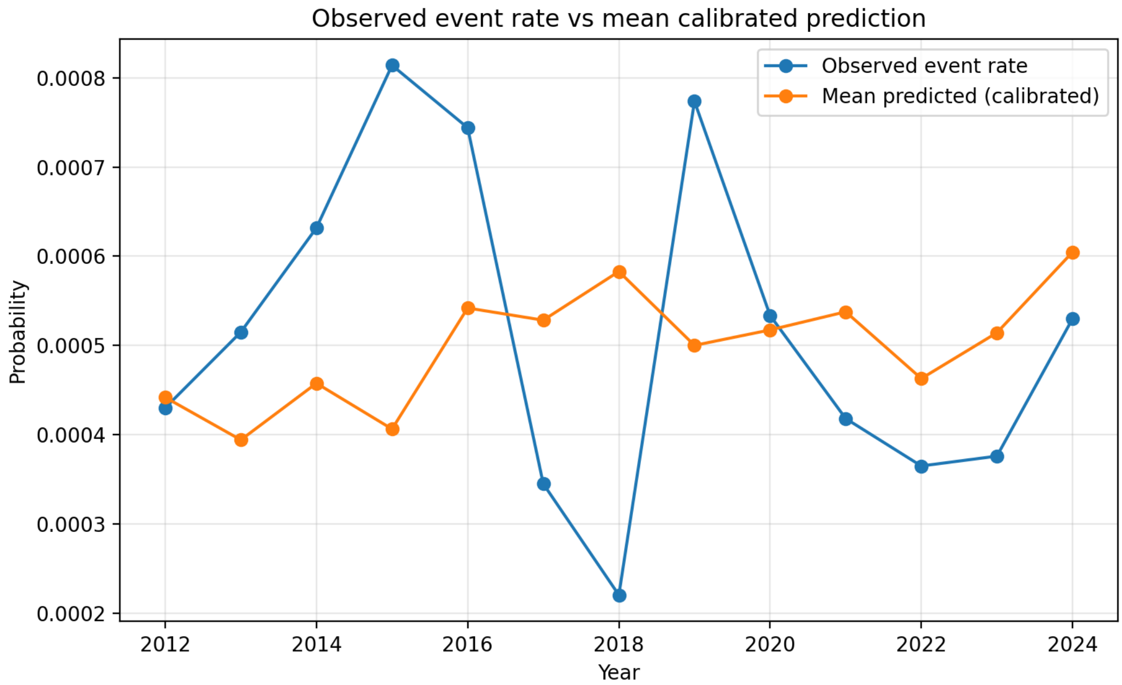


Figure 5.23 Observed Event Rate vs Mean Calibrated Prediction

Discrimination assesses whether the index assigns systematically higher scores to event site-days than to non-event site-days, independently of any fixed threshold. **AUC-ROC** is reported for global ranking ability, and Precision-Recall performance through PR-AUC is computed as AvgPrec. In this rare-event setting, AvgPrec values are necessarily small in absolute terms because the random baseline equals the yearly event rate; therefore, plotting AvgPrec on the same axis as AUC-ROC is visually uninformative. To obtain a more interpretable view, PR performance is expressed as **AvgPrec enrichment**, defined as AvgPrec divided by the yearly event rate (i.e., how many times better than random ranking).

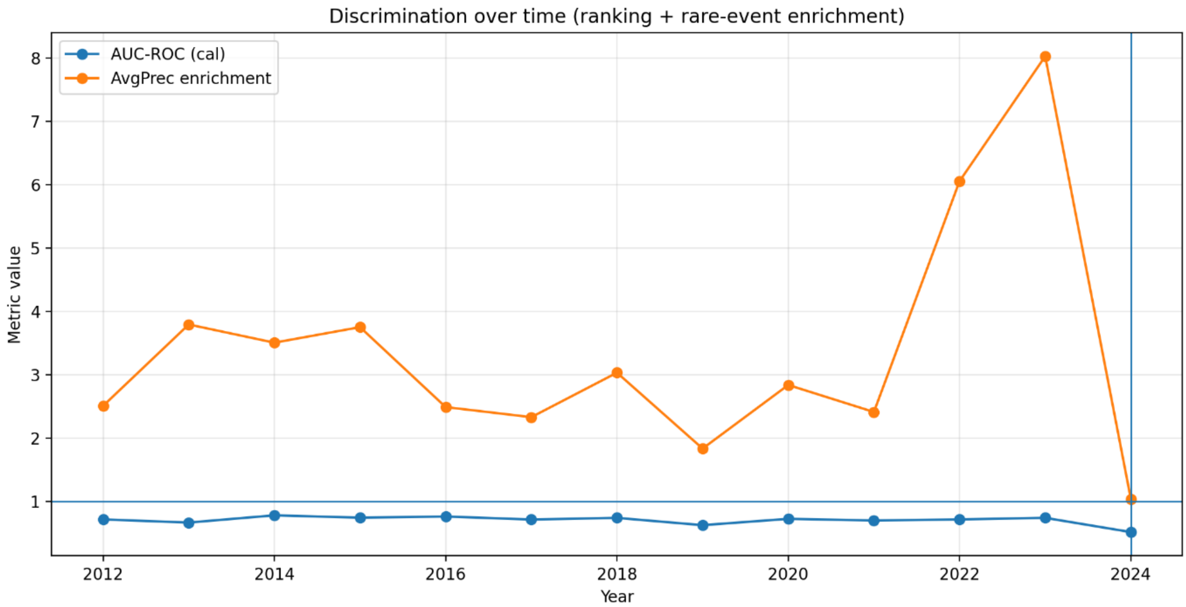


Figure 5.24 Discrimination Over Time

As shown in Figure 5.24, AUC-ROC remains consistently above random throughout 2012-2023, typically around the 0.7 range, indicating stable discrimination across historical years. In parallel, AvgPrec enrichment stays clearly above 1× over 2012-2023 (typically several-fold, with peaks in the most informative years), confirming that observed failures are meaningfully concentrated toward the top of the ranked list rather than being uniformly spread across site-days.

The year **2024** deviates sharply from this historical regime. AUC-ROC drops close to 0.5, and AvgPrec enrichment collapses to ~1×, indicating performance close to random ordering. This joint degradation is consistent with the broader evidence of a **distribution shift** in 2024 relative to the historical structure learned by the index, and it anticipates the collapse observed in top-tail capture diagnostics reported later. Because enrichment normalizes by the yearly base rate, spikes can partly reflect lower prevalence rather than a proportional jump in absolute AvgPrec. For this reason it is interpreted as a relative 'better-than-random' indicator rather than a standalone performance measure.

Because the primary operational use is prioritisation, the most interpretable validation is how many observed failures are captured within the highest risk fraction of site-days. Figure 5.25 reports capture rates for several monitoring budgets, including top **1%, 5%, 10%, 20%, and 30%.**

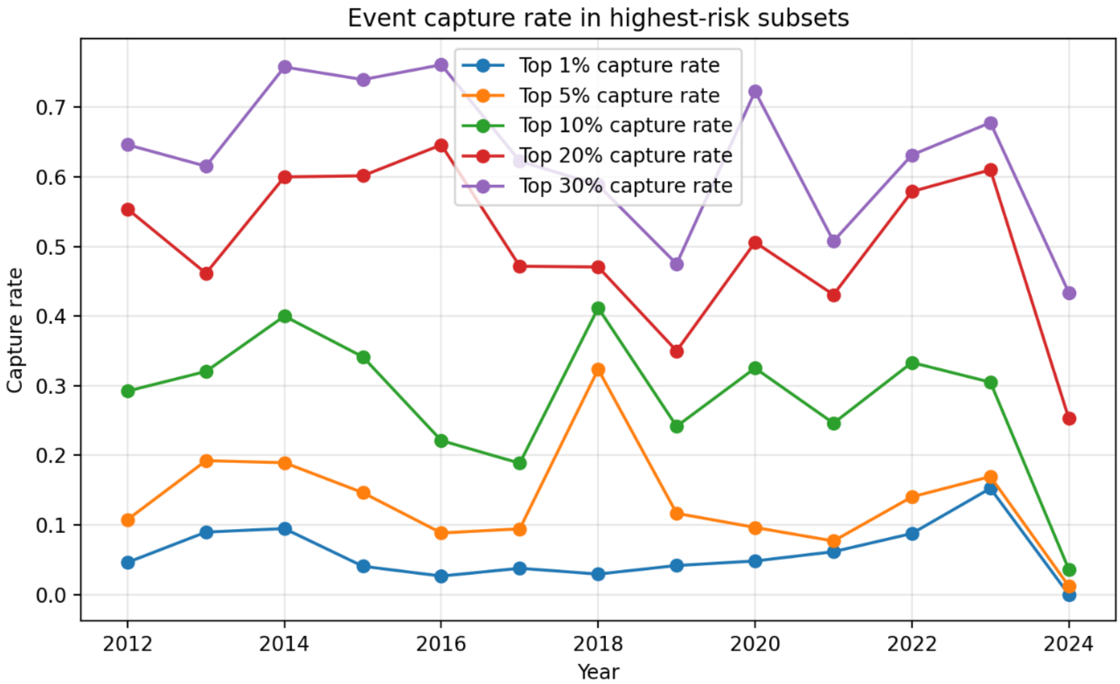


Figure 5.25 Event Capture Rate in Highest-Risk Subsets Over Time

Across 2012 to 2023, the curves show stable enrichment in the upper tail. In practical terms, failures concentrate disproportionately in the highest risk subsets, so monitoring a small fraction of site-days captures a meaningful share of events. This confirms that the index supports the intended operational role of focusing attention on a manageable subset of site-days.

In **2024**, this concentration collapses in the upper tail. The **top 1% capture drops to zero** in the evaluation split, and the top 10% capture becomes close to what would be expected under near random ranking. Larger fractions such as top 20% or top 30% still retain some weak enrichment, suggesting that part of the risk structure remains detectable, but the **extreme-tail mechanisms do not transfer** well to 2024. It indicates that under a limited monitoring capacity, the index would no longer reliably surface failures early in the ranked list in 2024.

Reliability is assessed using decile-based calibration plots comparing the mean predicted probability in each decile against the observed failure rate. In this regime, uncertainty is unavoidable because each decile contains a small number of events, and CIs are therefore wide.

Figure 5.26 shows a representative historical year, 2014, where the decile relationship remains **broadly monotonic**. Higher predicted deciles correspond to higher observed failure rates, and the curve remains qualitatively consistent with the perfect calibration line within uncertainty. The highest decile can even show observed rates exceeding the mean prediction, which is plausible when truly extreme meteorological conditions occur.

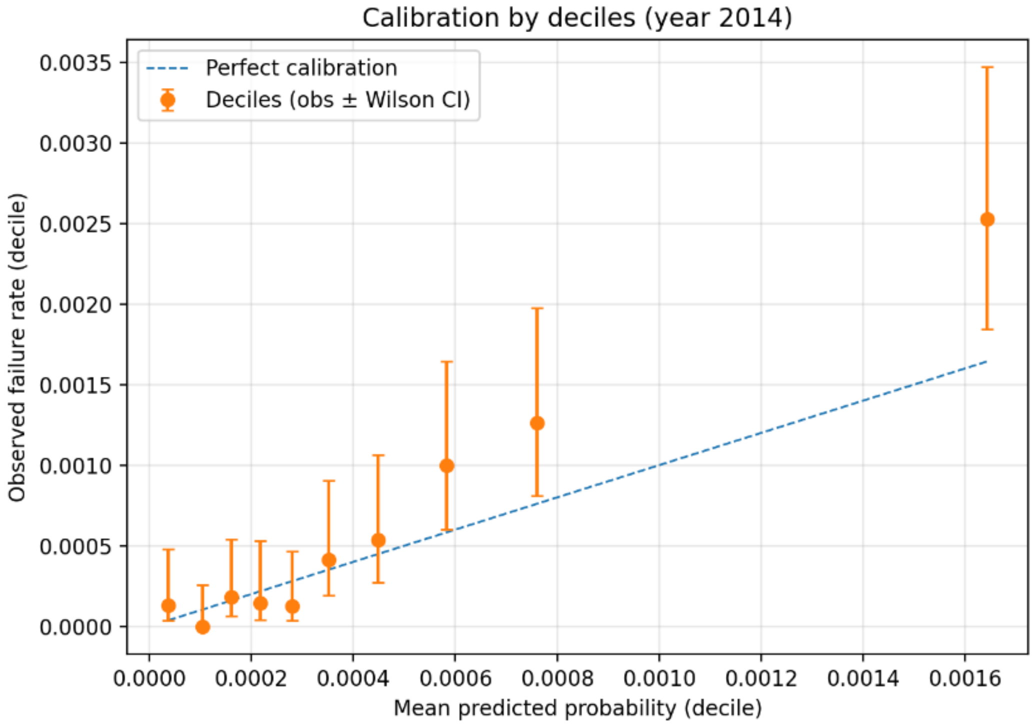


Figure 5.26 Calibration by Deciles (2014)

Figure 5.27 shows the corresponding diagnostic for 2024. The decile relationship becomes weaker and less monotonic. Higher predicted deciles do not consistently correspond to higher observed failure rates, and the highest predicted decile is not enriched in failures. This behaviour is coherent with the drop in AUC-ROC and PR-AUC and with the collapse of upper tail capture, reinforcing the interpretation of 2024 as a distribution shift year.

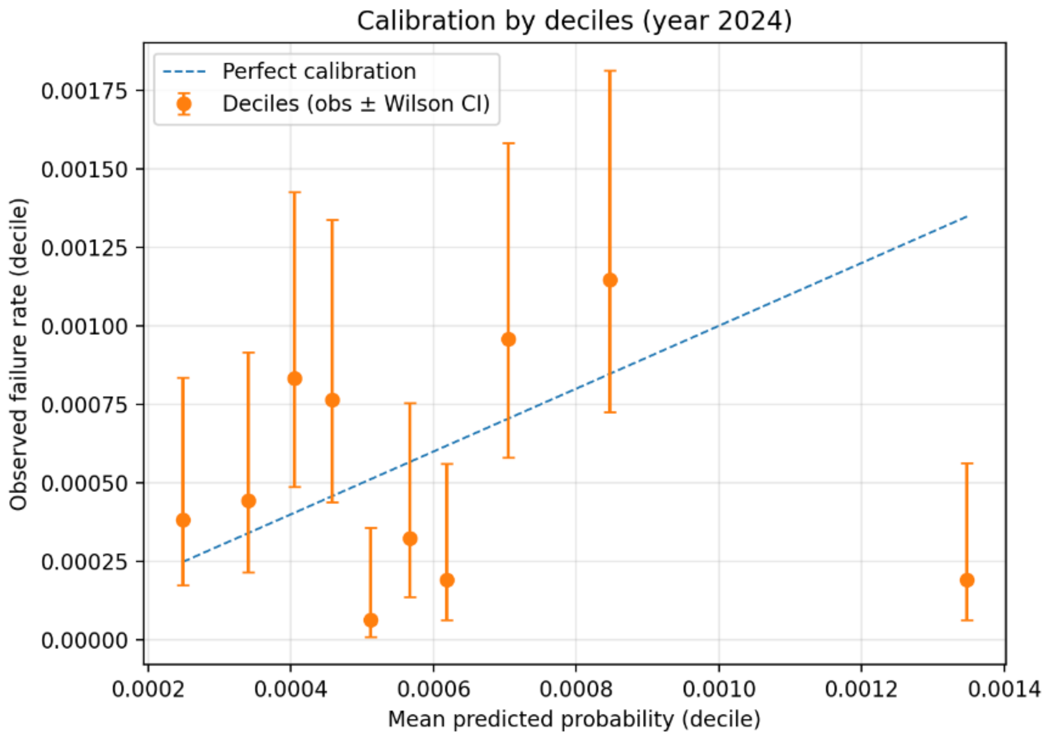


Figure 5.27 Calibration by Deciles (2024)

Finally, Brier score is reported as a global measure of probabilistic accuracy (mean squared error of predicted probabilities). In an extreme rare-event regime, its absolute magnitude is necessarily small, so it is displayed on a logarithmic scale to make year-to-year differences readable. As shown in Figure 5.28, the **calibrated score** is consistently **lower than or comparable to** the raw score, indicating that Platt scaling mitigates the overconfidence introduced by multiplicative odds stacking.

Figure 5.28 Brier Score Over Time

6 Discussion and Conclusions

6.1. Key Findings

Using the site-day database and the reconstruction pipeline, the work shows that it is possible to study weather-failure relationships in a way that is both repeatable and interpretable, even in a region with complex terrain like Piedmont. The key support for this is the explainability audit. 94.58% of Terna-labelled weather-related failure site-days coincide with at least one reconstructed adverse condition under the BROAD definition (80.37% under STRICT), with only 4.66% remaining unexplained when lightning is available (and 0.76% undetermined due to missing lightning). This matters because in Piedmont weather can change quickly over short distances (especially with altitude), so the closest station can be at a different elevation or on the other side of a ridge and record quite different conditions than the user site. A site-level reconstruction reduces this mismatch by translating heterogeneous station observations into comparable exposure variables at the grid site.

A second clear finding is that weather-related failures are better described by compound conditions than by isolated single-variable extremes. Among failure site-days, rain (moderate+) appears in 71.1%, wet soil (wet+) in 65.9%, and lightning in 57.1%, while composite regimes are also frequent (e.g., wet & windy 37.1%, wet soil & wind 26.1%, ROS 18.9%). Looking at co-occurrence, single-driver days exist but are not dominant (13.8%), whereas more than half of failures (~52.6%) happen with 3-4 concurrent main moderators. The most common combination is Rain + Wet soil + Lightning (N=131), followed by Lightning alone (N=90) and Rain + Wind + Wet soil + Lightning (N=87), confirming that “background” wetness and persistence often set the stage, while other hazards act as triggers.

This compound picture is consistent with which drivers show the strongest risk gradients at daily scale. Rainfall intensity displays the steepest monotonic increase (RR rising from 8.02 in moderate rain to 29.94 in very-heavy rain, relative to no rain). Soil wetness, as a persistence proxy, is also strong (RR up to 10.22 in very wet conditions relative to dry). Wind has a clearer but smaller effect (RR up to 4.61 in very strong wind days), and temperature highlights a critical near-zero band (RR 3.15). Interactions reinforce the same message: the highest-risk corners are found where hazards combine (e.g., heavy rain with strong wind, wetness with near-zero temperatures, and lightning risk amplified under wet and/or windy conditions).

Finally, the FPI shows how the evidence can be operationalised into a daily ranking tool. In walk-forward backtesting, the FPI's discrimination is consistently above random throughout 2012-2023 (AUC-ROC typically around the 0.7 range), and failures concentrate toward the top of the ranked list, so monitoring a small fraction of site-days captures a meaningful share of events. At the same time, the backtest highlights the limits expected in an extreme rare-event setting (only 34-123 events/year). Since the raw FPI multiplies odds contributions, it can become overconfident in the upper tail when several sparse multipliers coincide, so Platt scaling is needed to compress extremes while preserving ranking. The 2024 split is a practical stress test of stability. Discrimination drops close to random (AUC-ROC near 0.5) and the top 1% capture drops to zero, signalling that performance can degrade sharply under regime change and/or missing key drivers.

6.2. Limitations

The first limitation is the daily time resolution. Many weather events that can affect the grid are short and intense (e.g., a convective storm lasting one hour, a short wind burst, or a brief period of very intense rainfall). With daily data, these peaks are smoothed out. This means a day can look moderate in the database even if a severe event happened within a short time window, and some failures may therefore look weakly linked to weather simply because the input cannot capture the intensity and timing within the day.

A second limitation is linked to the set of weather drivers represented in the reconstructed exposure. In a small fraction of cases, Terna-labelled weather-related failures occur on site-days where the reconstructed site-level meteorology does not show any clear moderate or extreme condition among the modelled drivers. This does not imply that the label is wrong; instead, it highlights that the reconstructed feature set may be incomplete for some mechanisms. The current pipeline focuses on a defined set of moderators, but other plausible factors are not explicitly represented. Examples include saturated ground combined with vegetation fall, local drainage and runoff dynamics, detailed soil properties and land cover and vegetation proximity.

A third limitation concerns spatial representativeness and interpolation uncertainty. Site exposure is reconstructed from a network of stations that is uneven in space and elevation. Uncertainty is not uniform, it increases in complex terrain and for variables with strong microscale variability, especially wind. Even when the reconstruction is robust on average, some local site-days remain harder to represent, and this uncertainty can propagate to subsequent analyses.

There are also limitations specific to the FPI. Given the strong imbalance of the problem, predicted probabilities are small and can vary across years, and calibration remains noisy when yearly failure counts are limited. Furthermore, the FPI is built by combining multiple drivers, so that the upper tail can become sensitive when several rare drivers occur together, and performance can degrade under distribution shift or when key drivers are missing in a given year. For this reason, the FPI is best interpreted as a practical ranking signal with year-to-year stability that must be monitored rather than assumed.

6.3. Future Work

A natural next step is to apply the same workflow to other regions. Rebuilding the same site-day exposure layer and running the same monthly and daily analyses would make it possible to compare regimes across Italy, checking which drivers are consistent, and understanding how results change with different climates, station densities, and network characteristics. This would also strengthen the value of the database as a common platform for cross-regional benchmarking.

It would also be useful to explore other interpolation options. Other interpolators (or hybrid methods) could provide better accuracy in complex terrain or for specific variables. Testing more methods with the same validation logic would clarify whether the extra complexity brings real improvements.

Finally, the FPI can be improved and extended. The current index is presented as an interpretable, ranking-oriented score built from the generated moderators, but there is clear scope to refine its components (feature definitions, binning strategy, interaction handling), strengthen calibration, and make it more robust to year-to-year changes. Testing alternative formulations while keeping the same walk-forward protocol would clarify which design choices best improve stability and operational utility.

Bibliography

[1] Terna, 'Metodologia per il calcolo del beneficio per l'incremento della resilienza della Rete di Trasmissione Nazionale', RSE, Roma, Presentation / Webinar, May 2021.

[2] European Commission. Joint Research Centre., *Climate change and critical infrastructure: storms.* LU: Publications Office, 2019. doi: 10.2760/986436.

[3] V. Do *et al.*, 'Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA', *Nat Commun*, vol. 14, no. 1, p. 2470, Apr. 2023, doi: 10.1038/s41467-023-38084-6.

[4] Autorità di Regolazione per Energia Reti e Ambiente (ARERA), 'Relazione annuale all'Agenzia internazionale per la cooperazione fra i regolatori nazionali dell'energia e alla Commissione europea sull'attività svolta e i compiti dell'Autorità di regolazione per energia reti e ambiente', Milano, 354/2025/I, Jul. 2025.

[5] F. Cavalleri *et al.*, 'Hourly Precipitation Patterns and Extremization over Italy using convection-permitting reanalysis data', *Nat. Hazards Earth Syst. Sci.*, vol. 26, no. 1, pp. 279–297, Jan. 2026, doi: 10.5194/nhess-26-279-2026.

[6] E. Ciapessoni, A. Pitto, and D. Cirio, 'An Application of a Risk-Based Methodology to Anticipate Critical Situations Due to Extreme Weather Events in Transmission and Distribution Grids', *Energies*, vol. 14, no. 16, p. 4742, Aug. 2021, doi: 10.3390/en14164742.

[7] M. Beniston and M. Stoffel, 'Rain-on-snow events, floods and climate change in the Alps: Events may increase with warming up to 4 °C and decrease thereafter', *Science of The Total Environment*, vol. 571, pp. 228–236, Nov. 2016, doi: 10.1016/j.scitotenv.2016.07.146.

[8] F. Z. Maina and S. V. Kumar, 'Global patterns of rain-on-snow and its impacts on runoff from past to future projections', *Nat Commun*, vol. 16, no. 1, p. 4731, May 2025, doi: 10.1038/s41467-025-59855-3.

[9] O. Hotovy, O. Nedelcev, J. Seibert, and M. Jenicek, 'Rain-on-snow events in mountainous catchments under climate change', *Hydrol. Earth Syst. Sci.*, vol. 29, no. 17, pp. 4199–4217, Sep. 2025, doi: 10.5194/hess-29-4199-2025.

[10] W. T. Sittner, C. E. Schauss, and J. C. Monro, 'Continuous hydrograph synthesis with an API-type hydrologic model', *Water Resources Research*, vol. 5, no. 5, pp. 1007–1022, Oct. 1969, doi: 10.1029/WR005i005p01007.

[11] X. Li, Y. Wei, and F. Li, 'Optimality of antecedent precipitation index and its application', *Journal of Hydrology*, vol. 595, p. 126027, Apr. 2021, doi: 10.1016/j.jhydrol.2021.126027.

[12] T. Page, K. J. Beven, B. Hankin, and N. A. Chappell, 'Interpolation of rainfall observations during extreme rainfall events in complex mountainous terrain', *Hydrological Processes*, vol. 36, no. 11, p. e14758, Nov. 2022, doi: 10.1002/hyp.14758.

[13] A. Napoli, A. Crespi, F. Ragone, M. Maugeri, and C. Pasquero, 'Variability of orographic enhancement of precipitation in the Alpine region', *Sci Rep*, vol. 9, no. 1, p. 13352, Sep. 2019, doi: 10.1038/s41598-019-49974-5.

[14] A. Moschetti, L. Martiri, C. Laurano, L. Cristaldi, D. Martinenghi, and M. Faifer, 'A Data Preparation Pipeline for the Analysis of the Italian Power Grid Reliability', in *2025 International Conference on Clean Electrical Power (ICCEP)*, Villasimius, Italy: IEEE, Jun. 2025, pp. 1002–1007. doi: 10.1109/ICCEP65222.2025.11143620.

[15] S. Duvnjak Žarković, X. Weiss, and P. Hilber, 'Addressing data deficiencies in outage reports: A qualitative and machine learning approach', *Electric Power Systems Research*, vol. 236, p. 110901, Nov. 2024, doi: 10.1016/j.epsr.2024.110901.

[16] Z. Zhang, H. Zhang, S. Yue, and W. Zeng, 'A Review of Icing and Anti-Icing Technology for Transmission Lines', *Energies*, vol. 16, no. 2, p. 601, Jan. 2023, doi: 10.3390/en16020601.

[17] O. Chabart and J. L. Lilien, 'Galloping of electrical lines in wind tunnel facilities', *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 74–76, pp. 967–976, Apr. 1998, doi: 10.1016/S0167-6105(98)00088-9.

[18] M. Li *et al.*, 'Thunderstorm total lightning activity behavior associated with transmission line trip events of power system', *npj Clim Atmos Sci*, vol. 7, no. 1, p. 148, Jun. 2024, doi: 10.1038/s41612-024-00697-z.

[19] A. A. Salem, K. Y. Lau, Z. Abdul-Malek, S. A. Al-Gailani, and C. W. Tan, 'Flashover voltage of porcelain insulator under various pollution distributions: Experiment and modeling', *Electric Power Systems Research*, vol. 208, p. 107867, Jul. 2022, doi: 10.1016/j.epsr.2022.107867.

[20] M. M. Hussain, S. Farokhi, S. G. McMeekin, and M. Farzaneh, 'Mechanism of saline deposition and surface flashover on outdoor insulators near coastal areas part II: Impact of various environment stresses', *IEEE Trans. Dielect. Electr. Insul.*, vol. 24, no. 2, pp. 1068–1076, Apr. 2017, doi: 10.1109/TDEI.2017.006386.

[21] C. Yin, Z. Xiao, Y. Guo, C. Shi, X. Zhang, and G. Wu, 'Method for detecting the pollution degree of naturally contaminated insulator based on hyperspectral

characteristics', *High Voltage*, vol. 6, no. 6, pp. 1031–1039, Dec. 2021, doi: 10.1049/hve2.12076.

[22] J. Zuo, J. Fan, Y. Ouyang, H. Liu, C. Yang, and C. Hao, 'Transmission Line Sag Measurement and Simulation Research Based on Non-Contact Electric Field Sensing', *Sensors (Basel)*, vol. 22, no. 21, p. 8379, Nov. 2022, doi: 10.3390/s22218379.

[23] M. T. Bedialauneta, E. Fernandez, I. Albizu, and A. J. Mazon, 'Sag-tension evaluation of high-temperature gap-type conductor in operation', *IET Generation Trans & Dist*, vol. 16, no. 1, pp. 19–26, Jan. 2022, doi: 10.1049/gtd2.12288.

[24] D. Cerrai, P. Watson, and E. N. Anagnostou, 'Assessing the effects of a vegetation management standard on distribution grid outage rates', *Electric Power Systems Research*, vol. 175, p. 105909, Oct. 2019, doi: 10.1016/j.epsr.2019.105909.

[25] C. Manning, S. Wilkinson, H. J. Fowler, and E. J. Kendon, 'Antecedent rainfall, wind direction and seasonal effects may amplify the risk of wind-driven power outages in the UK', *Commun Earth Environ*, vol. 6, no. 1, p. 217, Mar. 2025, doi: 10.1038/s43247-025-02176-6.

[26] R. B. Duffey, 'Power Restoration Prediction Following Extreme Events and Disasters', *Int J Disaster Risk Sci*, vol. 10, no. 1, pp. 134–148, Mar. 2019, doi: 10.1007/s13753-018-0189-2.

[27] V. Do, L. B. Wilner, N. M. Flores, H. McBrien, A. J. Northrop, and J. A. Casey, 'Spatiotemporal patterns of individual and multiple simultaneous severe weather events co-occurring with power outages in the United States, 2018–2020', *PLOS Clim*, vol. 4, no. 1, p. e0000523, Jan. 2025, doi: 10.1371/journal.pclm.0000523.

[28] J. Li and A. D. Heap, 'Spatial interpolation methods applied in the environmental sciences: A review', *Environmental Modelling & Software*, vol. 53, pp. 173–189, Mar. 2014, doi: 10.1016/j.envsoft.2013.12.008.

[29] D. Shepard, 'A two-dimensional interpolation function for irregularly-spaced data', in *Proceedings of the 1968 23rd ACM National Conference*, in ACM '68. New York, NY, USA: ACM Press, 1968, pp. 517–524. doi: 10.1145/800186.810616.

[30] N. A. C. Cressie, *Statistics for spatial data*, Rev. ed. in Wiley series in probability and mathematical statistics. New York: Wiley, 1993.

[31] T. Hengl, *A practical guide to geostatistical mapping*, 2nd extended ed. Amsterdam: Hengl, 2009.

[32] M. F. Hutchinson, 'Interpolating mean rainfall using thin plate smoothing splines', *International journal of geographical information systems*, vol. 9, no. 4, pp. 385–403, Jul. 1995, doi: 10.1080/02693799508902045.

[33] N. Hofstra, M. Haylock, M. New, P. Jones, and C. Frei, 'Comparison of six methods for the interpolation of daily, European climate data', *J. Geophys. Res.*, vol. 113, no. D21, p. 2008JD010100, Nov. 2008, doi: 10.1029/2008JD010100.

[34] J. Hauke and T. Kossowski, 'Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data', *Quaestiones Geographicae*, vol. 30, no. 2, pp. 87–93, Jun. 2011, doi: 10.2478/v10117-011-0021-1.

[35] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, 'STL: A Seasonal-Trend Decomposition Procedure Based on Loess', vol. 6, no. 1, Journal of Official Statistics, pp. 3–33, 1990.

[36] S. Greenland, 'Model-based Estimation of Relative Risks and Other Epidemiologic Measures in Studies of Common Outcomes and in Case-Control Studies', *American Journal of Epidemiology*, vol. 160, no. 4, pp. 301–305, Aug. 2004, doi: 10.1093/aje/kwh221.

[37] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, 'A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise', presented at the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996, pp. 226–231.

[38] P. J. Clark and F. C. Evans, 'Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations', *Ecology*, vol. 35, no. 4, pp. 445–453, Oct. 1954, doi: 10.2307/1931034.

[39] M. Panteli and P. Mancarella, 'Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies', *Electric Power Systems Research*, vol. 127, pp. 259–270, Oct. 2015, doi: 10.1016/j.epsr.2015.06.012.

[40] D. W. Wanik, E. N. Anagnostou, B. M. Hartman, M. E. B. Frediani, and M. Astitha, 'Storm outage modeling for an electric distribution network in Northeastern USA', *Nat Hazards*, vol. 79, no. 2, pp. 1359–1384, Nov. 2015, doi: 10.1007/s11069-015-1908-2.

[41] D. W. Wanik *et al.*, 'A Case Study on Power Outage Impacts from Future Hurricane Sandy Scenarios', *Journal of Applied Meteorology and Climatology*, vol. 57, no. 1, pp. 51–79, Jan. 2018, doi: 10.1175/JAMC-D-16-0408.1.

[42] L. Souto, R. Neal, J. O. Pope, P. L. M. Gonzalez, J. Wilkinson, and P. C. Taylor, 'Identification of weather patterns and transitions likely to cause power outages in the United Kingdom', *Commun Earth Environ*, vol. 5, no. 1, p. 49, Jan. 2024, doi: 10.1038/s43247-024-01217-w.

[43] G. Adinolfi, R. Ciavarella, and A. Merola, 'Analisi e valutazione preliminare delle problematiche di affidabilità delle reti di distribuzione ibride AC/DC', ENEA, RdS/PTR2019/159, 2019.

[44] R. Otto, S. Susanne, H. Jouni, and L. Jukka, 'Crown snow load outage risk model for overhead lines', *Applied Energy*, vol. 343, p. 121183, Aug. 2023, doi: 10.1016/j.apenergy.2023.121183.

[45] H. Liu, R. A. Davidson, and T. V. Apanasovich, 'Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms', *Reliability Engineering & System Safety*, vol. 93, no. 6, pp. 897–912, Jun. 2008, doi: 10.1016/j.ress.2007.03.038.

[46] S. Han, S. D. Guikema, and S. M. Quiring, 'Improving the Predictive Accuracy of Hurricane Power Outage Forecasts Using Generalized Additive Models', *Risk Analysis*, vol. 29, no. 10, pp. 1443–1453, Oct. 2009, doi: 10.1111/j.1539-6924.2009.01280.x.

[47] J. He, D. W. Wanik, B. M. Hartman, E. N. Anagnostou, M. Astitha, and M. E. B. Frediani, 'Nonparametric Tree-Based Predictive Modeling of Storm Outages on an Electric Distribution Network', *Risk Analysis*, vol. 37, no. 3, pp. 441–458, Mar. 2017, doi: 10.1111/risa.12652.

[48] Terna S.p.A., 'Elenco disalimentazioni 2024'. 2024.

[49] A. Derdouri and Y. Murayama, 'A comparative study of land price estimation and mapping using regression kriging and machine learning algorithms across Fukushima prefecture, Japan', *J. Geogr. Sci.*, vol. 30, no. 5, pp. 794–822, May 2020, doi: 10.1007/s11442-020-1756-1.

[50] G. Y. Lu and D. W. Wong, 'An adaptive inverse-distance weighting spatial interpolation technique', *Computers & Geosciences*, vol. 34, no. 9, pp. 1044–1055, Sep. 2008, doi: 10.1016/j.cageo.2007.07.010.

[51] T. Hengl, G. B. M. Heuvelink, and D. G. Rossiter, 'About regression-kriging: From equations to case studies', *Computers & Geosciences*, vol. 33, no. 10, pp. 1301–1315, Oct. 2007, doi: 10.1016/j.cageo.2007.05.001.

[52] S. P. Serbin and C. J. Kucharik, 'Spatiotemporal Mapping of Temperature and Precipitation for the Development of a Multidecadal Climatic Dataset for Wisconsin', *Journal of Applied Meteorology and Climatology*, vol. 48, no. 4, pp. 742–757, Apr. 2009, doi: 10.1175/2008JAMC1986.1.

[53] G. E. Fasshauer, *Meshfree approximation methods with MATLAB*. in Interdisciplinary mathematical sciences, no. v. 6. Singapore ; Hackensack, N.J: World Scientific, 2007.

[54] P. Goovaerts, 'Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall', *Journal of Hydrology*, vol. 228, no. 1–2, pp. 113–129, Feb. 2000, doi: 10.1016/S0022-1694(00)00144-X.

[55] I. O. A. Odeh, A. B. McBratney, and D. J. Chittleborough, 'Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and

regression-kriging', *Geoderma*, vol. 67, no. 3–4, pp. 215–226, Aug. 1995, doi: 10.1016/0016-7061(95)00007-B.

[56] D. R. Cox, 'The Regression Analysis of Binary Sequences', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 20, no. 2, pp. 215–232, Jul. 1958, doi: 10.1111/j.2517-6161.1958.tb00292.x.

[57] N. Mantel and W. Haenszel, 'Statistical aspects of the analysis of data from retrospective studies of disease', *J Natl Cancer Inst*, vol. 22, no. 4, pp. 719–748, Apr. 1959.

[58] M. J. Knol and T. J. VanderWeele, 'Recommendations for presenting analyses of effect modification and interaction', *International Journal of Epidemiology*, vol. 41, no. 2, pp. 514–520, Apr. 2012, doi: 10.1093/ije/dyr218.

[59] F. J. Anscombe, 'On estimating binomial response relations', *Biometrika*, vol. 43, no. 3–4, pp. 461–464, 1956, doi: 10.1093/biomet/43.3-4.461.

[60] J. Platt, 'Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods', vol. 10, Adv. Large Margin Classif., Jun. 23, 2000.

Acknowledgments

I would like to sincerely thank my co-supervisors, Andrea Moschetti and Luca Martiri, for their constant support, availability, and valuable guidance throughout this work. Their feedback and encouragement were essential in shaping the work and bringing it to completion. I also thank Prof. Loredana Cristaldi for the opportunity to carry out this project.