Project 2413.1 | September 2026

SJSU SAN JOSÉ STATE UNIVERSITY

MINETA TRANSPORTATION INSTITUTE

Identification of At-Risk Rail Network Segments and Development of Fragility-based Resilience Metrics

Haizhong Wang, PhD
Brian M. Staes, PhD
Benyamin Ghoreishi, MSc

Oregon State University

CLEMSON UNIVERSITY

MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1 and the Rail Resilience to Severe Weather: Training and Research Program (RR2SW) funded by the Federal Railroad Administration. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and goods, we must prepare the next generation of skilled transportation professionals who can lead a thriving, forward-thinking transportation industry for a more connected world. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 26-27

Identification of At-Risk Rail Network Segments and Development of Fragility-based Resilience

Haizhong Wang, PhD
Brian M. Staes, PhD
Benyamin Ghoreishi, MSc

September 2026

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

10.31979/mti.2026.2413.1

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2413.1

This research was supported by the Federal Railroad Administration (FRA) through Consolidated Rail Infrastructure and Safety Improvements (CRISI) Grant No. 69A36523420190CRSCA.

Contents

Illustrations

Tables

1. Introduction

Clemson University, as a sub-awardee under a Federal Railroad Administration (FRA) Consolidated Rail Infrastructure and Safety Improvement (CRISI) grant number 69A36523420190CRSCA, awarded to San Jose State University, developed and demonstrated new methods to evaluate natural hazard risk across the U.S. rail network. This report presents the results of that work, including a national flash flood risk framework linked to rail infrastructure, an extreme heat fragility approach for thermal buckling risk, and the foundational components of a resilience quantification system. The analyses integrate national environment and meteorological datasets with FRA infrastructure records to identify at-risk rail segments, characterize hazard exposure, and support evidence-based resilience planning. The research team conducted the research between May 2024 and November 2025.

1.1 Background

The U.S. rail network is a critical component of the national transportation system, supporting both freight and intercity passenger travel across various and often hazard-prone environments. Natural hazards, including flash flooding, riverine flooding, extreme heat, extreme cold, and landslide, pose recurring threats to rail infrastructure. These hazards can cause derailments, damage trackbeds and structures, and disrupt operations.

Despite the recognized risk, consistent national methods for quantifying where and how severely hazards affect rail infrastructure have been limited. FRA accident and incident records provide partial coverage but are subject to monetary reporting thresholds that exclude many weather-driven disruptions. Maintenance records, slow order logs, and outage duration data, which are essential for building complete resilience curves, are generally not publicly available. This gap limits the ability of federal agencies, owners, and operators to identify the most vulnerable corridors, prioritize investments, and estimate the consequences of future hazard scenarios.

This research was conducted to address those gaps by developing data-driven, nationally scalable methods that combine publicly available meteorological datasets with FRA infrastructure records.

1.2 Objectives

This project had three primary objectives, each corresponding to a subtask of the research:

- Locate the natural hazard data and information needed from rail owners and operators to better understand current risk management practices and operational resilience strategies (see Chapter 2).
- Identify potential at-risk locations, vulnerable segments, and critical links in the U.S. rail network, and develop metrics to measure network-level impacts from natural hazards (see Chapter 3).
- Develop railway fragility functions and establish loss-of-function relationships to support the creation of restoration curves for resilience quantification (see Chapter 4).

Together, these objectives provide a scalable, integrated hazard and resilience assessment framework that federal agencies and rail owners can apply nationally.

1.3 Overall Approach

The research team built the framework by integrating national datasets from multiple federal agencies, including NOAA, NCEI, NREL's National Solar Radiation Database (NSRDB), the National Weather Service (NWS), the U.S. Geological Survey (USGS), and FRA infrastructure sources with the FRA Form 54C accident and incident database and the North American Rail Network (NARN) shapefile.

For flash flood risk, the researchers linked more than 53,000 historical flash flood records from the NCEI Storm Events Database to the rail network using 12-digit Hydrologic Unit Code (HUC-12) watersheds as the spatial reference unit. They paired these records with hourly precipitation data from approximately 1,489 Local Climatological Data (LCD) stations and applied extreme value methods. This includes the Generalized Pareto Distribution (GPD) for peaks-over-threshold analysis and the Generalized Extreme Value (GEV) distribution for intensity-duration-frequency (IDF) curve generation. To estimate event intensity, duration, and return period these metrics were translated into two segment-level risk scores: a maximum return-period score reflecting acute, single-event exposure and a cumulative score capturing chronic, repeated exposure.

For extreme heat, the researchers used FRA Form 54C derailment records to identify thermal buckling events (cause code T109) and reconstructed rail-track temperatures for each event day using a simplified temperature model calibrated against Amtrak monitoring data. Estimated rail temperatures were placed in their historical context using NSRDB data extending back to 2000, allowing each derailment to be ranked as a percentile within the local temperature distribution. The researchers then developed a reduced-form fragility framework using principal component analysis (PCA) to combine rail-track temperature and train speed into a single intensity measure.

Network criticality analysis was explored using graph-theoretic methods applied to the NARN shapefile, including node-based and link-based measures. Where data limitations prevented full implementation, the report describes the required inputs and a practical pathway using Bureau of Transportation Statistics Freight Analysis Framework (FAF) data as an interim approach.

Together, these analyses provide a foundation for multi-hazard, link-level assessment of rail resilience at the national scale. The researchers converted meteorological and operational data into exposure measures and risk scores that agencies can use for resilience planning. The results also point to next steps: a network-wide criticality analysis (node/link/system) that accounts for natural hazards and human-caused disruptions, and decision-support tools that prioritize inspection, investment, and adaptation using up-to-date environmental datasets.

1.4 Scope

This research focused on four natural hazards: flash flooding, riverine flooding, extreme heat (thermal buckling), and landslides. The analysis covers the contiguous United States rail network and the study period generally spans 2000 through 2024, aligned with the availability of the NSRDB and NCEI datasets.

The flash flood and extreme heat analyses were developed to the point of producing national risk scores and a fragility relationship, respectively. Full resilience curves, which require outage duration and restoration time data by damage state, were not completed due to the absence of publicly available operational disruption records. Network criticality metrics were explored conceptually and in prototype form, but meaningful application was limited by the lack of origin-destination freight flow data at the segment level.

The research does not include a comprehensive assessment of extreme cold or landslides beyond characterization; those hazards are identified and described but are prioritized for future work. The framework does not replace FRA operational guidance or owner-specific maintenance planning tools; rather, it provides a national screening layer to identify where more targeted assessment is warranted.

In alignment with FRA's mission, this study combines national climate datasets with FRA infrastructure data to describe historical risk while also supporting forward-looking planning. The framework provides a scalable, data-driven approach to strengthen rail safety and transportation security.

1.5 Organization of Report

Section 2 describes our data acquisition process, the main limitations of the current study, and the data needed to support national-scale rail risk assessment. The researchers identify key data gaps and outline what is required to estimate the rail network's adaptive capacity to flash flooding, riverine flooding, extreme heat, and landslides/slope instability.

Section 3 presents the flash-flood risk framework for NARN developed by Clemson University. The researchers linked more than 53,000 historical flash-flood records from NCEI to the rail network and combined them with observations from about 1,400 meteorological stations. The researchers then applied extreme value methods to estimate event intensity, duration, and frequency near rail segments. Using these results, the researchers computed risk scores based on maximum return periods and cumulative exposure. The maps show concentrated flash-flood exposure in the Southeast, Midwest, and parts of the Northeast. These areas represent priorities for inspection, reinforcement, and targeted capital investment.

Section 4 evaluates extreme-heat risk and the relationships between track temperature, train speed, and track buckling-related derailments. The researchers developed a reduced-form fragility model that estimates buckling likelihood as a function of observed thermal conditions and operating speed. The results separate incidents that occurred under typical operating conditions (suggesting limited track resistance or other contributing factors) from incidents associated with high track temperatures and higher speeds (consistent with elevated thermal-buckling risk).

The appendix presents state-based PCA results and state-level maps of the PCA fragility results for the nine states with the highest concentrations of thermal buckling derailments.

2. Nationwide Survey of Natural Hazards that Affect US Rail Network

Task 1 of this subproject, published in Wang et al. (2025), documented the range of natural hazards that affect the U.S. rail network. These hazards include pluvial and fluvial flooding, extreme heat that can trigger thermal buckling, extreme cold that can contribute to rail cracking, and landslides or washouts. These events have affected rail infrastructure across the country.

To map where exposure occurs, the research team combined FRA Accident/Incident records with external datasets from the National Weather Service (NWS), the National Oceanic and Atmospheric Administration (NOAA), the U.S. Geological Survey (USGS), and other agencies. Our goal was not to replace FRA reports. Instead, the researchers used these external records to identify rail corridors that face recurring hazard exposure.

FRA reporting requirements depend in part on a monetary reporting threshold tied to labor and equipment costs. As a result, many hazard-related disruptions since 2000 may not appear in the FRA accident database because they did not exceed the threshold, did not cause a derailment, or were not recorded.

This limitation is visible in accident codes T001, T002, and T109, where the total number of reported events is close to the total number of derailments. Many hazard-driven impacts may be handled as routine maintenance rather than reported as accidents, but maintenance records are not publicly available. This creates a gap in verification data when the researchers try to estimate how often hazards affect rail infrastructure and operations, and it adds uncertainty when the researchers translate historical experience into future risk.

For these reasons, this section summarizes the expected information and potential outputs that a nationwide survey of Class I and short line owners and operators would produce to satisfy the existing data limitations gap. The outputs of a survey would provide additional evidence on hazard impacts and strengthen the results from the other tasks in this report.

2.1 Limitations and Required Data for Internal Project Tasks

2.1.1 Limitations - Identify the Potential At-Risk Vulnerable Segments and Critical Link Locations

While the research team considered external natural hazard data sources in conjunction with the FRA incident and derailment dataset, an important limitation is the absence of owner-maintained datasets on natural hazards and historical hazard-related damage to rail assets. This limits the ability to identify and evaluate rail segments that may be most vulnerable or at risk, and to assess the extent of rail owners' preparation for hazards within already exposed areas. To address this limitation, the research team compiled generalized design criteria from American Railway Engineering and Maintenance-of-Way Association (AREMA) and other design and guidance manuals developed and used by rail owners. These criteria were organized into the three dominant natural hazards affecting rail systems and are presented in Table 1.

Table 1. Design Criteria and Standards for Natural Hazards

Railroad	Flooding / Drainage	Extreme Heat / CWR Neutral-Temp (DRNT)	Landslide / Rockfall Prevention
BNSF	Guidelines for Industry Track Projects – Drains for Steep Slopes (Sheet A–9)[1]; AREMA Manual Ch. 6[2]	FRA CWR Generic Plan (±20 °F around DRNT)[3]; AREMA Ch. 30[2]	BNSF-UP Joint Shoring Guidelines[5]; AREMA Ch. 6[2]
Union Pacific	UPRR/BNSF Joint Shoring Guidelines[5]; AREMA Ch. 6[2]	FRA CWR Plan (±20 °F)[3]; 49 CFR 213.118–119[4]	UPRR/BNSF Shoring Guidelines[5]; AREMA Ch. 6[2]
CSX	CSX Pipeline Occupancy Specs – 100-yr storm culvert design[6]; AREMA Ch. 6[2]	FRA CWR Generic Plan (±20 °F)[3]	OptaSense slide detection[7]; AREMA Ch. 6[2]
Norfolk Southern	NS Industrial-Track Specs[8]; AREMA Ch. 6[2]	FRA CWR Generic Plan (±20 °F)[3]; CWR Plan filing[4]	NS geohazard partnerships[9]; AREMA Ch. 6[2]
Canadian National	CN Culvert Crossings Spec[10]; AREMA Ch. 6[2]	Neutral temp 105 °F per AREMA/FRA[10] [3]	InSAR slope monitoring[11]; AREMA Ch. 6[2]
CPKC	Grade-Separation Guidelines[12]; AREMA/FEMA culverts; AREMA Ch. 6[2]	FRA CWR Generic Plan (±20 °F)[3]; AREMA Ch. 30[2]	Slope-stability in grade-seps[12]; AREMA Ch. 6[2]
Canadian Pacific	Public Projects Manual[13]; AREMA drainage[2]	AREMA/FRA CWR (±20 °F)[3] [2]	Slope-stabilization practices[13]; AREMA Ch. 6[2]

Notes: [1] BNSF Railway (2020). [2] AREMA (2021). [3] FRA (2021). [4] FRA Federal Register (2019). [5] Union Pacific and BNSF (2022). [6] CSX Transportation (2021). [7] OptaSense (2019). [8] Norfolk Southern (2022). [9] Norfolk Southern (2023). [10] Canadian National (2020). [11] Wasowski & Bovenga (2014). [12] CPKC (2022). [13] Canadian Pacific (2021).

The researchers found general guidance documents, but they needed rail owner and operator specific internal surveys and better data access to estimate hazard costs accurately. The biggest gap is the lack of information on routine maintenance and preparation practices for natural hazards. Without these records, the researchers could not reliably connect hazard exposure to repair actions, costs, or service impacts.

If the research team obtained maintenance and repair records, they could build a baseline for future scenario analysis. That baseline would help us estimate which links are likely to be disrupted, how long outages may last, and which commodities would be affected under different hazard scenarios. It would also allow us to separate routine maintenance from hazard-driven repairs.

In Task 2, the researchers identified at-risk locations and vulnerable segments and linked them to hazard-specific datasets, but our validation relied almost entirely on derailment and incident reports. This likely undercounts many hazard impacts that disrupt operations but do not exceed reporting thresholds or do not produce derailments. For example, their analysis suggested that extreme rainfall impacts are substantially underrepresented in the incident data, and they saw a similar gap for extreme heat impacts.

To close this gap, the researchers need survey responses and datasets that describe maintenance actions, repair costs, and repair locations over time. With those data, they could build a predictive model that takes hazard intensity as an input and estimates affected locations, expected outage duration, and expected costs as outputs.

Critical links-nodes and network disconnections

To develop link criticality, the research team analyzed how the FRA North American Rail Network (NARN) shapefile represents nodes across the U.S. rail system. The researchers used the "Major Industrial Lead (I)" node designation to identify freight loading and unloading sites that connect to mainline rail segments.

The researchers then grouped these nodes by Class I owner and by the railroads that have operating rights on each track segment. This step allowed us to build a complete set of plausible origin–destination (OD) pairs for industrial freight movements across the network. Figure 1 summarizes this workflow.

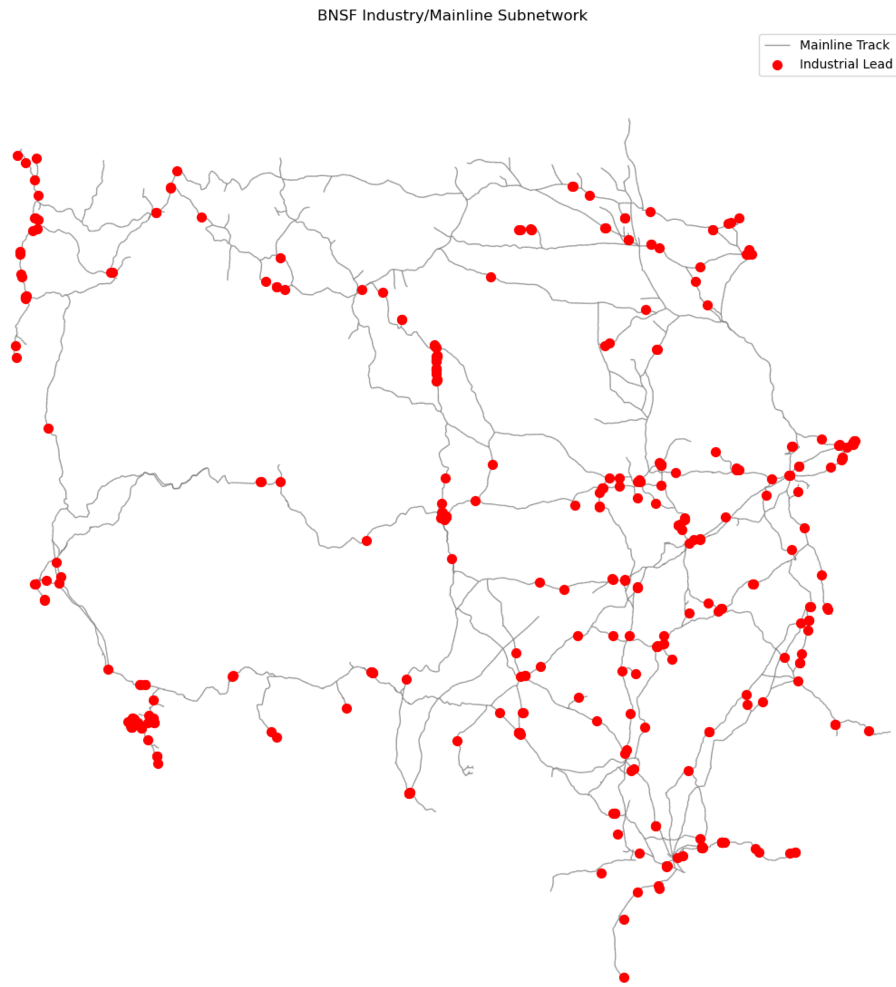

Figure 1. Example of BNSF critical nodes derived from the FRA North American Rail Network.

Figure 1 shows the BNSF industry/mainline rail subnetwork extracted from the FRA North American Rail Network. Mainline track segments are displayed as gray lines, and major industrial lead intersections are displayed as red points. These red points identify candidate critical nodes used to construct plausible industrial freight origin–destination pairs for the link criticality analysis.

This network could support standard disruption analyses. For example, the researchers could remove a link and measure how much it increases travel distance, then combine those results with the hazard-prone locations identified elsewhere in this report. In principle, this would highlight track segments where a disruption would cause the largest detours.

In practice, this approach requires realistic freight flows and routing. The researchers did not have access to true origin–destination (OD) data or information on typical commodity pathways, and that uncertainty dominates the results. The network contains many candidate nodes, so the number of possible OD pairs grows rapidly and becomes unrealistic if the researchers treat them all as equally likely.

For example, even in the BNSF subset shown above (yards plus mainline–industrial intersections), there are hundreds of nodes. Treating every node pair as a plausible OD movement produces an inflated and misleading picture of criticality. Without industry input on representative OD pairs and typical routes, the researchers could credibly estimate link-, node-, or network-level criticality, even if they overlay natural hazard exposure.

Node-based criticality assessment

From the standpoint of network science and graph theory, the researchers constructed the entire network into a graph that has nodes and links to which different performance metrics are commonly utilized to investigate the characteristics of the graph. From the viewpoint of the nodes within the graph, common approaches include:

- Degree of Centrality
    - Links connected to each node
    - This can be weighted by load or flow through nodes as well
- Betweenness Centrality
    - Fraction of shortest paths for the entire network that path through each node
- Eigenvector Weighting
    - Scoring ranking through normalized variables

Link-based criticality assessment

Instead of considering individual nodes, these metrics consider the edges themselves, or rail segments:

- Edge Betweenness
    - Fraction of shortest paths going through a link
- Edge Connectivity
    - Number of OD pairs disconnected in the event of edge removal
- Link Flow

System-based criticality assessment

This looks into systemwide metrics from a holistic perspective.

- Giant Component
    - Largest connected component in a network
- Percolation Threshold
    - Proportional of nodes or links that can fail prior to network collapse

These criticality metrics are common in network analysis, and several are relevant to rail operations. However, the main takeaway of this limitations section is simple: researchers did not have the flow and routing data needed to apply most of these metrics in a meaningful way. Without freight flow data, node- and link-level centrality measures become generic and offer limited insight. In many cases, node degree is not informative because rail nodes often connect to a single inbound and a single outbound mainline segment.

Betweenness centrality and edge betweenness can be useful, but they depend on defined origin–destination (OD) pairs to compute shortest paths. Because the researchers did not have OD data or representative routing choices, they could not calculate these metrics credibly. The same limitation applies to eigenvector-based measures: the researchers can compute the mathematical values but did not have meaningful weights (such as traffic, tonnage, or value) to make the rankings operationally relevant.

System-level measures are the most aligned with this project because they can represent how the network degrades when hazard-exposed links fail. Metrics such as the size of the giant component and percolation thresholds can quantify how connectivity and operational capacity change as links are removed. However, without OD paths and segment-level commodity flows, these results remain qualitative. The researchers can show that connectivity decreases, but they cannot translate that change into impacts on service, commodities, or cost.

With OD data and commodity values on each segment, the researchers could produce actionable criticality metrics and directly link disruption consequences to natural hazard exposure. Because those inputs are not currently available, this remains a priority for future work.

One alternative that mitigates lack of access to direct industry data is to use the Bureau of Transportation Statistics Freight Analysis Framework (FAF). FAF reports commodity flows between FAF zones often at the scale of states or large regions. The researchers could assign those flows to rail links using shortest-path or route-directness methods, recognizing that this would reduce accuracy. This provides a practical direction for future work and a pathway to approximate flow-informed criticality until more detailed OD and routing data become available.

2.1.2 Limitations - Loss of Function, Fragility and Restoration Curves to Quantify Resilience

Loss of function, fragility, and restoration curves describe what the researchers needed to estimate rail resilience link by link across the U.S. network. To build these curves, they need historical records that show when specific hazards disrupted specific track links.

If the researchers could match a hazard event to a damaged link, they could then relate hazard intensity to outcomes such as repair cost, physical damage, and outage duration. The researchers define "loss of function" as the operational impact of the event on a track segment, based on damage to track components and resulting restrictions on service.

A thermal-buckling incident illustrates the process. When a buckle occurs, the railroad must replace or repair track components and mobilize materials (ties, ballast, rail), crews, and equipment. Those requirements determine how long the segment remains out of service before operations resume. By contrast, extreme heat without a buckle may only trigger slow orders, which reduces performance without fully closing the line.

With survey responses and hazard-specific disruption records from Class I railroads, the researchers connected event characteristics (intensity, frequency, duration, and related variables) to observed outcomes. This would allow them to estimate expected outages and costs as a function of hazard magnitude.

In Section 3, the researchers demonstrated a fragility-style approach for sun kink risk. However, they still lack the loss-of-function and restoration data needed to complete resilience curves. Section 3 also relies on derailment cases where a buckle is confirmed. A full fragility and resilience framework requires exposure-based data, including events that affected track operations without producing derailments.

Based on our review of FRA Form 54C and our understanding of rail operations, the researchers expect substantial underreporting of track damage and disruption in public datasets. That missing information limits our ability to estimate standard fragility and restoration curves. Even so, Section 3 provides a reduced-form fragility method for sun kinks; producing full resilience curves will require additional industry data.

2.2 Nation-Implications on an industry survey and data acquisition from historical natural hazard events

In Task 1 (previously published), the researchers matched historical derailments from the FRA derailment dataset with multiple government natural hazard datasets. The researchers found many cases where hazard events overlapped with derailments or occurred close in time and location.

However, the researchers could not verify the full extent of damage in either physical or monetary terms. The FRA incident and derailment records include fields for equipment and accident costs, but these values can under- or overestimate actual damages. They also lack key operational details, including outage duration and the length of track affected. These gaps limit hazard-specific damage assessment for each hazard examined in this study.

2.2.1 Flooding – Pluvial and Fluvial Extreme Events

When dealing with flooding events, the research team utilized two primary datasets:

- United States Geological Survey (USGS) - High Water Marks
- National Centers for Environmental Information (NCEI)
- FRA derailment and incident database, targeting T002 (Washout/rain/slide/flood/snow/ice damage to track) and M103 (Extreme environmental condition – FLOOD)

In both flood datasets, each event is represented by a single latitude–longitude point that the researchers spatially assigned to the nearest rail segment. To support this assignment, they also used 12-digit Hydrologic Unit Code (HUC12) watersheds, which typically represent subwatersheds smaller than about 40 square miles. The HUC12 framework helps link flooding to rail segments because surface water within a watershed drains toward a shared outlet.

However, this approach requires a key assumption: before water reaches the outlet, flooding occurs roughly uniformly across the HUC12 area. The researchers used this simplification so they could scan the full U.S. rail network for potential flood exposure and then compare candidate locations to derailment and incident records.

Even when the researchers found matches, they still lacked the information needed to quantify damage. In particular, they did not know how much track was inundated, how deep flooding was at the track, or the extent of washouts. Without these details, the researchers could not build reliable relationships between flood intensity and expected damage to track and equipment.

More detailed watershed-level performance measures would also improve future projections because these datasets would allow the researchers to connect projected rainfall extremes to local hydrologic response more precisely. Overall, the same issue appears across hazards: lack of consistent, link-level disruption data that describes how natural hazards affect rail operations and infrastructure.

2.2.2 Extreme Temperature

FRA accident code T109 which pertains to track failures due to sun kinks/track alignment irregularities shows an almost one-to-one relationship with derailments in the incident and derailment database. This hazard can occur in every state and across any rail corridor; it is not limited to river basins or specific watersheds. Because it is widespread, it creates a broad safety risk and can reduce rail operating capacity.

For extreme heat, the researchers used data from NCEI records and the National Solar Radiation Database (NSRDB) and focused on National Weather Service (NWS) zones that issued excessive heat or extreme heat conditions. These datasets compile information from federal and state agencies, meteorological offices, weather stations, and news sources. Their primary purpose is to warn the public about heat exposure.

Those same public-facing heat triggers do not always translate directly to rail risk. Slow orders and thermal buckling depend on rail-specific factors, including track condition, maintenance history, rail neutral temperature, and material and structural properties. To improve verification and prediction, the researchers need operational records that show when and where railroads imposed slow orders and when buckling-related repairs occurred.

Railroads monitor rail temperatures in real time and use short-range forecasts. The researchers did not have the supporting data that explain how railroads manage heat risk. In particular, the researchers lacked information on preventive maintenance strategies and the distribution of rail neutral temperatures by segment, both of which influence when slow orders are triggered and how buckling risk evolves.

The researchers also needed clearer evidence of how railroads translate extreme-heat exposure into decision making. That includes where railroads classified assets as heat-prone, how railroads prioritize short-term actions versus long-term retrofits, and how railroads incorporate meteorological information beyond near-term forecasts, including long-term regional heat trends.

2.2.3 Land Slides

The final hazard the research team examined was landslides. Landslide risk is more localized than flooding or extreme heat, but it still poses a serious safety threat and often requires early detection to prevent derailments.

The researchers reviewed derailment cause codes and incident narratives that referenced landslides. Most of these events occurred near steep slopes in Washington, California, Arkansas, Kentucky, West Virginia, and Pennsylvania.

The researchers then extracted high-resolution elevation data from USGS (down to 1 m × 1 m) and used additional sources to compute slope and related terrain metrics. However, the researchers did not have records for smaller rockfalls or slope failures that did not cause reportable disruptions. This limited their ability to validate where landslide activity occurs and how often it affects rail operations.

Finally, the researchers used the USGS landslide inventory to identify the mapped landslide footprint as a polygon. If they were able to also obtain operational impact details, such as the affected track length, outage duration, and the volume of material removed, then the researchers could link landslide characteristics to disruption consequences. The available data allowed the researchers to estimate future landslide-related outages, expected clearance times, and other inputs needed for network-scale slope stability and resilience analysis.

3. Identify the Potential At-Risk Locations, Vulnerable Segments, and Critical Links of the U.S. Rail Network

The research team identified which natural hazard events led to derailments and incidents in the FRA database. They also developed a method to assign historical hazard events to rail segments even when those events were not recorded in the FRA incident data.

The researchers published this work as an academic paper under Task 1 of this subproject (Wang et al., 2025, Ghoreishi et al., 2025). The analysis identified nine geographic areas with the highest counts of riverine flooding, flash flooding, extreme heat, and landslides.

Under this project, the researchers provided the resulting dataset as a shapefile, found at (Ghoreishi et al., 2025). It augments the standard U.S. rail links-and-nodes network by adding hazard event attributes (including event timing and duration) as new fields. Because the baseline, North American Rail Network shapefile (BTS, 2024) already includes ownership and operating rights, each railroad can use this dataset to summarize the historical number of hazard events affecting its network.

3.1 Intensity & Frequency Mapping of Each Hazard

To extend the hazard-prone corridor analysis, the researchers estimated both hazard intensity and hazard frequency. The researchers used annual exceedance probability curves to characterize the likelihood of different event magnitudes.

For rainfall and flooding, the researchers expressed intensity in standard return-period terms (for example, 1-in-100-year or 1-in-500-year events). They then assigned these intensity and frequency estimates to events linked to FRA records and to additional hazard events that may affect track even when they are not captured in FRA reports.

For extreme heat, the researchers analyzed conditions back to the year 2000 at each location where a thermal buckling incident or derailment occurred. Using a rail temperature estimation equation and historical meteorological data, the researchers estimated rail-track temperature for each event day and recorded the event's historical percentile at that location.

3.1.1 Flooding

Flash floods are among the most destructive and deadly natural hazards, with major impacts on people and infrastructure (Chung et al., 2025; Jonkman & Kelman, 2005; Kotlyakov et al., 2013). Railroads face high exposure because their fixed alignments often pass through cut sections, low-lying corridors, steep slopes, and dense urban areas with limited drainage capacity (Ilalokhoin et al., 2025; Varra et al., 2024).

As short-duration, high-intensity rainfall events intensify, rail disruptions have increased. These events can trigger electrical failures, ballast and embankment erosion, landslides, washouts, and structural damage to tracks, tunnels, and bridges, which in turn causes delays and costly outages (Kellermann et al., 2015; Koks et al., 2019; Lamb et al., 2019; Liu, Wang, Cao, Zhu, & Yang, 2018). Because rail supports national and regional mobility for freight and intercity travel, improving resilience to localized, fast-evolving flood hazards has become a priority (Bubeck et al., 2019). Globally, floods and flash floods were associated with an estimated $56 billion in losses in 2016 alone (Munich Re, 2017). A recent review focused on railways also reports that flood events can require months of recovery and cause substantial economic impacts (Ochsner et al., 2023).

Flood impacts on rail infrastructure occur through several mechanisms. Fluvial flooding occurs when rivers overtop their banks (Costabile et al., 2021). Pluvial flooding occurs when intense rainfall exceeds surface drainage capacity (Schanze, 2018; Srivastava et al., 2023). Flash floods often combine these processes and produce fast-moving water and debris surges that damage trackside assets such as culverts, ditches, and river crossings through blockage, scour, and erosion. These processes can degrade track geometry and undermine stability (Liu, Wang, Cao, Zhu, Wu, et al., 2018).

Rail networks have limited rerouting flexibility compared with road networks, so even localized outages can propagate into broader system disruption (Mattsson & Jenelius, 2015). FRA has recognized these risks for decades. After the 1997 Amtrak derailment near Kingman, Arizona, FRA Safety Advisory 97-1 documented many flood-related washout derailments and recommended mitigation actions, with emphasis on higher-class tracks and passenger routes (Federal Railroad Administration, 1997). Despite this early recognition, consistent nationwide assessment methods and spatially integrated flood-risk data for rail remain limited.

A key barrier is limited data and measurement. National hazard records such as NCEI Storm Events often omit basic hydrometeorological details, including rainfall intensity, duration, and recurrence interval. They also typically report events as points without clear spatial footprints, which limits their direct use for rail risk evaluation (NCEI/NGDC, 2024). In parallel, the FRA accident database misses many weather-related disruptions that fall below its monetary reporting threshold. This gap hides recurring maintenance needs and slow-order burdens associated with flooding (FRA, 2024a; FRA, 2024 threshold). As a result, agencies lack a comprehensive, quantitative view of where rail assets repeatedly intersect flash-flood-prone conditions.

To address these gaps, the researchers must add hydrologic context and use observations at the right spatial and temporal scales. They were able to map events to HUC-12 subwatersheds to link flood occurrence to rail corridors using a hydrologically meaningful unit (USGS, 2024). The researchers also paired event reports with Local Climatological Data (LCD) to reconstruct storm characteristics, intensity, duration, and frequency, using annual exceedance probability (AEP) and intensity–duration–frequency (IDF) analyses. These approaches capture sub-daily extremes relevant to rail performance (National Centers for Environmental Information (NCEI), 2024; Chow et al., 1988; Serinaldi & Kilsby, 2014; Papalexiou & Koutsoyiannis, 2013).

A national spatio-temporal framework that connects these metrics to rail alignments can identify two types of exposure: acute exposure (rare, high-consequence events) and chronic exposure (frequent, moderate events). This distinction supports practical decision-making, including maintenance prioritization, targeted retrofits, drainage upgrades, and resilience-oriented planning across owners and operators (Bubeck et al., 2019; Koks et al., 2019).

Data

Historical Local Climatological Data (LCD) Sensors (1975–2024)

The researchers obtained hourly precipitation data from NOAA’s Local Climatological Data (LCD) archive for 1975–2024 (National Centers for Environmental Information (NCEI), 2024). They used these data to identify independent storm events and to compute station-specific rainfall statistics, including annual exceedance probability (AEP) and intensity–duration–frequency (IDF) curves.

The researchers first compiled 4,583 unique LCD stations across the United States, and removed stations located offshore or over water bodies to ensure spatial relevance, which reduced the set to 3,911 land-based stations. They then kept only stations that report hourly precipitation, which reduced the sample to 2,783 stations.

Next, the researchers required each station to have at least 19 complete years of hourly data. They selected this threshold to balance spatial coverage with statistical robustness. This filter reduced the dataset to 1,867 stations.

The researchers delineated discrete storm events using a minimum inter-event time (MIT) threshold applied to each station’s hourly precipitation record. They then identified extreme events using a peaks-over-threshold (POT) approach, with the threshold set to the 95th percentile of non-zero hourly rainfall at each station. The researchers fit a generalized Pareto distribution (GPD) to the exceedances to estimate rainfall intensity and frequency metrics.

To support reliable GPD fitting, the researchers required each station to have at least 30 storm events above the threshold. This reduced the number of qualifying stations from 1,867 to 1,818.

Finally, the researchers limited each HUC12 watershed to one representative LCD station to avoid over-weighting watersheds with multiple nearby sensors. For watersheds with more than one station (86 total), the researchers selected the station with the best GPD fit, defined as the lowest RMSE. Of these 86 watersheds, 81 contained two stations and 5 contained three stations. After this step, the final dataset included 1,726 stations.

Figure 2 shows the spatial distribution of the final LCD stations and the number of years of hourly precipitation data available at each site.

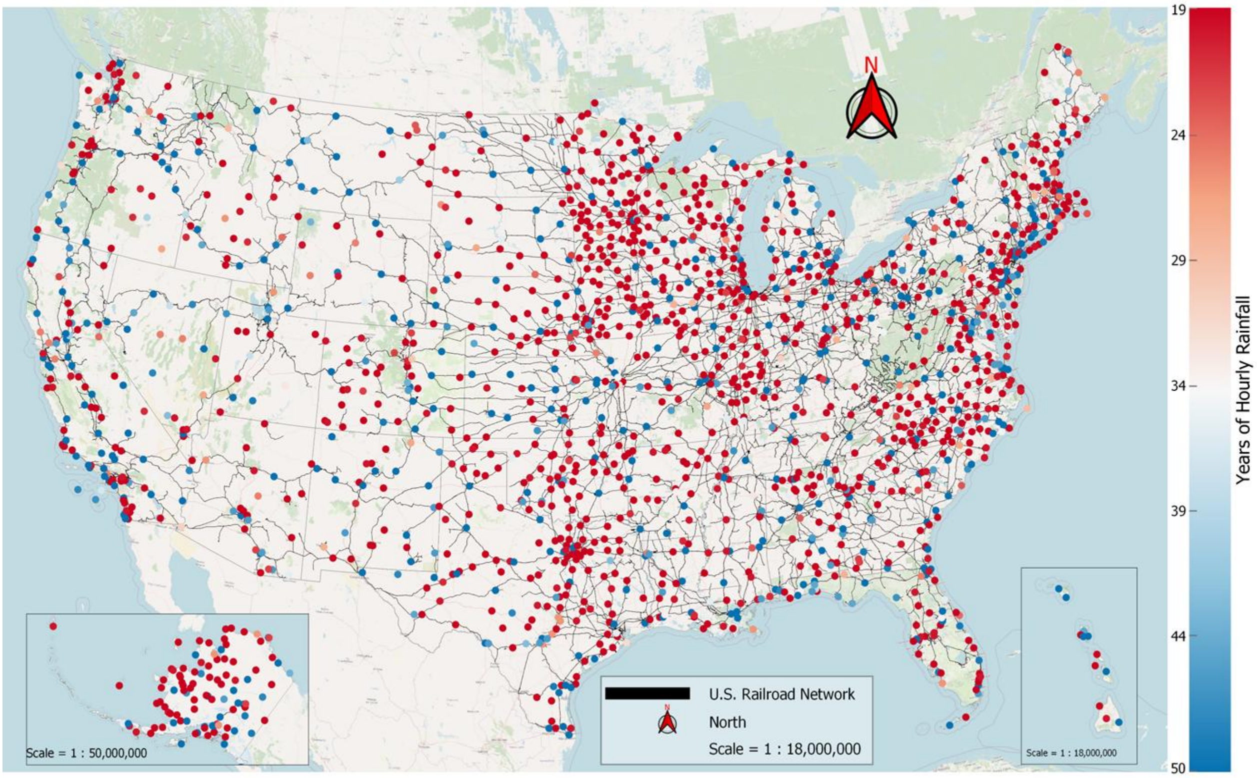

Figure 2. Spatial distribution of the final LCD stations (1975–2024) and their corresponding data availability in years.

Flash Flood Event Records: NCEI Storm Events Database

The researchers obtained historical flash flood records from the NCEI Storm Events Database, which compiles officially reported hazardous weather events across the United States (National Centers for Environmental Information, 2024). They extracted all events labeled "Flash Flood" from 2000 to 2024 to align with the period covered by the LCD hourly precipitation record. Each report includes the event date and time, magnitude (when available), a general location (typically county or warning zone), and narrative descriptions of impacts.

The researchers assigned each flash flood report to a Hydrologic Unit Code (HUC-12) watershed so that they could align event locations with LCD rainfall records and intersecting rail infrastructure segments. This watershed-based assignment provides a consistent spatial unit for linking hazards, precipitation observations, and rail corridors.

The researchers then linked the flash flood reports to observed rainfall events using the nearest LCD stations. Out of 64,271 flash flood reports, they matched 53,990 events (84%) to identifiable storms recorded at nearby LCD sensors. The researchers could not match the remaining 10,281 events (16%), most often because no nearby station reported hourly precipitation for the relevant time and location.

Figure 3 shows the spatial distribution and timing of flash flood reports from 2000 to 2024. Each point represents one event and is colored by year, which highlights regional differences and clustering in flood activity. Although the NCEI dataset provides broad national coverage, it does not include key hydrometeorological descriptors such as rainfall depth, storm duration, or return period.

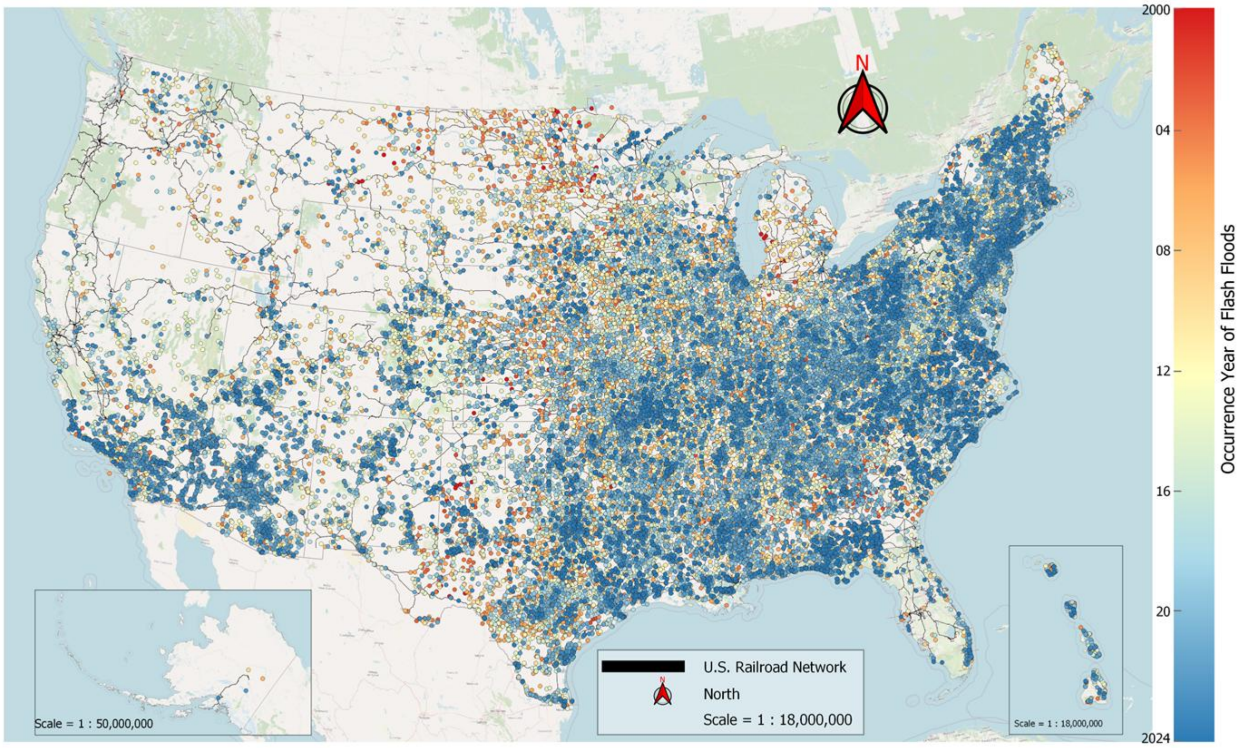

Figure 3. Historical flash flood reports from the NCEI Storm Events Database (2000–2024).

Railway Infrastructure and Incident Data:

The research team obtained geospatial data for U.S. freight and passenger rail lines from the Bureau of Transportation Statistics (BTS) National Transportation Atlas Database (NTAD) (Bureau of Transportation Statistics, 2024). The shapefiles include ownership and operational attributes. The team intersected these rail layers with the rainfall-based hazard layers to quantify exposure along the network.

The team also extracted flood-related rail disruptions from the FRA Rail Equipment Accident/Incident Database for 2000–2024. They filtered the records using cause codes T002 and M103 and retained incidents with valid locations. This process produced 113 geolocated incidents. These events are not exclusive to flash floods, but they demonstrate the broader ways flooding affects rail infrastructure.

Figure 4 shows the geographic distribution of these incidents across the U.S. rail network.

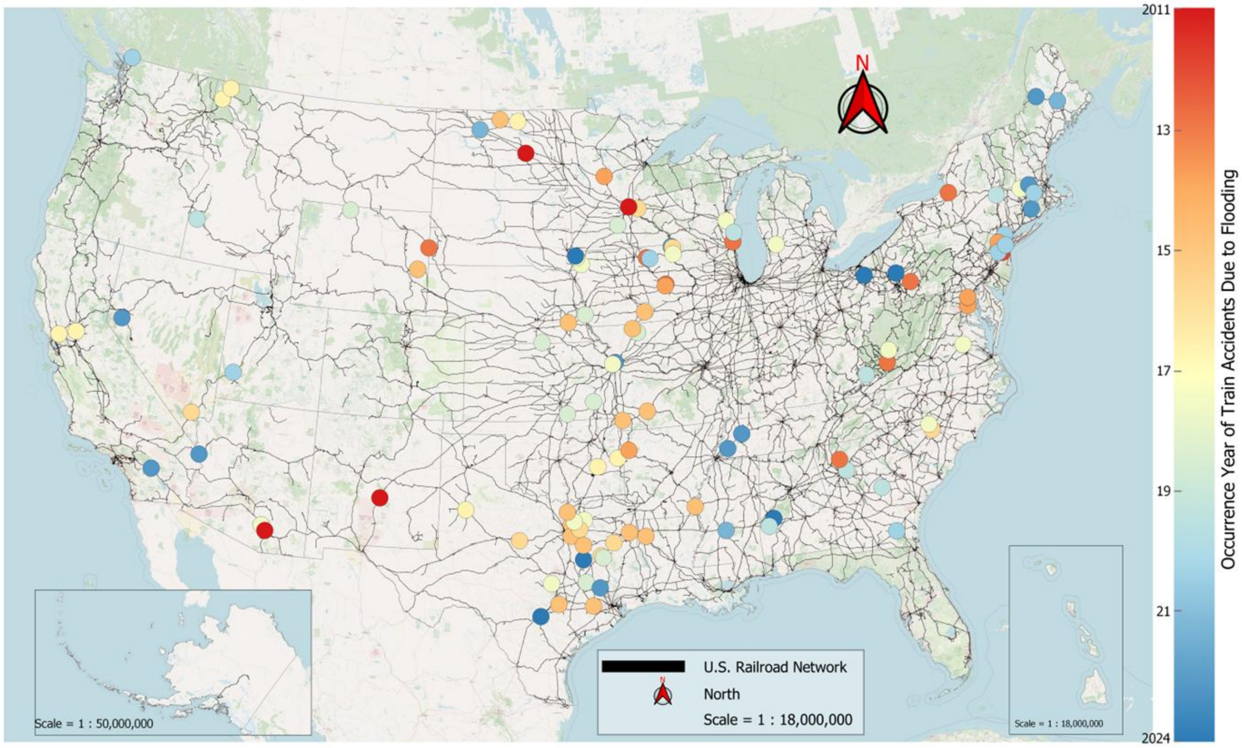

Figure 4. Flood-related rail incidents (2011–2024) from the FRA database, filtered by cause codes T002 and M103 (n=113).

Hydrologic Reference Units (HUC-12):

To enable hydrologically relevant regional aggregation, all rainfall and railway datasets were mapped to the U.S. Geological Survey's Hydrologic Unit Code level 12 (HUC-12) boundaries (U.S. Geological Survey, 2024) (Figure 5). This spatial referencing supports subwatershed-level analysis and improves the interpretability of infrastructure flood exposure within topographically consistent areas.

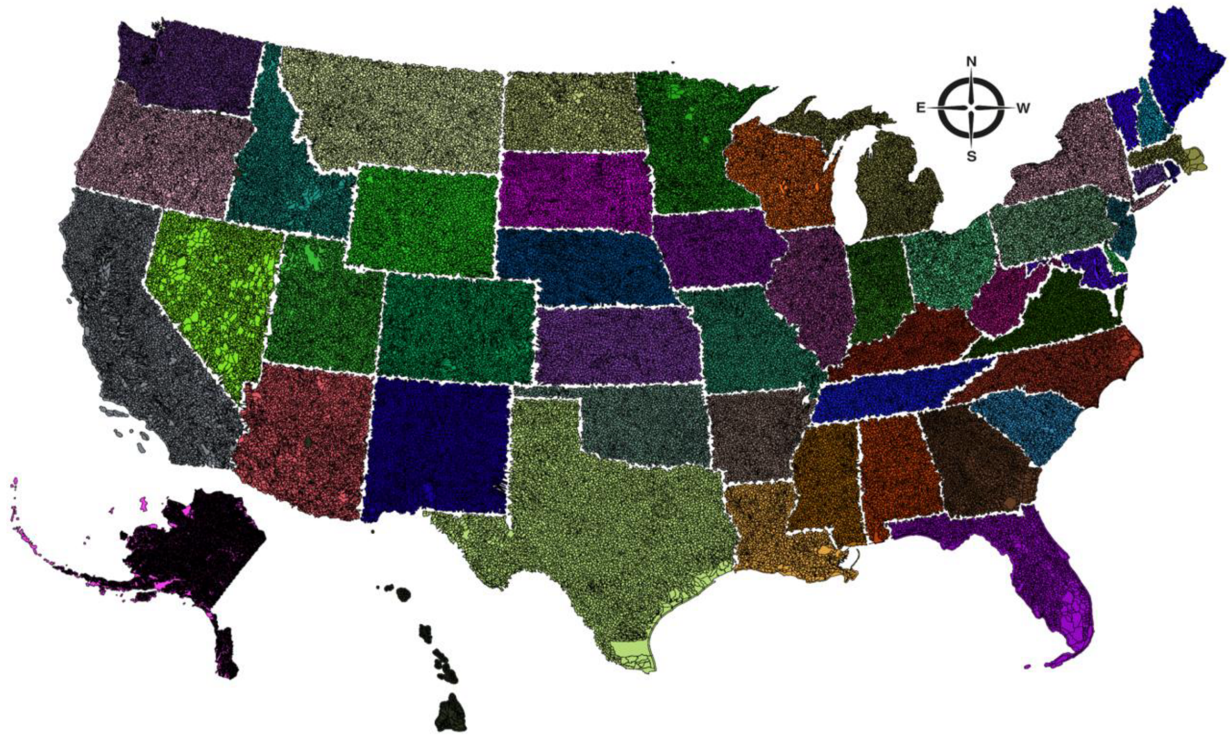

Figure 5. Hydrologic Unit Codes (HUCs) - 12-Digit

Methodology

The team developed a multi-stage framework to assess flash flood exposure along U.S. railways. The framework combines hourly rainfall observations, historical flash flood reports, and geospatial rail network data. Using NOAA Local Climatological Data (LCD), the researcher team estimate rainfall intensity, duration, and return periods. They then linked these metrics to flash flood events from the NCEI Storm Events Database and mapped the results to rail alignments to classify exposure at the segment level.

Figure 6 summarizes the four-stage workflow. It integrates rainfall and flood data, models rainfall extremes using GPD and generalized extreme value (GEV), and maps risk scores to rail segments based on exposure. Stage 1 compiles the input datasets: LCD rainfall records, NCEI flash flood reports, FRA accident and incident logs, and geospatial shapefiles for rail infrastructure and watersheds. Stage 2 identifies hydrologically independent storm events using a 6-hour minimum inter-event time (MIT), applies a peaks-over-threshold (POT) filter to isolate extreme rainfall, and assigns each storm to a watershed. Stage 3 applies extreme value methods by fitting a GPD to develop annual exceedance probability (AEP) curves and fitting GEV models to generate IDF curves, using stationary or non-stationary forms based on trend testing. Stage 4 links rainfall metrics and flood reports to intersecting rail segments and FRA incident locations, producing spatial risk maps and percentile-based flash flood risk scores that support both near-term response and long-term resilience planning.

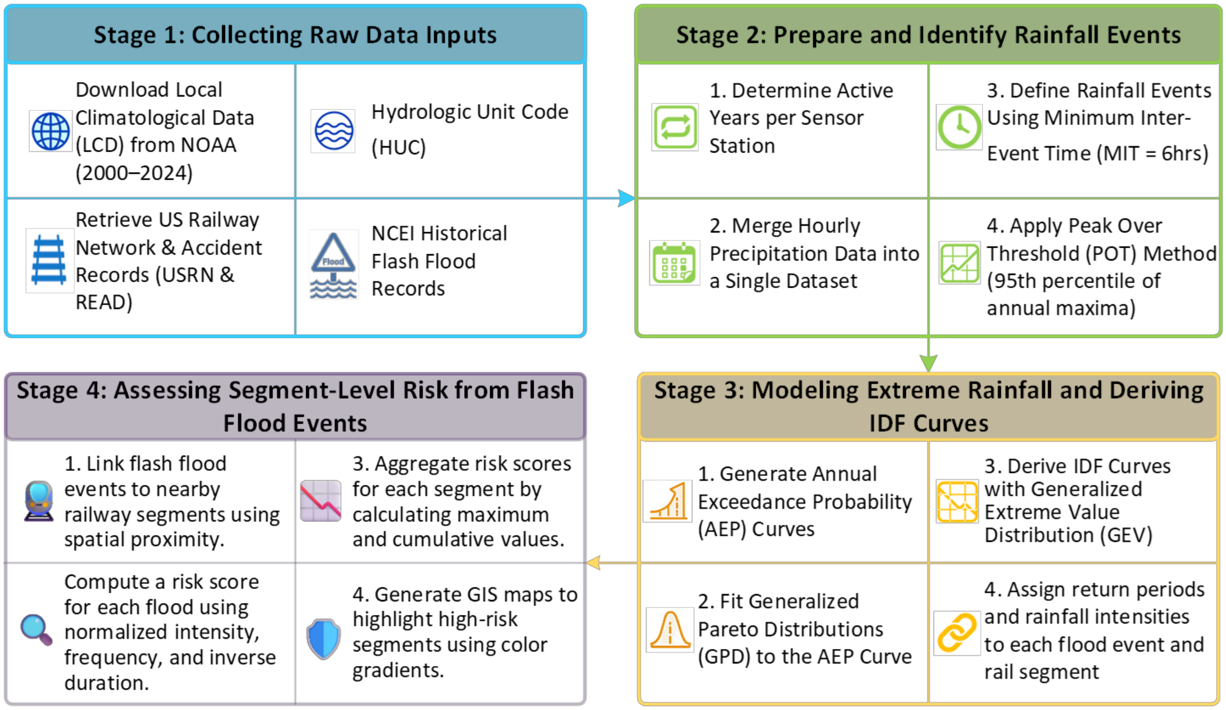

Figure 6. Four-stage framework for assessing flash flood exposure in U.S. railway infrastructure.

The following subsections describe each component of the framework in detail, including event separation, extreme value modeling, IDF generation, and risk assignment to rail segments.

Rainfall Event Identification and Event Delineation:

Identifying hydrologically independent rainfall events is a key step in extreme precipitation analysis. The research team divided each hourly LCD station record into discrete storm events using a 6-hour minimum inter-event time (MIT). Under this rule, they treat two storms as independent when at least six consecutive hours with zero rainfall separate them.

Hydrology studies commonly use a 6-hour MIT because it balances event independence with preservation of storm structure. Prior work has applied similar thresholds in semiarid settings (Brasil et al., 2022), Mediterranean basins (Medina-Cobo et al., 2016), and humid regions (Tu et al., 2023). Some studies recommend longer intervals for specific climates or objectives (e.g., 10 hours; Wang et al., 2019), but the researchers used 6 hours because it is widely accepted and fits the project's national-scale application.

Peak Over Threshold (POT) Approach for Modeling Rainfall Extremes

The research team used the peaks-over-threshold (POT) approach from extreme value theory to model short-duration, high-intensity rainfall (Ferreira & Haan, 2013; Gomes & Guillou, 2015). POT uses all storms that exceed a chosen threshold, rather than only the single annual maximum. This approach increases the amount of information available for fitting extremes and improves statistical efficiency.

The researchers set the threshold at the 95th percentile of non-zero hourly rainfall at each station; this equates to a minimum hourly threshold of 1.7 inches of rainfall. This choice captures rare events while still retaining enough exceedances for stable model fitting. Figure 7 illustrates the process for the Louis Armstrong New Orleans International Airport station (ID: 72231012916). The left panel shows the 1.7-inch threshold, which aligns with the lower bound of the annual block maxima. This threshold is shown as the red dashed line in the left side graph. The right panel shows all storms in the historical record that exceed this threshold; the black dashed line shows the threshold limit used for the GPD.

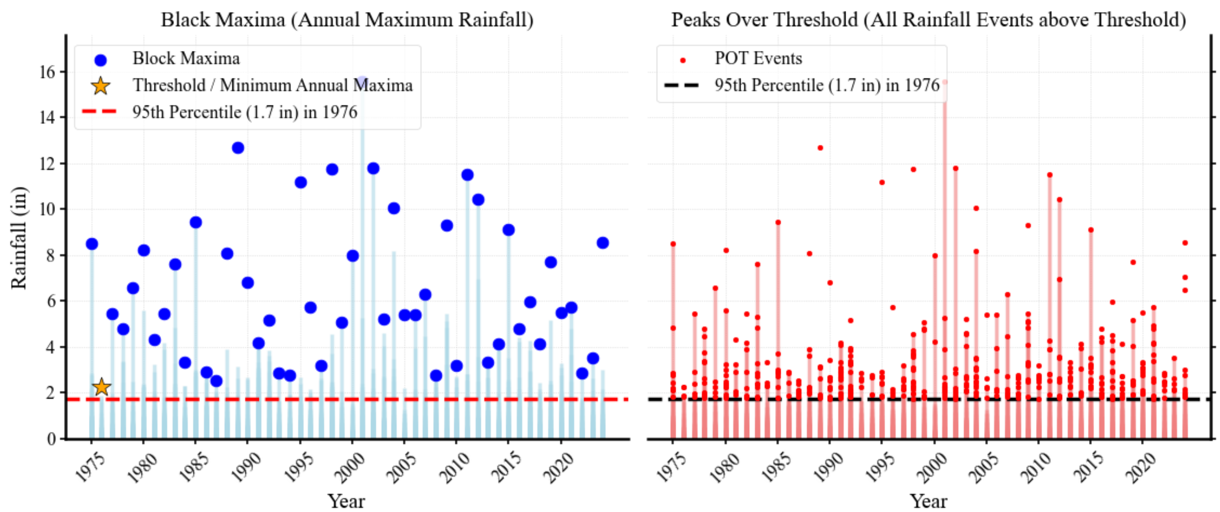

Figure 7. POT-based threshold exceedance analysis at Louis Armstrong New Orleans International Airport (ID: 72231012916).

To ensure statistical robustness of the POT analysis, stations were filtered to retain only those with a minimum of 30 independent exceedance events. This criterion helps ensure reliable estimation of the tail behavior, as recommended by previous studies on extreme precipitation modeling (Coles, 2001). Each station's exceedances above the threshold were modeled using the GPD.

Generalized Pareto Distribution for AEP Estimation and Return Period Interpretation

The GPD is a standard model for extremes above a high threshold in hydrology and environmental risk analysis (Pickands, 1975; Davison & Smith, 1990; Castillo & Hadi, 1997). Pickands (1975) showed that under the (POT) approach, in extreme value theory, threshold exceedances converge to a GPD after appropriate scaling. Researchers have used the GPD to model extremes in rainfall, streamflow, and wave height (Hosking & Wallis, 1987; Grimshaw, 1993; Castillo & Hadi, 1997).

In this study, the research team fitted a GPD to rainfall events that exceed the 95th percentile of non-zero hourly precipitation at each LCD station. This percentile-based threshold captures rare events while still providing enough exceedances for reliable parameter estimation. By modeling exceedances, the researchers estimate return levels and associated probabilities, which supports a probabilistic assessment of hydrologic extremes relevant to rail infrastructure.

The GPD probability density function is:

$$f(x;\xi,\mu,\sigma) = \{\frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{\xi}},\ for\ \xi \neq 0, \frac{1}{\sigma}ex\left(-\frac{x-\mu}{\sigma}\right),\ for\ \xi = 0, \qquad where\ x > \mu\ and\ \sigma > 0\#(1)$$

- $\mu \in R$: Location parameter, defines the lower bound (threshold) of the distribution.
- $\sigma > 0$: Scale parameter, controls the spread of the distribution.
- $\xi \in R$: Shape parameter, determines the tail behavior.

When ξ> 0, the GPD exhibits a heavy (Pareto-type) tail, allowing for unbounded extreme values, a situation often encountered in hydrologic extremes. In contrast, ξ= 0 yields the exponential distribution, representing a light-tailed case. When ξ< 0, the distribution has a finite upper bound (Weibull-type), indicating a natural physical limit to the observed extremes (Davison & Smith, 1990; Grimshaw, 1993).

This theoretical flexibility has rendered the GPD a preferred model for applications requiring accurate estimation of rare event probabilities and their associated return levels. Moreover, its applicability to a broad range of threshold exceedance phenomena, as documented by numerous studies (Beirlant et al., 2004; Castillo & Hadi, 1997; Hosking & Wallis, 1987, p. 1997), underlines its central role in modern extreme value analysis. The CDF of the Generalized Pareto Distribution (GPD), a foundational model within extreme value theory, is expressed as follows (Castillo & Hadi, 1997; Davison & Smith, 1990):

$$F(x;\xi,\mu,\sigma) = \{1-\left(1+\frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi},\ for\ \xi \neq 0, 1-ex\left(-\frac{x-\mu}{\sigma}\right),\ for\ \xi = 0, \qquad where\ x > \mu\ and\ \sigma > 0\#(2)$$

- $\mu \in R$: Location parameter, defines the lower bound (threshold) of the distribution.
- $\sigma > 0$: Scale parameter, controls the spread of the distribution.
- $\xi \in R$: Shape parameter, determines the tail behavior.

From the CDF, the AEP, defined as the probability of observing a rainfall event with magnitude exceeding x in any given year, is derived as:

$$AEP(x;\xi,\mu,\sigma) = 1 - F(x;\xi,\mu,\sigma).\#(3)$$

- $AEP(x)$: Annual Exceedance Probability, the probability that a variable exceeds a threshold x in a given year.
- $F(x;\xi,\mu,\sigma)$: Cumulative distribution function (CDF) of the Generalized Pareto Distribution.
- μ,σ,ξ : Location, scale, and shape parameters, respectively (as previously defined).

To estimate the parameters of the GPD, namely the shape (ξ), location (μ), and scale (σ), a two-stage fitting procedure was implemented. The first stage involves maximum likelihood estimation (MLE), a standard approach for statistical parameter estimation in extreme value models (Castillo & Hadi, 1997; Smith, 1984). Given a sample of n exceedances $\{x_i\}_{i=1}^{n}$, the log-likelihood function is given by:

$$log\,L(\xi,\mu,\sigma) = \sum_{i=1}^{n} \quad log\,f(x_i;\xi,\mu,\sigma)\#(4)$$

- $L(\xi,\mu,\sigma)$: Likelihood function of the Generalized Pareto Distribution (GPD).
- $f(x_i;\xi,\mu,\sigma)$: Probability density function of the GPD for observation x_i.
- x_i : Observed exceedances above threshold μ, for $i = 1,2,\ldots,n$.
- n : Sample size (number of exceedances).

where $f(x;\ \xi,\mu,\sigma)$ is the GPD probability density function. This optimization yields the initial MLE parameter estimates: $(\hat{\xi}_{\text{MLE}},\hat{\mu}_{\text{MLE}},\hat{\sigma}_{\text{MLE}})$. While MLE provides efficient and asymptotically unbiased estimates under ideal conditions, its performance may be unreliable near the distribution tails, particularly where empirical AEPs become sparse or noisy. To address this limitation, a constrained optimization routine was implemented in the second stage. This method penalizes discrepancies between model-derived and empirical AEP values, particularly at the upper and lower boundaries of the observed data (Castillo & Hadi, 1997; Hosking & Wallis, 1987). The objective function minimized during this stage is defined as:

$$J(\xi,\mu,\sigma) = \sum_{i=1}^{n} \quad \left[\left(1 - F(x_i;\xi,\mu,\sigma)\right) - y_i\right]^2 + \left[\left(1 - F(x_1;\xi,\mu,\sigma)\right) - y_1\right]^2\#(5)$$
$$+ \left[\left(1 - F(x_n;\xi,\mu,\sigma)\right) - y_n\right]^2$$

- $J(\xi,\mu,\sigma)$: Objective (loss) function.
- $F(x_i;\xi,\mu,\sigma)$: Cumulative distribution function (CDF) of the Generalized Pareto Distribution at x_i.
- $1 - F(x_i;\xi,\mu,\sigma)$: Survival function (exceedance probability) at x_i.
- y_i : Empirical exceedance probability associated with x_i.
- x_i : Ordered sample values, $i = 1,2,\ldots,n$.

Where y_i is the empirical AEP associated with the exceedance value x_i, and x_1 and x_n are the minimum and maximum exceedances in the sample, respectively. The optimization was initialized with the MLE-derived parameters and solved using the L-BFGS-B algorithm, a quasi-Newton method capable of incorporating box constraints to preserve parameter validity (Zhu et al., 1997). Upon convergence, the final parameter estimates are recorded as: $(\hat{\xi}_{\text{opt}}, \hat{\mu}_{\text{opt}}, \hat{\sigma}_{\text{opt}})$. To evaluate the fit, the log-likelihood of the optimized parameters was recorded, along with the root mean squared error (RMSE) between the predicted and empirical AEP values. The RMSE metric is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left[\left(1 - F(x_i; \xi^{\wedge}, \mu^{\wedge}, \sigma^{\wedge})\right) - y_i\right]^2}. \#(6)$$

- RMSE: Root Mean Square Error, a measure of goodness-of-fit.
- $F(x_i; \xi^{\wedge}, \mu^{\wedge}, \sigma^{\wedge})$: Estimated CDF of the Generalized Pareto Distribution at observation x_i.
- $1 - F(x_i; \xi^{\wedge}, \mu^{\wedge}, \sigma^{\wedge})$: Estimated survival function (exceedance probability).
- y_i : Empirical exceedance probability for x_i.
- n : Number of observations.
- $\xi^{\wedge}, \mu^{\wedge}, \sigma^{\wedge}$: Maximum Likelihood Estimates (MLEs) of the shape, location, and scale parameters.

The researchers fit the GPD in two steps using maximum likelihood estimation (MLE) followed by penalty-based optimization. This approach improves the fit to the observed exceedances while also stabilizing the tail behavior. A reliable tail fit is essential because return levels and risk estimates depend on the upper tail of the distribution.

Figure 8 illustrates the results for the Louis Armstrong New Orleans International Airport station. The figure shows the fitted GPD curve, the resulting annual exceedance probability (AEP) and return-period estimates, and the location of the sensor used in the analysis.

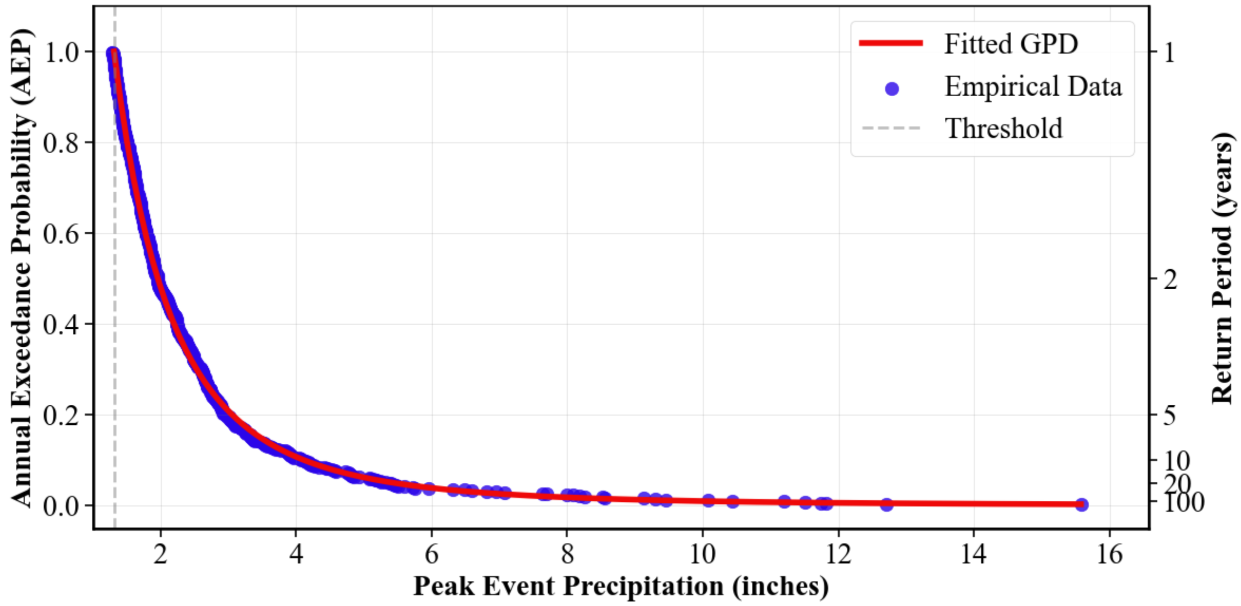

Figure 8. Illustrative GPD-AEP fitting for rainfall extremes at Louis Armstrong New Orleans International Airport, Louisiana (Sensor ID: 72231012916).

To produce defensible return-period estimates, the researchers limited how far they extrapolate beyond the observed record at each station. Station record lengths range from 19 to 50 years, and estimates far beyond these lengths become highly uncertain. Prior studies and guidance consistently warn that uncertainty grows quickly when return levels extend past the available record, and they recommend restricting interpretation to roughly two to three times the observation period (Benson, 1963; Council, 1975; Stedinger et al., 1993; Katz et al., 2002; American Society of Civil Engineers (ASCE), 1982). Other work also shows that models that fit well for common events can produce biased estimates when extrapolated to long return periods, and that key parameters can be unstable under short records (Landwehr et al., 1978, 1980; Data (IACWD), 1982; Griffis & Stedinger, 2007; Reis & Stedinger, 2005). Recent large-ensemble studies further confirm that estimates such as 1-in-100-year events can vary substantially when inferred from only a few decades of data (Wilby et al., 2023). The World Meteorological Organization (WMO) also advises flagging long return periods as highly uncertain when records are shorter than 30 years (Organization, 2009).

Based on this guidance, the researchers apply the following limits when interpreting return periods in this study:

- For stations with 19–30 years of data: return periods capped at 50–100 years
- For stations with 30–50 years of data: return periods capped at 100–150 years

The researchers keep return-period estimates that exceed the predefined caps in the dataset for completeness, but they did not interpret them because extrapolations in that range carry large uncertainty. This practice follows standard guidance in hydrologic and climate extremes analysis.

They also retain storms with precipitation below the GPD threshold (i.e., below the threshold level μ). For these events, they assigned a return period of 1 year. This conservative assignment reflects that these events are common and have a high probability of occurrence.

To avoid unrealistic behavior in the upper tail of the fitted distribution, the researchers cap return periods at three times the length of the observational record for each station. For consistency in reporting, they then round all return-period estimates to the nearest whole year.

Generating Historical IDF Curves with Generalized Extreme Value Distribution under Stationarity and Non-Stationarity

The researchers generated historical intensity–duration–frequency (IDF) curves from LCD station data to characterize how rainfall extremes vary across space and time. The researchers used a block-maxima approach rather than relying solely on NOAA PFDS curves, because the LCD record provides high-resolution observations that can better support infrastructure-focused assessments (National Oceanic and Atmospheric Administration (NOAA), 2024).

For each station, the research team aggregated hourly precipitation into rolling accumulation windows (for example, 1 to 24 hours). For each duration (d), they then identified the maximum accumulated precipitation depth in each year. This produced an annual maxima series for every duration (d) $\left\{X_d^{(y)}\right\}_{y=1}^{n}$ and was used to fit the GEV distribution.

The researchers tested each station's annual maximum precipitation series for a monotonic trend using the Mann–Kendall test (Kendall, 1975; Yue & Wang, 2004). The Mann-Kendall test is non-parametric and does not assume a specific distribution, which makes it well suited for climatological trend detection. Following Sam (2023), they used a significance level of (\alpha = 0.05).

Across 1,489 LCD stations, the Mann-Kendall test classified 459 series as non-stationary and 1,040 as stationary. These results were used to choose the IDF model form. We fit a stationary GEV model to stations classified as stationary, and we fit a time-varying (non-stationary) GEV model to stations classified as non-stationary.

The generalized extreme value (GEV) distribution provides the standard model for block maxima. Fisher and Tippett (1928) and Gnedenko (1943) established the limiting distribution for maxima, and Jenkinson (1955) provided the common parameterization. The GEV cumulative distribution function (CDF) is:

$$F(z;\xi,\mu,\sigma) = \{exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},\xi \neq 0\ exp\left\{-exp\left(-\frac{z-\mu}{\sigma}\right)\right\},\xi = 0\ \#(7)$$

- $\mu \in R$: Location parameter, shifts the distribution along the z-axis.
- $\sigma > 0$: Scale parameter, controls the spread of the distribution.
- $\xi \in R$: Shape parameter, determines the tail behavior.

Here, z denotes the precipitation depth, μ is the location parameter, $\sigma > 0$ the scale parameter, and ξ the shape parameter (Coles, 2001). The shape parameter governs the tail behavior: Fréchet ($\xi > 0$), Gumbel ($\xi = 0$), and Weibull ($\xi < 0$). For stationary stations, the parameters (μ, σ, ξ_i) were assumed constant and estimated via maximum likelihood estimation (MLE). The log-likelihood function for n observations $\{x_i\}_{i=1}^{n}$ is:

$$log\, L(\mu, \sigma, \xi) = -n log\, \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \quad log\left(1 + \xi \frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^{n} \quad \left(1 + \xi \frac{x_i - \mu}{\sigma}\right)^{-1/\xi} \#(8)$$

- $log\, L(\mu, \sigma, \xi)$: Log-likelihood function of the Generalized Extreme Value (GEV) distribution.
- n : Sample size (number of block maxima observations).
- x_i : Observed block maxima, for $i = 1,2, \ldots, n$.

subject to:

$1 + \xi \frac{x_i - \mu}{\sigma} > 0$ ∀i. MLE was performed using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization algorithm (Nocedal & Wright, 2006). For non-stationary stations, time-dependent parameterizations were adopted: the GEV parameters μ(t) and σ(t) were allowed to evolve linearly with time t, while ξ remained fixed. The four candidate GEV models are defined in Table 1: Candidate GEV model structures for IDF estimation. The Mann-Kendall test is essential to determine whether extreme rainfall behavior has shifted over time, which informs the selection of appropriate statistical models for frequency analysis.

Table 2. Candidate GEV model structures for IDF estimation.

Model Type	Parameterization	Remarks
GEV-0 (Stationary)	$\mu(t) = \mu, \sigma(t) = \sigma, \xi(t) = \xi$	No temporal trends
GEV-I	$\mu(t) = \mu_0 + \mu_1 t, \sigma(t) = \sigma, \xi(t) = \xi$	Trend in location only
GEV-II	$\mu(t) = \mu, \sigma(t) = \sigma_0 + \sigma_1 t, \xi(t) = \xi$	Trend in scale only
GEV-III	$\mu(t) = \mu_0 + \mu_1 t, \sigma(t) = \sigma_0 + \sigma_1 t, \xi(t) = \xi$	Trend in location and scale

To identify the most suitable non-stationary GEV models, we employ the corrected Akaike Information Criterion (AICc), which adjusts for model complexity and small sample size (Hurvich & Tsai, 1989). This method penalizes models with more parameters to reduce the risk of overfitting while still rewarding goodness of fit (Katz, 2013). The standard AIC is defined as:

$$AIC(k) = 2 \cdot nllh(k) + 2k \#(9)$$

- AIC (k) : Akaike Information Criterion for a model with k estimated parameters.
- $nllh(k)$: Negative log-likelihood of the fitted model.
- k : Number of estimated model parameters.

where nllh(k) is the minimized negative log-likelihood of the model, and k is the number of estimated parameters. For datasets with limited sample sizes, the corrected form, AICc is used:

$$AICc(k) = AIC(k) + \frac{2k(k+1)}{n-k-1} \#(10)$$

- $AICc(k)$: Corrected Akaike Information Criterion, adjusted for small sample sizes.
- $AIC(k)$: Standard AIC value.
- k : Number of estimated parameters.
- n : Sample size.

where n is the sample size. To ensure consistency across durations, a distinct approach was followed based on station classification. For stationary stations, the GEV-0 model was used exclusively. For non-stationary stations, the most frequently selected best-fit model (i.e., the one with the lowest AICc across durations) was chosen for the entire station. This approach avoids mixing model types across durations, thereby preserving the coherence of the resulting IDF curves while balancing model fidelity and simplicity.

The return level for a given return period T is derived by inverting the GEV CDF, as shown in Equation 11 (Cheng et al., 2014; Coles, 2001). In this context, z_t(t) directly represents the rainfall intensity for a duration d and return period T:

$$z_T(t) = \{\mu(t) + \frac{\sigma(t)}{\xi}\left\{\left[-ln\left(1-\frac{1}{T}\right)\right]^{-\xi} - 1\right\}, \xi \neq 0 \; \mu(t) - \sigma(t) ln\left(-ln\left(1-\frac{1}{T}\right)\right), \xi = 0 \#(11)$$

- $z_T(t)$: T-year return level at time t under the GEV model.
- $\mu(t)$: Time-varying location parameter.
- $\sigma(t)$: Time-varying scale parameter.

- ξ : Shape parameter of the GEV distribution.
- τ : Return period (e.g., 50-year, 100-year).

These IDF values support the representation of evolving rainfall intensities under both stationary and non-stationary assumptions. Figure 9 shows the IDF results for the Louis Armstrong New Orleans International Airport station, with intensities across 1-to-83-hour durations and 5-to-150-year return periods. The sharp drop at shorter durations and wider spread of rare events underscore their relevance for infrastructure design.

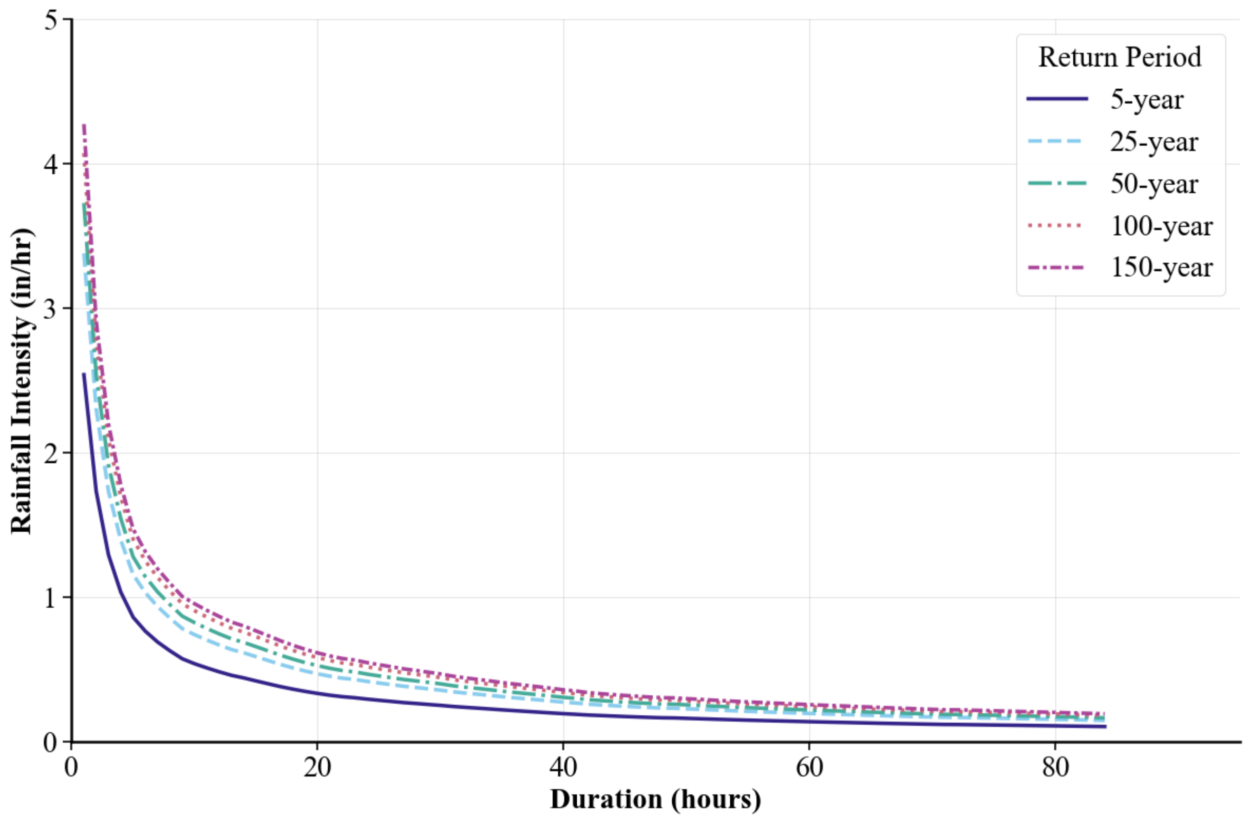

Figure 9. Intensity-Duration-Frequency (IDF) curves for the Louis Armstrong New Orleans International Airport station under stationary assumptions.

Flash Flood Risk Score Calculation:

We developed a composite risk score to compare the relative severity of flash flood events. The score combines three rainfall characteristics, intensity, duration, and return period, into a single index that we can compare across years, locations, and rail segments.

The score uses rainfall intensity I(inches per hour), event duration D(hours), and return period F(years), which represents how rare the event is. We normalize each variable to the range [1]so that no single term dominates due to units or magnitude. We include duration as (D_{norm})so that shorter events receive higher weight, because short-duration storms are typically more abrupt and can produce flash flooding.

We compute the flash flood risk score as:

$$Risk\ Score\ = I_{norm} + F_{norm} + (1 - D_{norm})\#(12)$$

- Risk Score : Composite hazard risk index.
- I_{norm} : Normalized hazard intensity.
- F_{norm} : Normalized hazard frequency.
- D_{norm} : Normalized distance/exposure factor (risk decreases as distance increases).

This formulation produces scores from 0 to 3. Higher scores indicate more hazardous events with higher rainfall intensity, shorter duration, and rarer recurrence. The score provides a consistent way to compare flood severity across locations and time periods.

Two complementary metrics were used to identify rail segments with the highest historical flash flood exposure. First, the researchers calculated the maximum risk score at each segment. This captures the single most severe event associated with that segment and helps flag safety-critical locations.

Second, they calculated the cumulative risk score by summing risk scores across all events linked to each segment. This metric highlights segments that experience repeated flooding, even if individual events are moderate.

Together, the maximum and cumulative metrics capture both acute risk (rare, high-consequence events) and chronic risk (frequent, moderate events) across the rail network.

Mapping Rainfall-Based Flood Hazard to Railway Infrastructure

The research team quantified extreme-rainfall exposure by linking each rail segment to a Hydrologic Unit Code 12 (HUC-12) watershed. For each HUC-12, we selected the nearest LCD station with sufficient data and fitted rainfall statistics. We then used that station's AEP and return-period curves to assign rainfall intensity metrics to all rail segments within the same watershed.

This station-driven approach provides a consistent, localized estimate of the probability that a rail segment will experience high-intensity rainfall over specific durations. The researchers mapped the resulting hazard indicators, intensity, duration, and return period, onto the national rail network using GIS methods. By transferring LCD metrics to rail lines through shared HUC-12 boundaries, we preserved hydrologic consistency across the network.

This mapping identifies rail segments exposed to either frequent storms or rare, high-severity events. The researchers combined these indicators into a national exposure map that summarizes flood-related risk patterns across U.S. railway segments. The map provides a practical tool for screening, prioritizing high-risk corridors, and supporting targeted risk assessment, infrastructure adaptation, and emergency planning.

Results

Rainfall Characteristics of NCEI Historical Flash Floods

Linking NCEI flash flood reports to LCD rainfall time series allowed the researchers to characterize the hydrometeorological conditions during each event. For every matched event, they extracted storm duration and maximum rainfall intensity. They then estimated the return period of each event using station-specific annual exceedance probability (AEP) curves, which quantify how rare each storm is relative to the local historical rainfall record.

Table 2: Summary statistics for rainfall intensity, storm duration, and return period of LCD-linked NCEI flash flood events (n = 53,990) presents the statistical distributions of key rainfall attributes across all 53,990 linked events. The rainfall intensity distribution exhibits a pronounced right-skewed pattern characteristic of extreme precipitation statistics. The majority of flash flood events (84.8%) were associated with moderate rainfall intensities between 0.01--1.0 in/hr, with the highest frequency occurring in the lowest intensity bin (0.01--0.25 in/hr, 52.8%, n = 28,493). This finding challenges the common assumption that flash floods require exceptionally high rainfall rates, instead demonstrating that local hydrological and topographical factors often amplify the impact of moderate precipitation. The exponential decay pattern is evident across higher intensity bins: 32.0%(n = 17,302) in the 0.25--0.50 in/hr range, 9.6%(n = 5,190) in the 0.50--0.75 in/hr range, and 3.3%(n = 1,792) in the 0.75--1.0 in/hr category. High-intensity events remain rare, with only 2.2%(n = 1,212) of events exceeding 1.00 in/hr.

Storm duration analysis reveals a pronounced bias toward short-duration precipitation events in flash flood generation. The temporal distribution demonstrates that 35.3% of events(n = 19,038) occurred within 9 hours, with peak frequency concentrated in the 5--9 hour window (15.2%, n = 8,216), followed by the 1--3 hour range (11.2%, n = 6,032). This pattern reflects the rapid hydrological response characteristic of flash flooding, where intense precipitation over brief periods overwhelms natural drainage capacity. However, the data also revealed a notable secondary mode in extended-duration events exceeding 48 hours (8.2%, n = 4,412), likely representing persistent weather systems or multi-day precipitation sequences that contribute to flash flood risk through cumulative soil saturation and reduced infiltration capacity. Moderate durations of 9 to 24 hours represent the largest category (35.3%, n = 19,075), indicating that many flash floods result from sustained precipitation events of intermediate duration.

Return period analysis using AEP curves reveals that flash floods span a wide spectrum of meteorological extremity. The distribution demonstrates a strongly right-skewed pattern, with nearly three-fifths of all events (58.6%, n = 31,663) corresponding to relatively frequent precipitation occurrences with 1-to-7-year return periods. This finding indicates that flash flooding often results from rainfall events that are not statistically exceptional within the local climatological context. However, a substantial proportion of events exceeded higher return period thresholds: 19.0% (n = 10,264) had return periods between 7 to 25 years, 9.6%(n = 5,207) exceeded 25 to 55 year levels, and 8.9% (n = 4,825) fell within the 55 to 103 year range. Particularly noteworthy is the presence of rarer extremes: 1.3% (n = 687) of events with return periods between 103 to 139 years, and 2.5% (n = 1,344) exceeding 139-year return periods.

These findings demonstrate that while many flash floods result from common meteorological conditions, a significant subset involves truly exceptional precipitation extremes that challenge infrastructure design standards and emergency preparedness protocols.

Table 3. Summary statistics for rainfall intensity, storm duration, and return period of LCD-linked NCEI flash flood events (n = 53,990).

Category	Range	Count	Percent (%)
Rainfall Intensity (in/hr)	0.01-0.25	28,493	52.8
	0.25-0.50	17,302	32.0
	0.50-0.75	5,190	9.6
	0.75-1.00	1,792	3.3
	>1.25	1,212	2.2
Storm Duration (hr)	1-3	6,032	11.2
	3-5	4,790	8.9
	5-9	8,216	15.2
	9-24	19,075	35.3
	24-48	11,465	21.2
	>48	4,412	8.2
Return Period (yr)	1-7	31,663	58.6
	7-25	10,264	19
	25-55	5,207	9.6
	55-103	4,825	8.9
	103-139	687	1.3
	>139	1,344	2.5

Spatial patterns in flash flood rainfall characteristics vary widely across the continental United States. Figure 10 maps rainfall intensity and shows that flash floods occur under a broad range of rainfall rates, with clear regional differences where low-, moderate-, and high-intensity events cluster.

Figure 11 maps storm duration and shows that flash floods are associated with both short and long storms. As duration increases, spatial distribution shifts, which indicates that different regions experience different types of flood-producing storms.

Figure 12 maps return period and shows that flash floods include both common events and rare events. The spatial patterns also change across return-period classes, which suggests that event rarity is not uniform across the country.

Taken together, Figures 10 to 12 show that flash floods are widespread, but their intensity, duration, and rarity vary systematically by region.

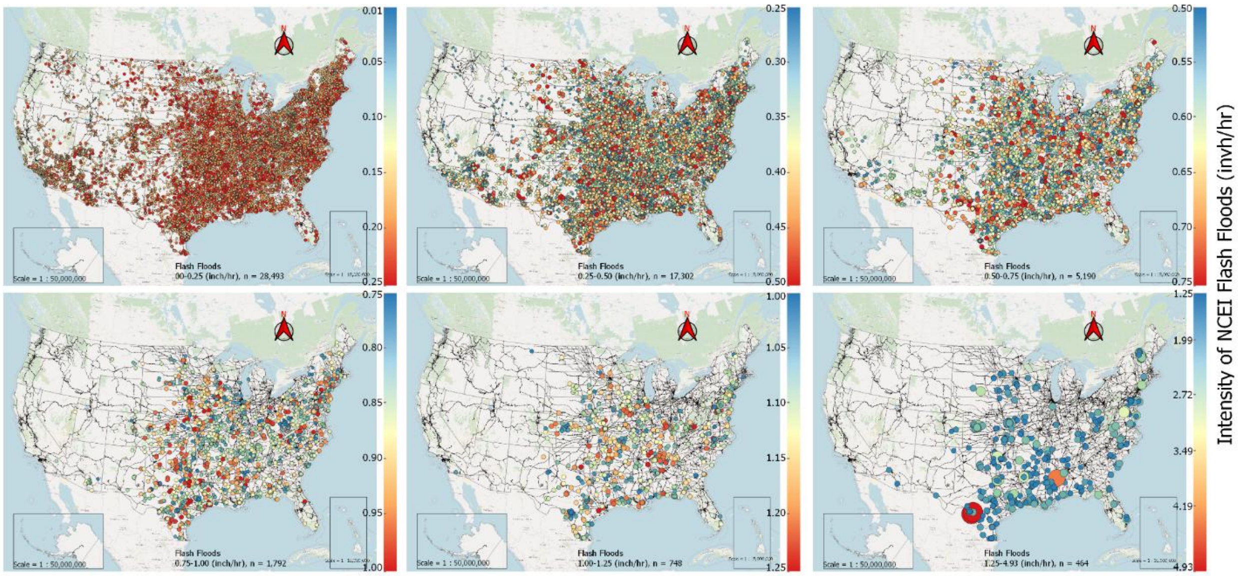

Figure 10. Spatial distribution of storm durations (hours) associated with NCEI flash flood events (2000-2024).

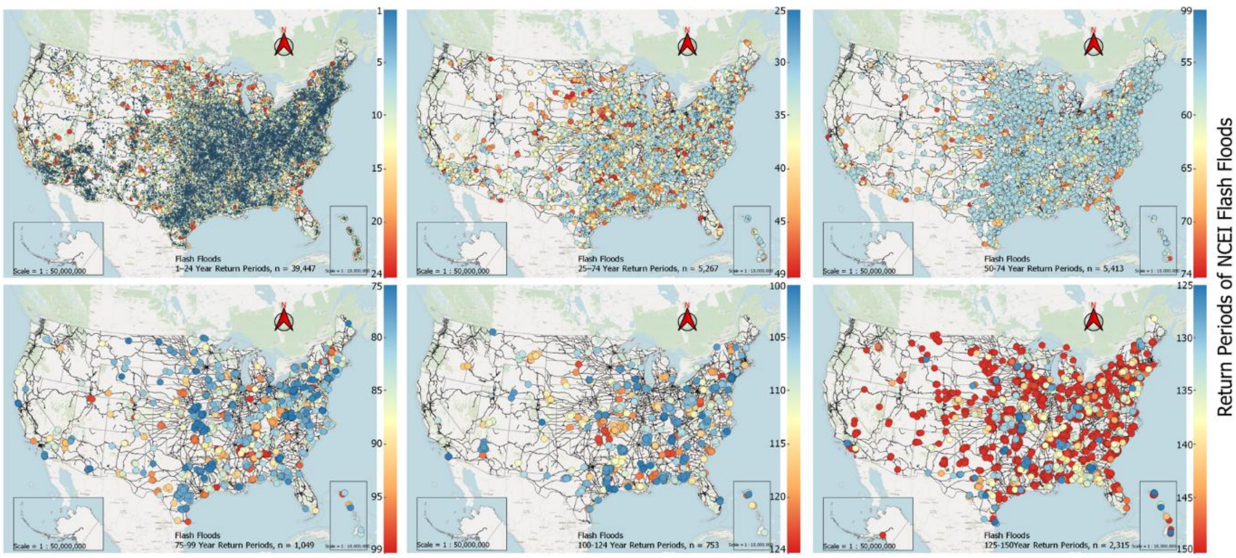

Figure 11. Spatial distribution of return periods (years) for NCEI flash flood events based on AEP curves (2000-2024).

These distributions inform both flash flood risk assessment and infrastructure vulnerability evaluation. Most events fall in the moderate-intensity, short-duration range, which suggests that local drainage, terrain, and watershed response often drive flash flood impacts as much as rainfall intensity itself.

At the same time, the historical record includes truly extreme events. These rare storms can produce high-consequence failures, so infrastructure design and retrofit planning must account for low-probability scenarios. The geographic clustering of the most extreme events also helps identify where targeted hardening and drainage upgrades are likely to deliver the greatest benefit.

Flash Flood Risk Scoring and Spatial Distribution of U.S. Rail Segments:

We quantified flood exposure at the rail-segment level using two metrics derived from the LCD-linked event mapping: (1) the maximum return period associated with each segment and (2) cumulative exposure based on all events linked to that segment. We then classified segments into risk categories using percentile thresholds calculated from the national distribution of scores. Using the same percentile-based approach for both metrics ensures consistent comparisons across the network.

We categorized rail segments as:

- No Risk: No observed flash flood event linked to the segment (score = 0).
- Low Risk: Risk scores below the 90th percentile (Maximum: 0.458 to 1.901; Cumulative: 0.458 to 18.335).
- Moderate Risk: Risk scores between the 90th and 99th percentiles (Maximum: 1.901 to2.056; Cumulative: 18.335 to 46.499).
- High Risk: Risk scores in the top 1% of the national distribution (Maximum: ≥2.056; Cumulative: ≥46.499).

The researchers created spatial maps to show how maximum and cumulative flash flood risk varies across the continental United States. The maximum risk map (Figure 13) highlights rail segments that have experienced at least one extreme event. Segments are classified using percentile thresholds. Maximum risk scores range from 0.458 to 2.400. Segments above the 99th percentile (≥ 2.056) represent the highest-risk locations based on single-event severity.

Figure 13 identifies several high-risk corridors (numbered on the map): 1) Colby, KS; 2) Del Rio, TX; 3) Tuscaloosa, AL; 4) St. Joseph, MO; 5) Lisbon, NH; 6) St. Johnsbury, VT; 7) Washington–Arlington, DC and Waldorf, MD; 8) New Orleans, LA; 9) Roswell, NM; and 10) Toms River, NJ. These corridors are strong candidates for targeted monitoring and resilience planning.

The cumulative exposure map (Figure 14) highlights rail segments that experienced repeated moderate-to-high rainfall events during the 24-year study period. This metric captures chronic flood exposure. Cumulative scores range from 0.458 to 80.508. Segments above the high-risk threshold (≥ 46.499) cluster in the Southeast, with additional corridors in the Northeast and Midwest.

Figure 14 highlights several high-risk locations (numbered on the map): 1) Columbia, SC; 2) New Orleans, LA; 3) Springfield, MO; 4) New York, NY; 5) Jackson, MS; 6) Hattiesburg, MS; 7) Washington, DC; and 8) Pittsburgh, PA. These corridors indicate chronic exposure and may warrant long-term mitigation and resilience investments.

The two maps capture different dimensions of flood risk. The maximum score flags segments vulnerable to rare, high-consequence events. The cumulative score identifies segments that experience frequent flooding that can degrade infrastructure over time. Scores represent the sum of all historical event exceedances, ranging from 0.458 to 80.508. Classification follows percentile thresholds with high-risk segments (≥46.499) indicating persistent flood exposure patterns. Gray segments had no linked flash flood events during the study period (2000 to 2024). Used together, they support two complementary decisions: the maximum score supports emergency planning and extreme-event response, while the cumulative score supports long-term asset management and resilience planning.

Segments that rank highly under both metrics represent the highest priorities for inspection, monitoring, and targeted upgrades. The percentile-based thresholds provide a consistent way to compare risk across regions with different climates and rainfall regimes, which supports national-scale prioritization and resource allocation.

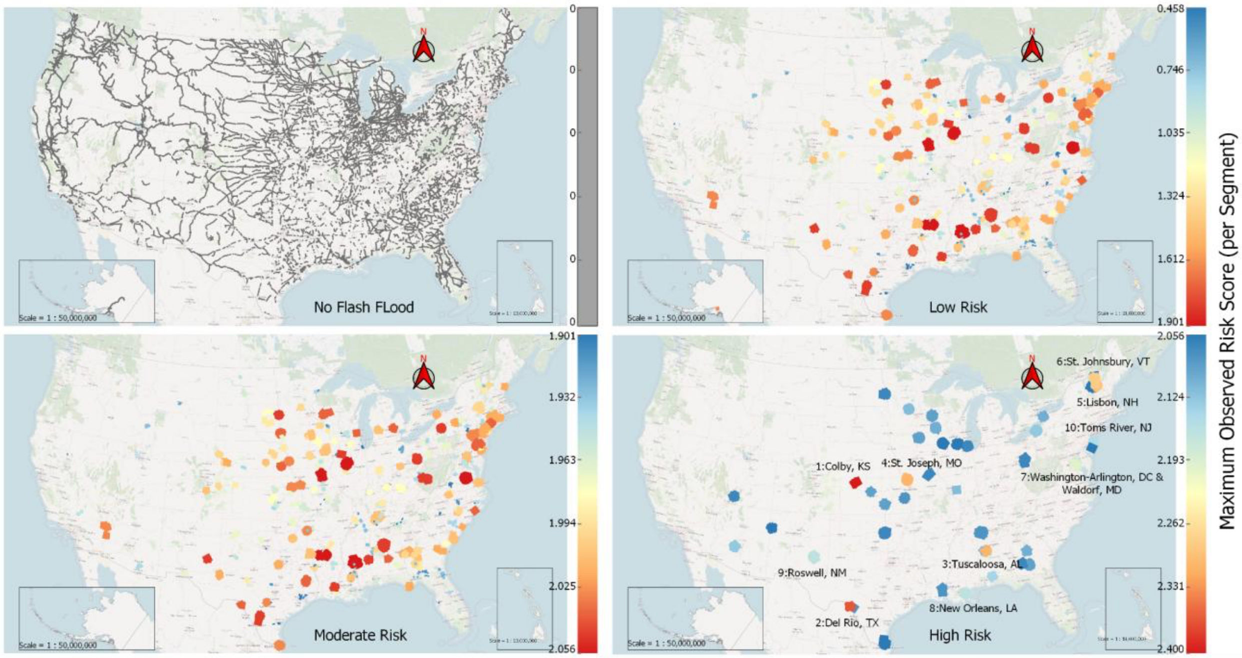

Figure 12. Spatial distribution of maximum flash flood risk scores across U.S. rail segments.

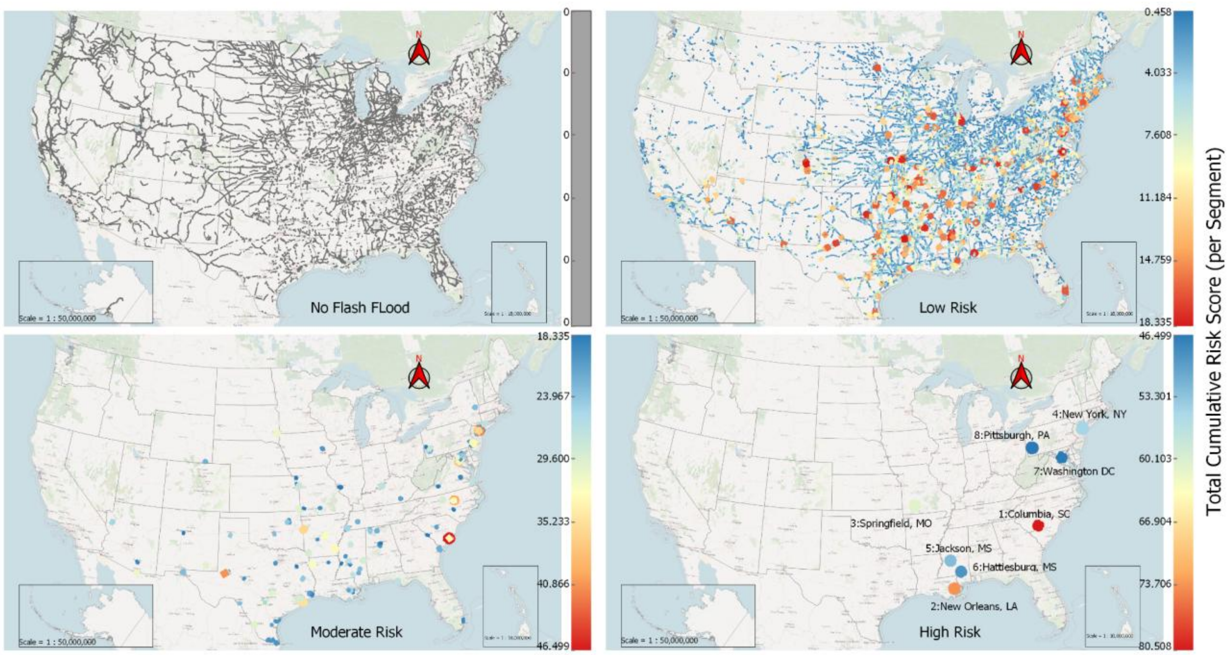

Figure 13. Spatial distribution of cumulative flash flood risk scores across U.S. rail segments.

3.1.2 Extreme Heat

For the extreme heat analysis, the researchers focused on heat events that coincided with reported derailments. They chose this scope because a network-wide heat analysis for every rail segment would require estimating daily operating conditions and rail temperatures across the entire U.S. system.

For each derailment location, the researchers reconstructed the rail-track temperature on the event day. To do this, they used a simplified rail temperature model that estimates the day's peak rail temperature from air temperature and solar radiation, shown in Eq. 13:

$$\alpha_r SR x_2 + T_i x_1 - \beta = T_r \tag{13}$$

In this model, α_r, is the rail conduction constant, which we set to 0.75. (SR) as the expected daily maximum solar radiation, (T_i) as the daily maximum air temperature, and (T_r) as the estimated daily maximum rail temperature. The coefficients (x_1) and (x_2) are fitted parameters. This reduced linear form follows prior rail-temperature forecasting approaches that relate peak rail temperature to forecast air temperature plus an adjustment term (Zhang et al., 2007).

The researchers then calibrated the parameters in Eq. 12 using observations from Amtrak's Engineering Track Testing Weather and Rail Temperature Monitoring program (AMTRAK, 2024). They used data from three monitoring stations to estimate the coefficients for Eq. 13. They then applied the calibrated model to estimate rail-track temperature on the days and at the locations of historical derailments in our secondary dataset.

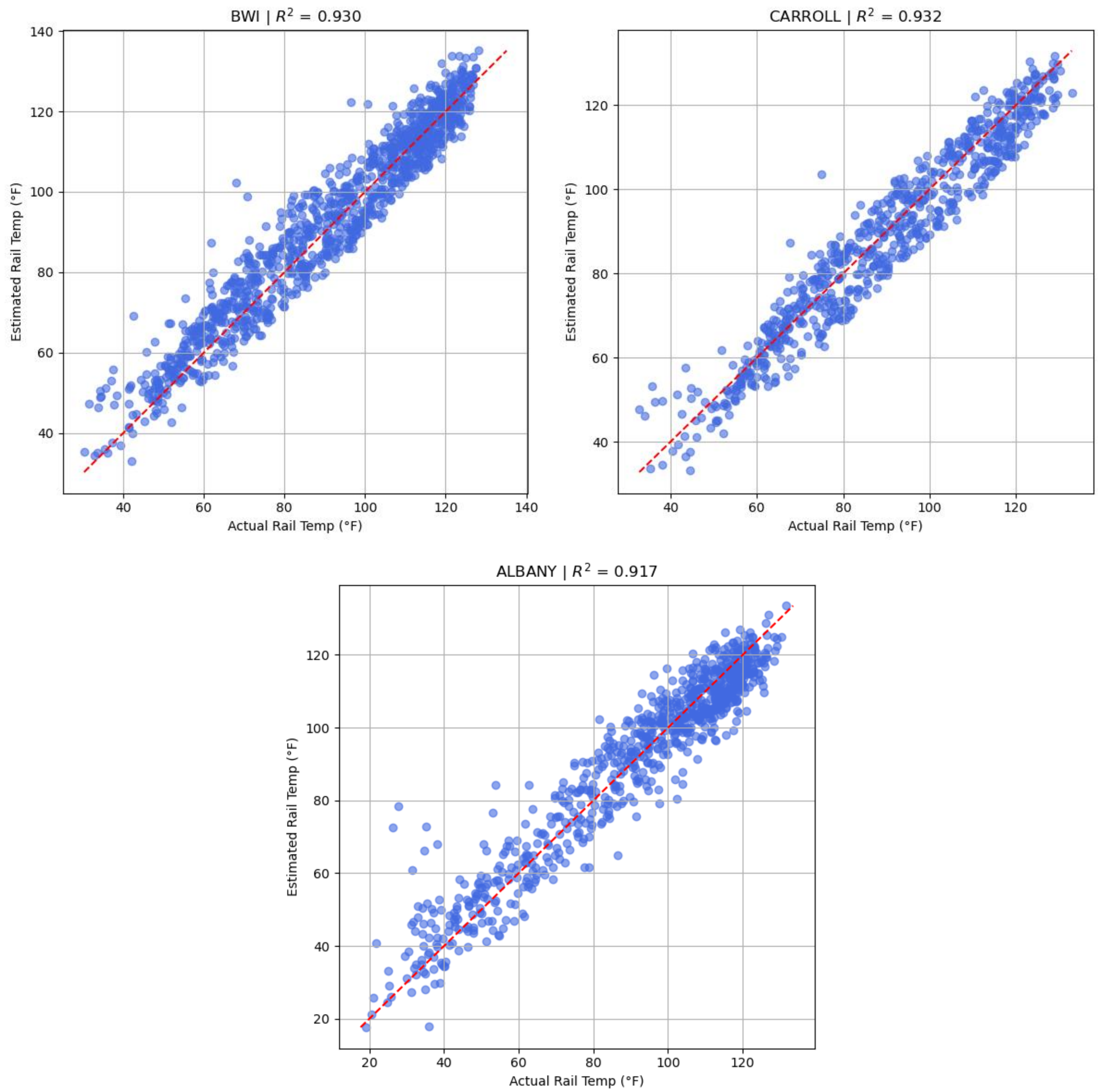

Figure 14. AMTRAK fitted rail track temperature comparison outputs

After calibration, the research team used an external dataset that provides both solar radiation and air temperature so we could apply the fitted coefficients from Eq. 13 at each derailment location. They used the National Solar Radiation Database (NSRDB) PSM v4 GOES Aggregated product, which provides 4-km gridded meteorological data extending back to at least 2000.

For each derailment site, researchers identified the nearest NSRDB grid cell and downloaded 30-minute solar radiation and air temperature records. They retrieved data for the event day and for all years back to 2000 so we could place the event in a long-term climatological context. They selected the GOES Aggregated v4 product because it provides a long, internally consistent time series suited to heat-risk analysis and includes the radiation components required by our filtering steps. It also provides radiation and air temperature together, which many standard meteorological datasets do not.

The research team queried the NSRDB CSV download API programmatically at 30-minute resolution, one year at a time, for the period 2000 through the event year. Each request targeted a single grid cell. They saved successful responses to:

./GOES_Query/<station>/<station>_<year>.csv.

Table 3 lists the full API query used in this study.

Table 4. NSRDB API Querry Parameters

Setting	Value
Dataset	nsrdb-GOES-aggregated-v4-0-0
Endpoint	CSV download API
Geometry	WKT point: POINT(lon lat)
Cadence	interval=30 minutes
Years	names=2000,…,2024
Attributes Requested	dni, dhi, ghi, clearsky_dni, clearsky_dhi, clearsky_ghi, air_temperature, wind_speed
Outputs per year	One CSV per year saved to ./GOES_Query/<station>/<station>_<year>.csv

They used the NSRDB data and Eq. 13 in two steps. First, they estimated peak rail-track temperature on the derailment day at the derailment location. Second, they ranked that temperature against the historical distribution at the same location using all available days back to 2000. This percentile rank shows how unusual the derailment-day rail temperature was relative to typical conditions at that site.

Figure 16 maps estimated rail-track temperatures for derailments that occurred on continuous welded rail (CWR) and have valid coordinates. Figure 16 also includes an example time series for a single derailment location, showing the historical distribution of estimated rail temperatures.

In that example, the derailment occurred on July 15, 2015. The research team estimated rail-track temperature for 5,672 prior non-derailment days at the same location. The derailment-day temperature fell at the 99.4th percentile of the 2000–2015 distribution, indicating that it occurred under unusually high thermal conditions.

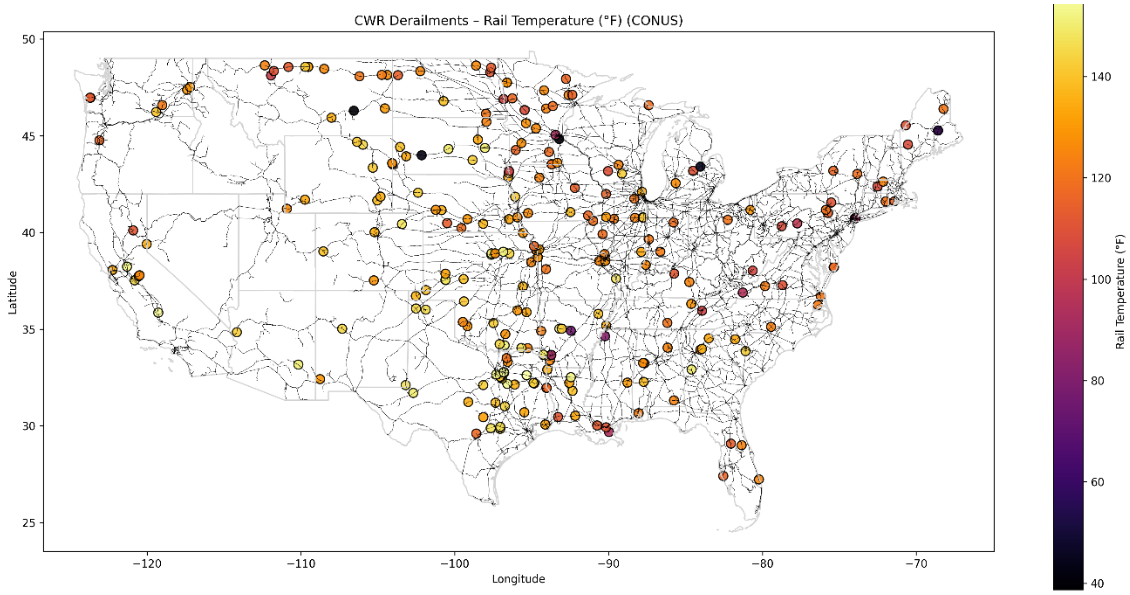

Figure 15. Rail track temperature map of historic T109 sun kink derailments.

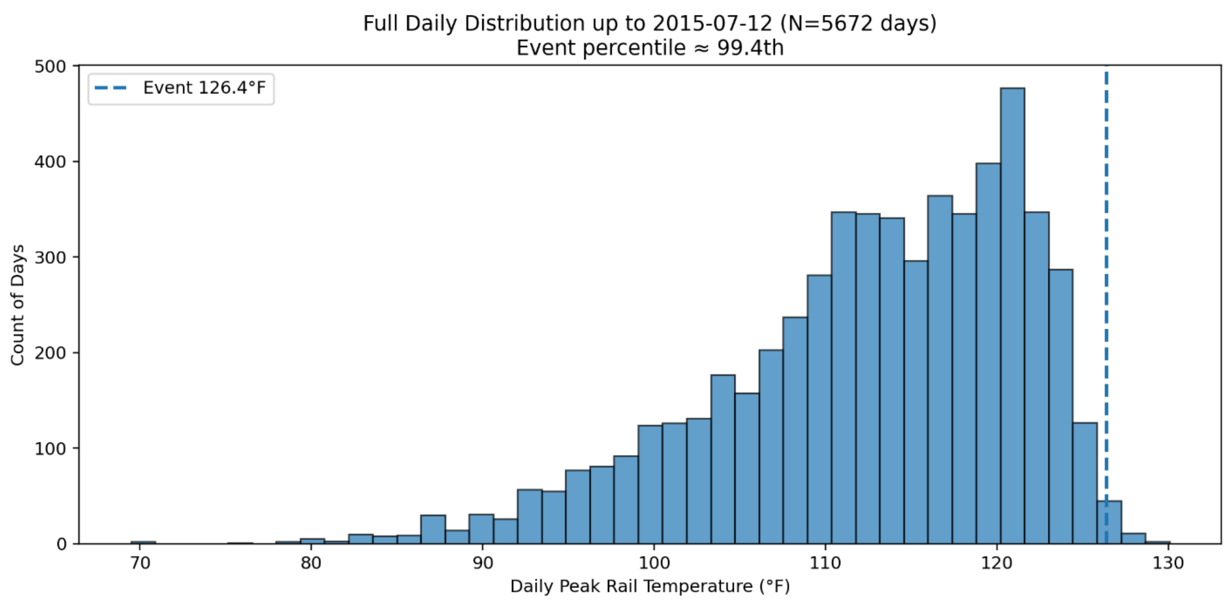

Figure 16. Demonstration on the estimated rail track temperature percentile for a specific location where a derailment took place with number of days estimated (n=5672).

Figure 17 and 18 provide a map on the percentile of each of the derailment based on their relative track temperature, and depict the majority of T109 derailments on CWR are indeed occurring during suppressive heat conditions, in the above 80th percentile over 100 degrees F. Lastly, Figures 19-22 demonstrate the CDF of the past derailments relative percentile demonstrating some occurred at lower percentile rankings and temperatures providing a viewpoint on an extreme lack of lateral resistance to buckling in those events.

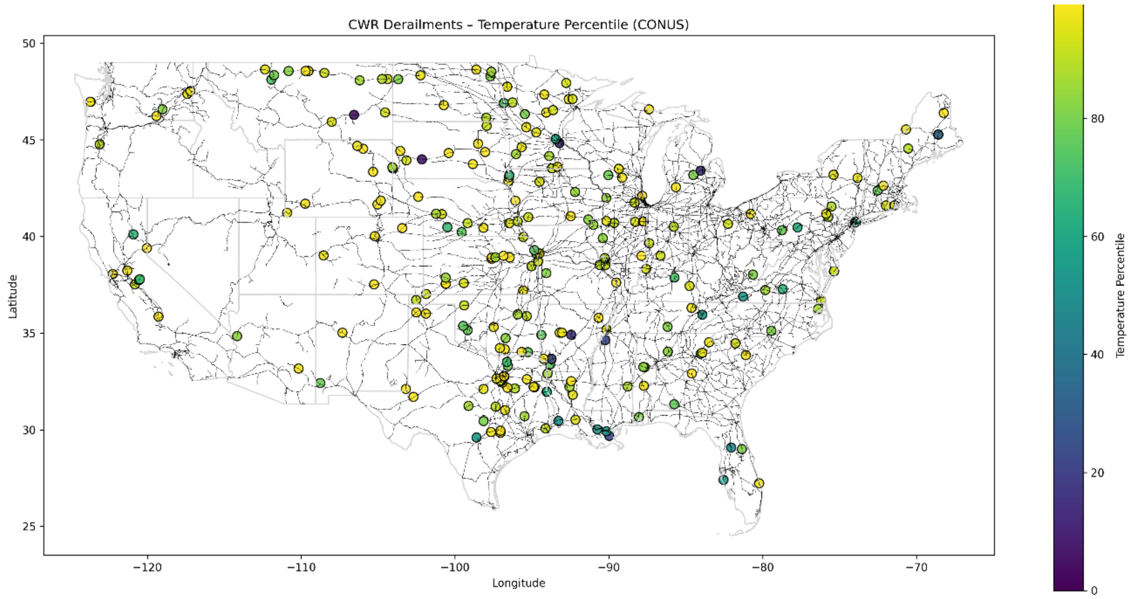

Figure 17. US rail network map of the relative percentile of rail track temperature that occurred the day of derailment vs. data going back to 2000 at that location.

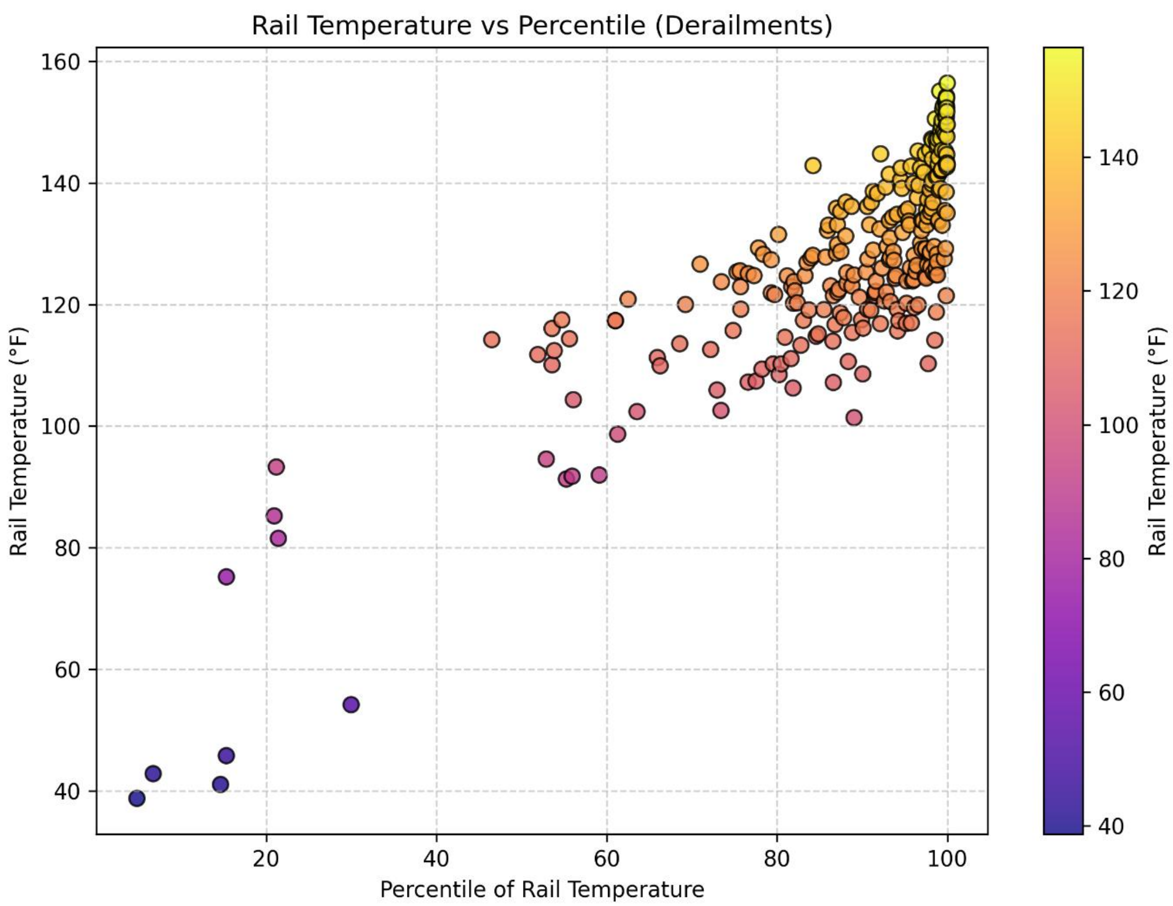

Figure 18. Estimated rail track temperature vs. rail temperature percentile derailment.

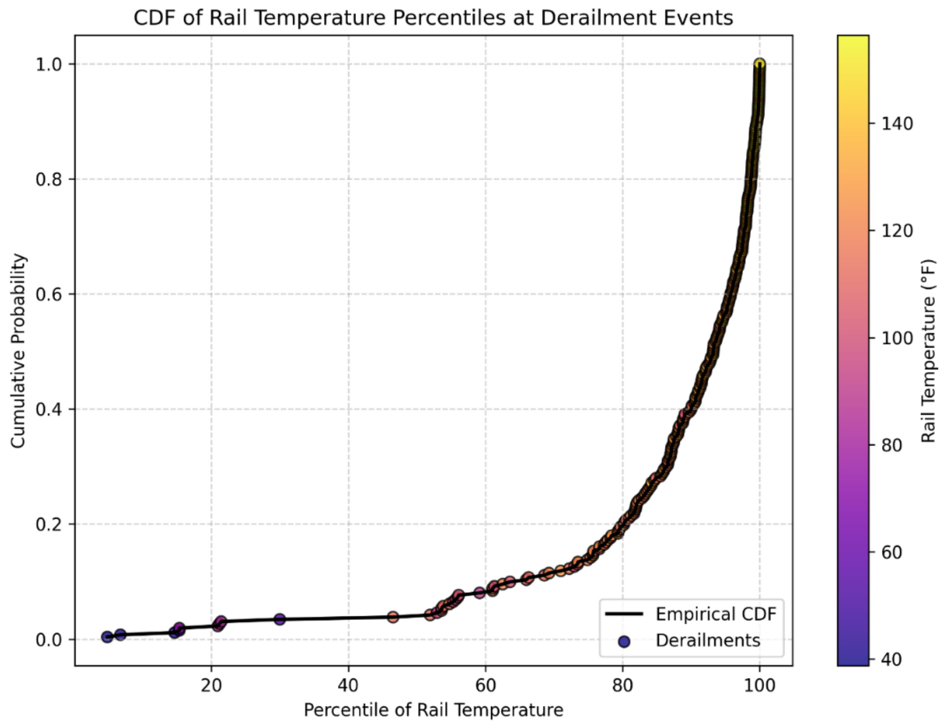

Figure 19. Cumulative density function for the percentile of rail temperature derailments.

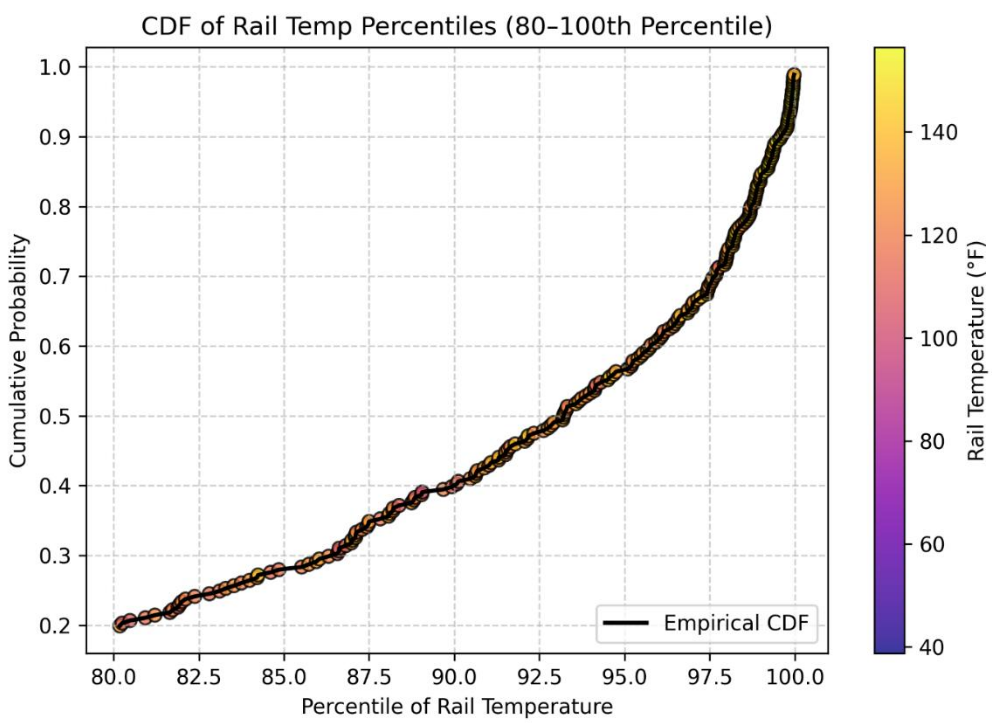

Figure 20. Cumulative density function for the percentile of rail temperature derailments: 80^{th}-100^{th} percentile.

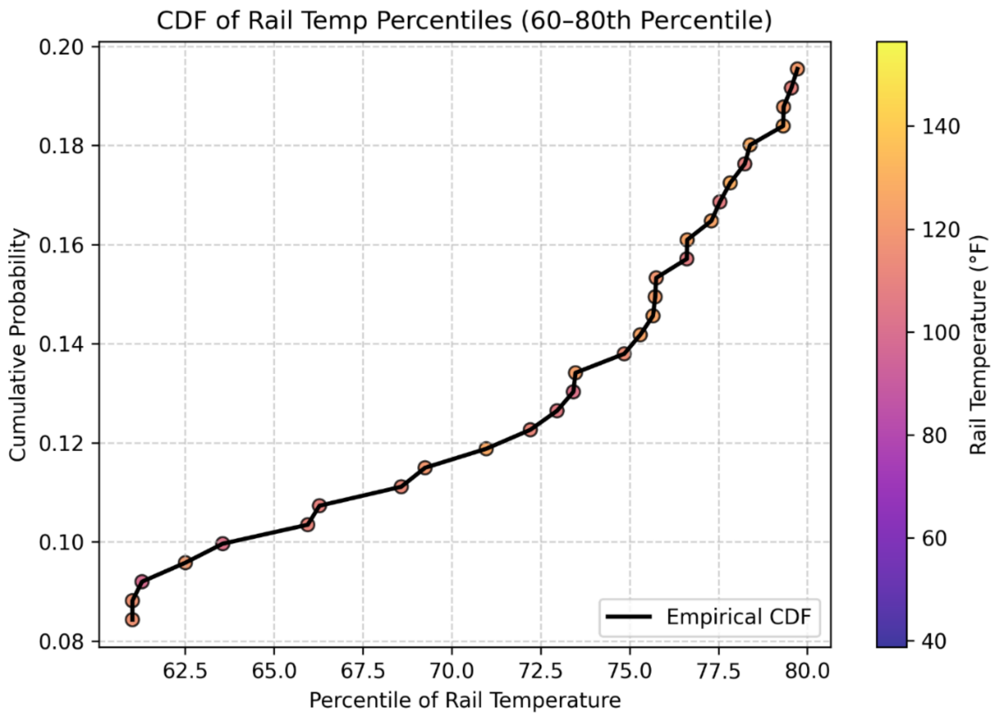

Figure 21. Cumulative density function for the percentile of rail temperature derailments: 60th-80th percentile.

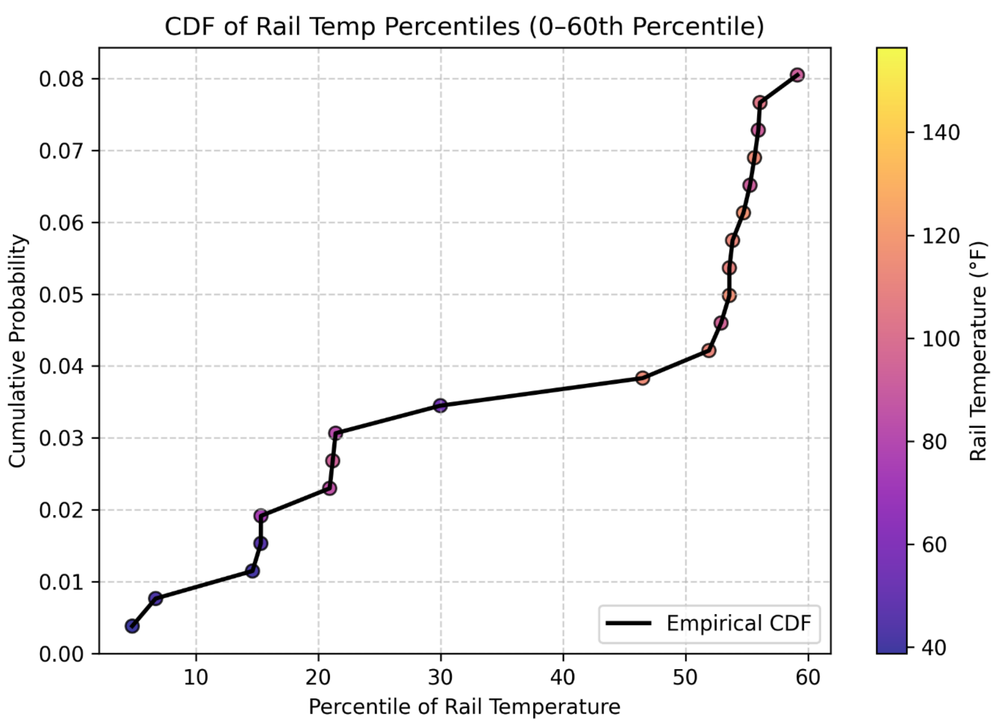

Figure 22. Cumulative density function for the percentile of rail temperature derailments: 0-60th percentile.

4. Development of railway fragility functions towards resilience quantification

A resilience assessment typically uses three linked curves. First, a loss-of-function curve describes how performance drops as hazard intensity increases. Second, a fragility function estimates the probability of reaching different damage states at each intensity. Third, a restoration profile describes how performance recovers over time after a disruption.

Figure 23 (Sathurshan et al., 2022) illustrates this performance-based resilience framework. Figure 24 (Poulin & Kane, 2021) shows how the same concept can be summarized as a cumulative resilience curve for an entire network.

For rail, the research team defined "network performance" in several ways. For example, they can measure connectivity as the number of links that remain connected between key nodes, or use other network-science measures to represent system function. They can also compute these metrics for the full U.S. rail network or for a subset defined by a specific owner or operating rights on particular segments.

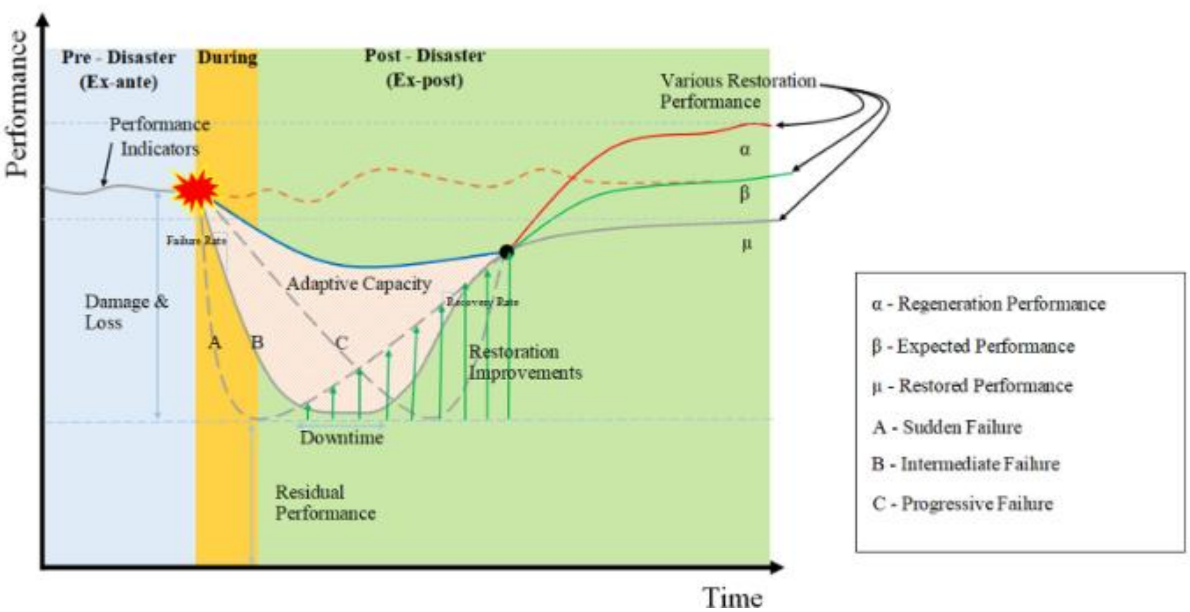

Figure 23. Performance based resilience curve (Sathurshan et al., 2022)

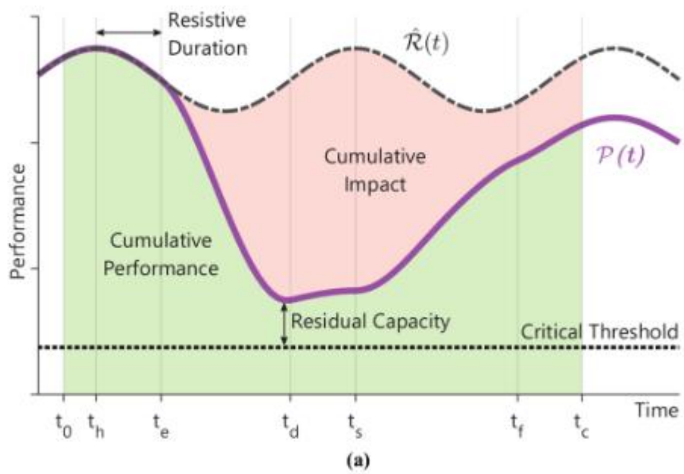

Figure 24. Cumulative Performance Resilience curve for a Network (Poulin & Kane, 2021)

Recent studies on resilience curves frame infrastructure performance under hazards using four phases: preparation, absorption, recovery, and adaptation (Poulin & Kane, 2021; Wells et al., 2022). Following this literature, we developed an internal resilience-curve framework for the U.S. rail network under different natural hazards. Figure 25 presents this framework.

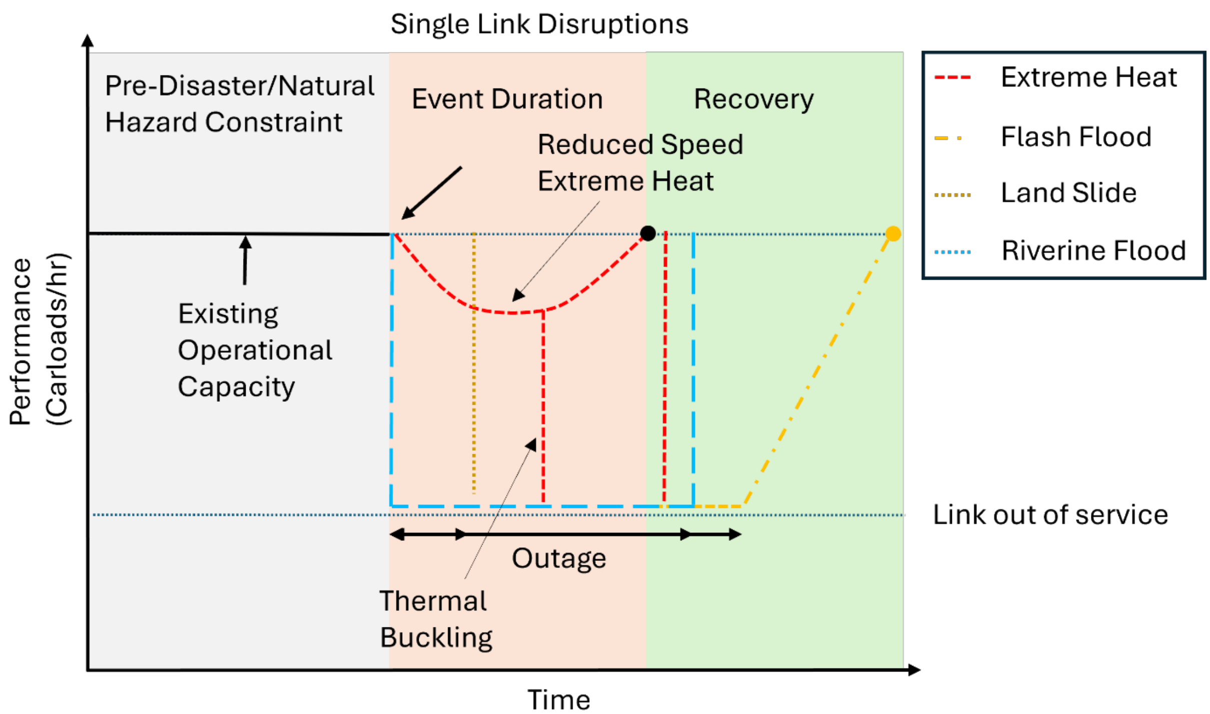

Figure 25. Resilience Curve Depiction for Rail Links in the Face of Natural Hazards.

The research team defined baseline operational capacity in terms of railcars per hour. Allowable speed can vary, but capacity is usually stable within a given track class until a disruption occurs. In this study, the disruptions of interest are extreme heat, flash flooding, riverine flooding, and landslides.

Event duration differs by hazard. Landslides occur quickly, so the "event" is short, but recovery can take much longer if crews must clear debris and repair the track. Flooding behaves differently. A riverine flood may inundate a yard or low-lying segment for hours, and operations may resume quickly if the track is not damaged. In that case, the disruption lasts as long as the water remains on or near the track, even if recovery is rapid once the flood recedes.

Extreme heat introduces a different mechanism: railroads may impose slow orders to reduce buckling risk. For example, BNSF manages heat risk by maintaining an installation range for rail neutral temperature. When rail temperature exceeds a threshold, BNSF can impose slow orders that reduce allowable speed on affected segments, in some cases to 25 mph (BNSF, 2023). In a resilience curve, this appears as an immediate drop in performance that persists for the duration of the slow order.

Heat-related performance loss depends on more than rail temperature alone. It also depends on track condition and the segment's lateral resistance relative to its rail neutral temperature. For this reason, we represent extreme heat using an "allowable temperature above neutral" concept that sets operational responses. In our framework, conditions can trigger (1) a slow order, (2) a stoppage, or (3) in rare cases, a thermal buckle. Each response produces a different resilience-curve shape because it changes both the magnitude and duration of the performance reduction and the time required for recovery.

4.1 Fragility curve creation for extreme heat events

The research team analyzed 153 derailments that occurred on continuous welded rail (CWR) segments using a principal component–based intensity measure. This PCA index combines peak rail-track temperature and train speed to represent the joint thermal and operational stress present at the time of the event.

They selected these 153 derailments using three criteria. First, the derailment had to be classified as a sun kink or buckle (FRA cause code T109). Second, the record had to include valid coordinates so we could link it to gridded meteorological data. Third, the derailment had to occur between 2000 and 2024 so that it aligned with the NSRDB query period described in Section 6.3.2.2.

The resulting PCA index captures the balance between environmental loading and operating conditions:

Low PCA values (0-1.5): Failures under extreme environmental stress (high temperature, high speed). Indicates the infrastructure held until conditions became severe.

High PCA values (3-5): Failures under mild to moderate stress. Indicates weak lateral resistance suggesting limited maintenance.

The research team used PCA instead of a Bayesian failure-probability model because derailments are rare at the segment level. A traditional fragility curve needs many failure and non-failure observations across a wide range of intensities to estimate how failure likelihood changes with increasing stress. In our data, most segments operate for decades with no derailments, and failures are extremely sparse. With data this limited, a probability-based model would be dominated by assumptions rather than observations.

Given these constraints, the research team focused on the conditions observed during known failures. Specifically, we examined the combined stress level at the time of derailment and used that information to infer relative differences in track resistance across events. They then applied PCA to the two key variables available for each derailment, peak rail temperature and train speed, to create a single intensity measure for comparing stress levels across the 153 cases.

4.1.1 Methodology

The research team considered two explanatory variables for derailment likelihood based on available data from the FRA form 54C, this included the speed of the train, and the now recreated rail track temperature (tonnage was also used however it did not improve the model and was therefore not used. Each of the two variables were then standardized through Equation 14:

$$T^* = \frac{T - \mu_T}{\sigma_T}, \qquad S^* = \frac{S - \mu_S}{\sigma_S} \tag{14}$$

PCA estimation follows in Equation 15

$$IM_{PCA} = a\,T^* + b\,S^* \tag{15}$$

where T^* and $S^*$are the standardized values for rail temperature and speed, *a,* and *b*, their coefficients and equation 16 establishes the PCA outputs to be positive.

$$IM_{PCA} = IM_{PCA} - (IM_{PCA}) + \varepsilon, \quad \varepsilon > 0 \tag{16}$$

Cumulative density function (CDF) for the now created PCA estimation data following equation 17.

$$\hat{F}(x) = \frac{1}{N}\sum_{i=1}^{N} \quad (IM_{PCA,i} \leq x) \tag{17}$$

From this now created CDF, we move forward to fit the curve following several standard distributions, names Weibull, Logistic and Normal ($F_W(x)$, $F_L(x)$, and $F_N(x)$) respectively within equations 18-20.

$$F_W(x) = 1 - exp\left[-\left(\frac{x}{\lambda}\right)^k\right], \qquad k > 0,\, \lambda > 0 \tag{18}$$

$$F_L(x) = \frac{1}{1 + exp\left(-\frac{x-\mu}{s}\right)} \tag{19}$$

$$F_N(x) = \Phi!\left(\frac{x-\mu}{\sigma}\right) \tag{20}$$

Lastly the model chosen utilizes the smallest RMSE and was found to be the standard Logistic distribution.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(\hat{F}(x_i) - F(x_i)\right)^2} \tag{21}$$

Table 5. Fitted PCA distribution results

Distribution	RMSE
Weibull	0.029
Logistic	0.023
Normal	0.024

From here the research team use the logistic form of the CDF, where it is possible to directly interpret where the PCA falls amongst all of the other derailment PCAs to provide insight on how likely a rail temperature and rail speed will result in a derailment respective to past derailments.

$$RP_{Logistic}(x) = \frac{1}{1 + exp\left(-\frac{x-\mu}{s}\right)} \tag{22}$$

where $\mu = 2.36$, $s = 0.60$.

Equation 22 maps the PCA intensity value (x) to a derailment exceedance probability. In future work, they can use the PCA parameters to translate operating conditions into relative risk. Given an estimated rail-track temperature, they can also evaluate how different train speeds align with the conditions observed during past derailments and identify speed–temperature combinations associated with higher exceedance probabilities.

We used train speed and estimated rail-track temperature because these were the only consistently available variables in the derailment records. Rail-track temperature also required a separate reconstruction workflow. If additional information becomes available, such as track geometry, maintenance condition, fastening systems, ballast condition, or other component-level attributes, the analysis can be extended using more detailed mechanistic or hybrid methods, including approaches such as CWR-SAFE (Kish and Samavedam, 2013).

4.1.2 Results

The IM_{PCA} values define an intensity scale that combines rail-track temperature and train speed into a single measure. Larger IM_{PCA} values indicate derailments that occurred under relatively mild thermal and operational conditions, which suggests lower lateral resistance and potentially weaker support conditions (for example, ballast or tie performance). Smaller IM_{PCA} values indicate derailments that occurred only under high stress, such as unusually high rail temperatures combined with higher train speeds. Intermediate values represent cases where both track condition and environmental loading likely contributed.

The fitted parameters provide a baseline, reduced-form fragility relationship. Users can apply these parameters to estimate derailment exceedance probabilities from combined thermal and operational conditions without re-fitting the model.

Figure 23 compares the empirical CDF of IM_{PCA} with Weibull, logistic, and normal distribution fits using Eqs. 17–20.

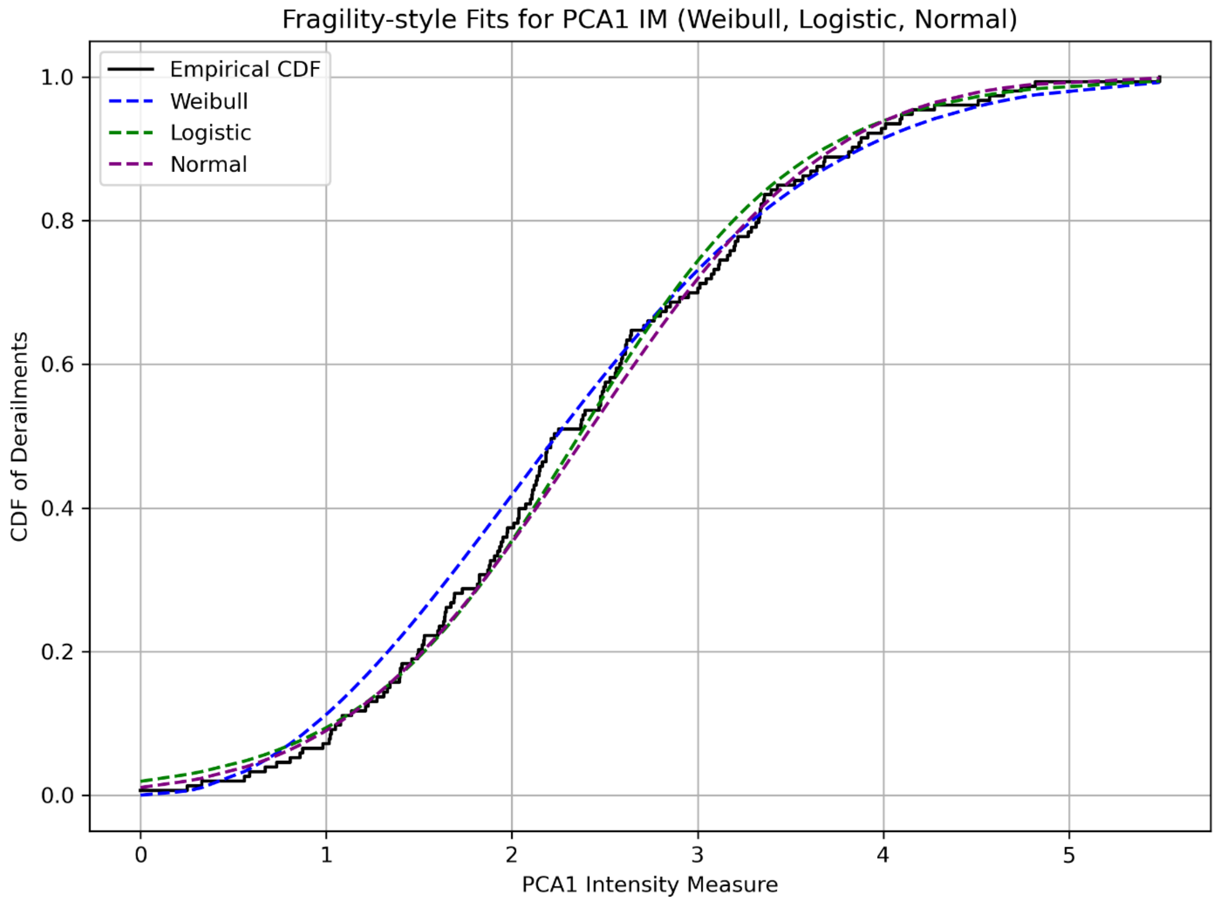

Figure 26. Extreme heat thermal buckling PCA intensity measure CDF for historic derailments.

Figure 24 provides a comparison of rail track temperature, speed, and the corresponding PCA that was estimated. For further verification one can observe the PCA value <1.0 that has a nearly 80 mph speed, whilst a relatively lower temperature ≈112 °F or a similar value in PCA that has a lower speed but higher rail track temperature. This interplay that is captured through the PCA is providing a metric to understand the relative risk of derailment based on past events. Lastly, Table 5 provides some statistical metrics on two ranges of the PCA, those less than 1.5, and those over 3.0, what can be interpreted is that the operating conditions on the higher PCA were not as severe and can almost entirely be attributed to limited lateral resistance or other infrastructure born limitations, as opposed to those driven from extreme environmental and operational conditions.

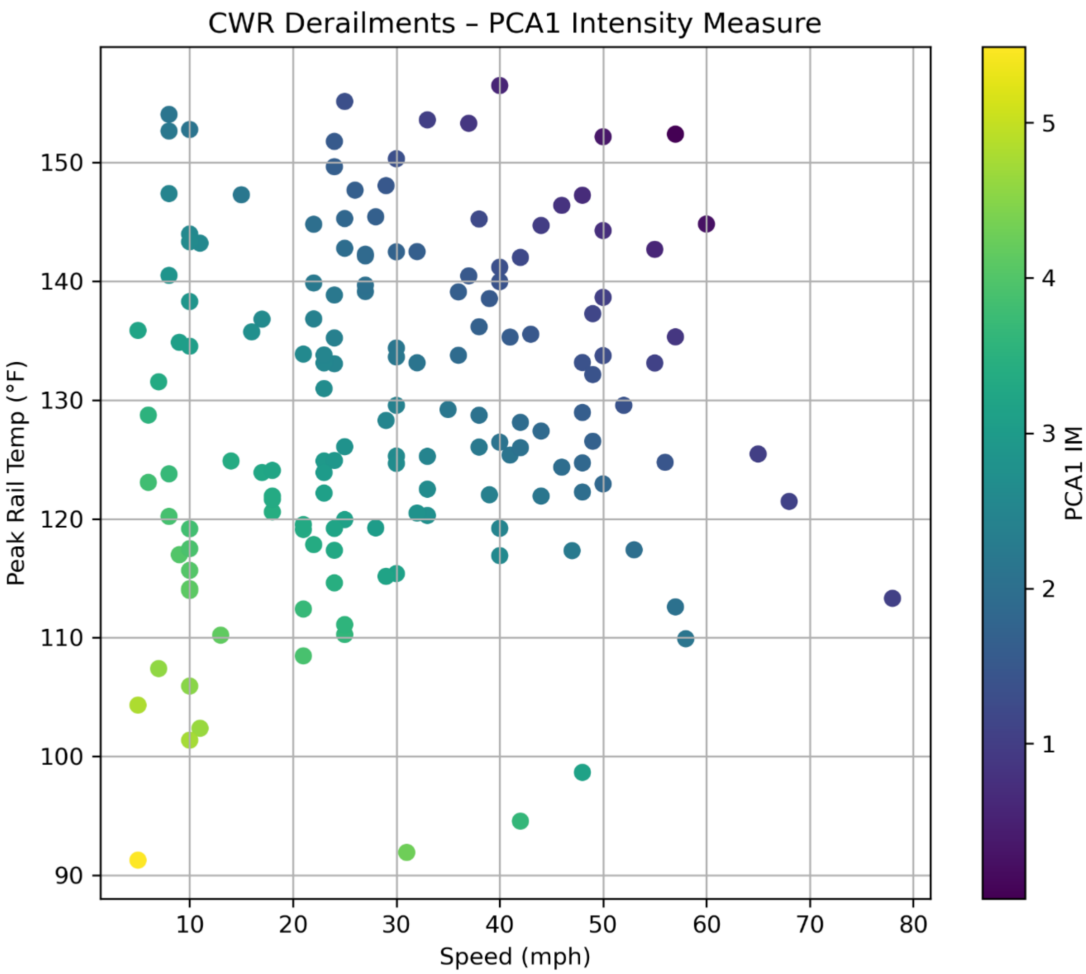

Figure 27. PCA distribution results: Rail Temperature vs. Speed at the time of derailment.

Table 6. PCA analysis results per category

	Mean Speed ± Std.(mph)	Mean Temp ± Std (°F).	Speed Range (mph)	Temp Range (°F).
$IM_{PCA} < 1.5$ (n = 31)	47.9 ± 11.6	140.4 ± 10.6	25 – 78	113.3 – 156.4
$IM_{PCA} > 3.0$ (n = 46)	17.2 ± 9.8	116.1 ± 10.5	5 – 48	91.3 – 135.8

Table 6 shows a clear spatial pattern: many high-PCA locations occur in the northern United States. These derailments occurred under relatively mild combined heat and speed conditions, which suggests that limited lateral resistance or other track-condition factors may have contributed more strongly than extreme thermal loading. One possible explanation is that northern locations experience large seasonal temperature swings between winter and summer.

These transitions may increase thermal stress or lead to less favorable rail neutral temperature conditions if the rail is not well matched to the local temperature regime.

Union Pacific (UP) shows a lower mean PCA, indicating that its observed derailments more often occurred under higher combined thermal and operational stress. This pattern is consistent with UP's network geography, which includes substantial mileage in the Southwest where extreme heat conditions are more common.

Table 7. PCA fragility owner results

Owner	Derailments	Mean PCA	Median PCA	Min PCA	Max PCA
BNSF	49	2.19	1.95	0.25	4.51
CN	4	3.32	3.52	2.25	4.01
CSXT	12	2.40	2.30	1.40	3.69
NS	9	2.72	2.62	1.46	4.57
UP	35	1.80	1.73	0.33	3.52
Non-Class-1	44	2.94	2.88	0.00	5.49

The research team next examined how the PCA-based fragility results vary geographically and by owner. They mapped results at three scales: national, state, and railroad ownership. Figure 29 shows the national pattern. Figures 29–35 show Class I results, Figure 35 shows Non–Class I results, and Appendix A presents state-level maps.

Figure 29 maps the spatial distribution of PCA intensity values for all 153 thermal buckling derailments across the contiguous United States. Each point represents a single derailment event, colored along a gradient from blue (low PCA, 0–1.5) to red (high PCA, 3.0–5.5). Blue points indicate derailments that occurred under severe combined thermal and operational stress, high rail-track temperatures and higher train speeds, meaning the infrastructure performed until conditions became extreme. Red points, by contrast, indicate derailments that occurred under relatively mild heat and low-speed conditions. Because environmental loading alone cannot explain these failures, they are most likely attributable to insufficient lateral resistance, degraded ballast, or deferred maintenance rather than extreme weather. Intermediate yellow and orange points represent events where both environmental loading and track condition likely contributed. Geographically, the CONUS map reveals a notable concentration of higher-PCA (red) events across the northern United States, consistent with the seasonal freeze-thaw cycles and neutral temperature mismatches discussed in the preceding analysis. Figures 30 through 34 disaggregate these same results by Class I railroad owners: Norfolk Southern, CSX, CN, Burlington Northern Santa Fe, and Union Pacific respectively, while Figure 35 presents results for Non-Class I carriers. The state-level breakdowns in Appendix A follow the same color scale and should be interpreted using this same framework.

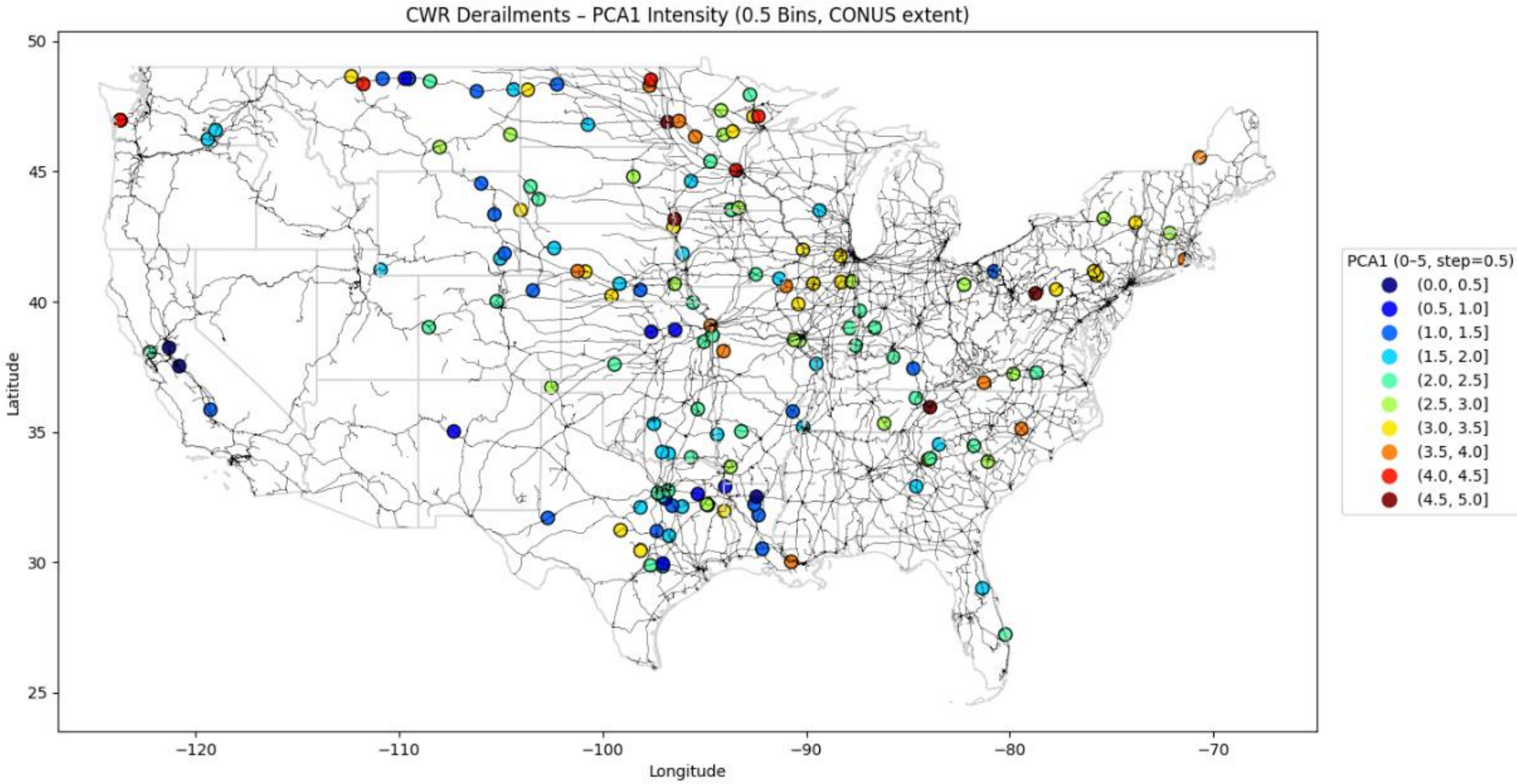

Figure 28. PCA results CONUS

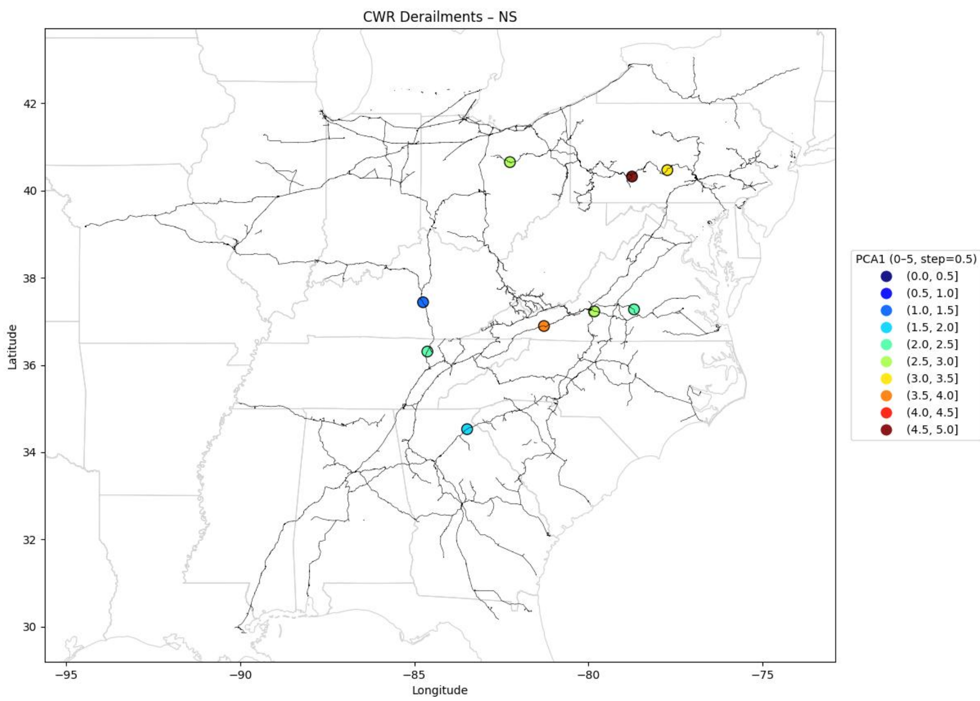

Figure 29. PCA results Norfolk Southern

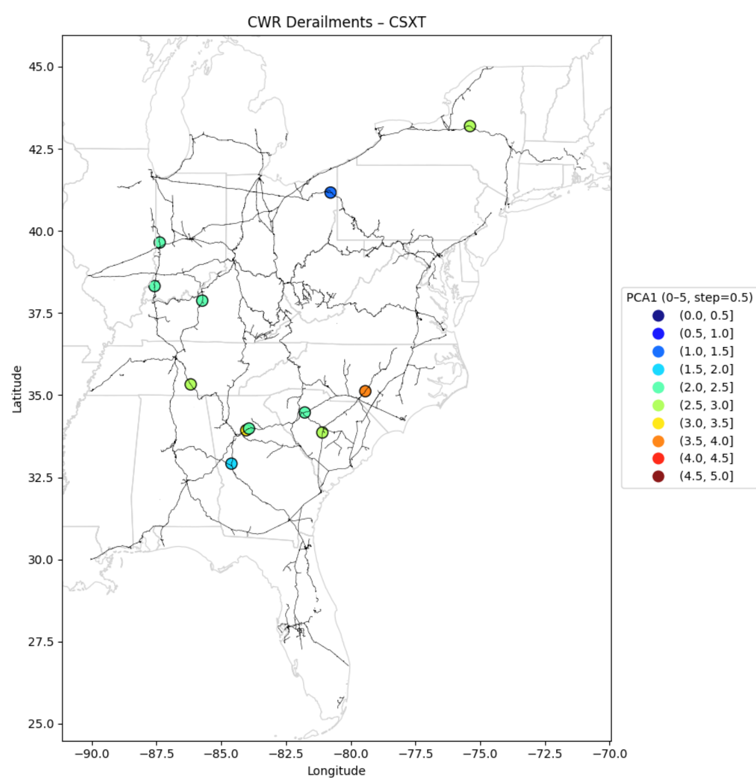

Figure 30. PCA results CSX

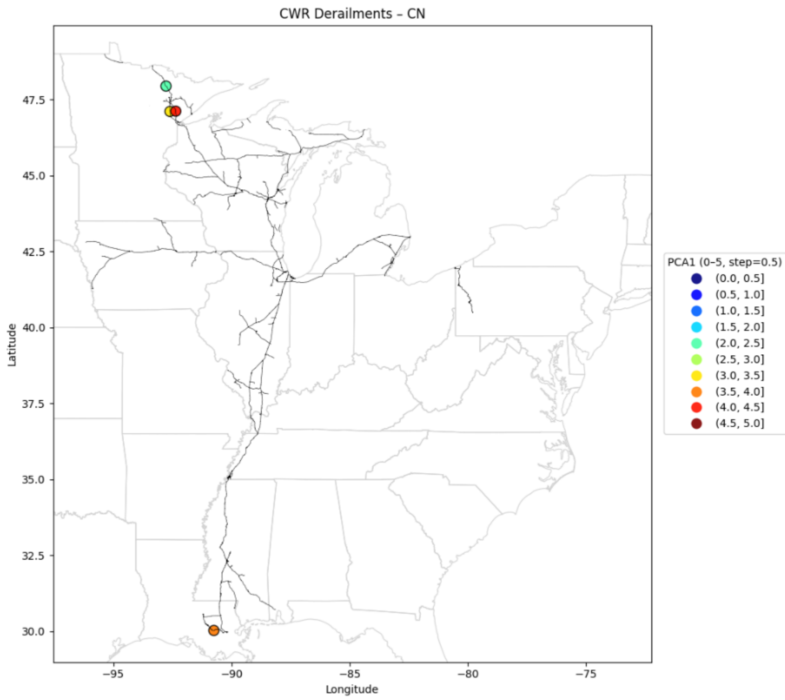

Figure 31. PCA results CN

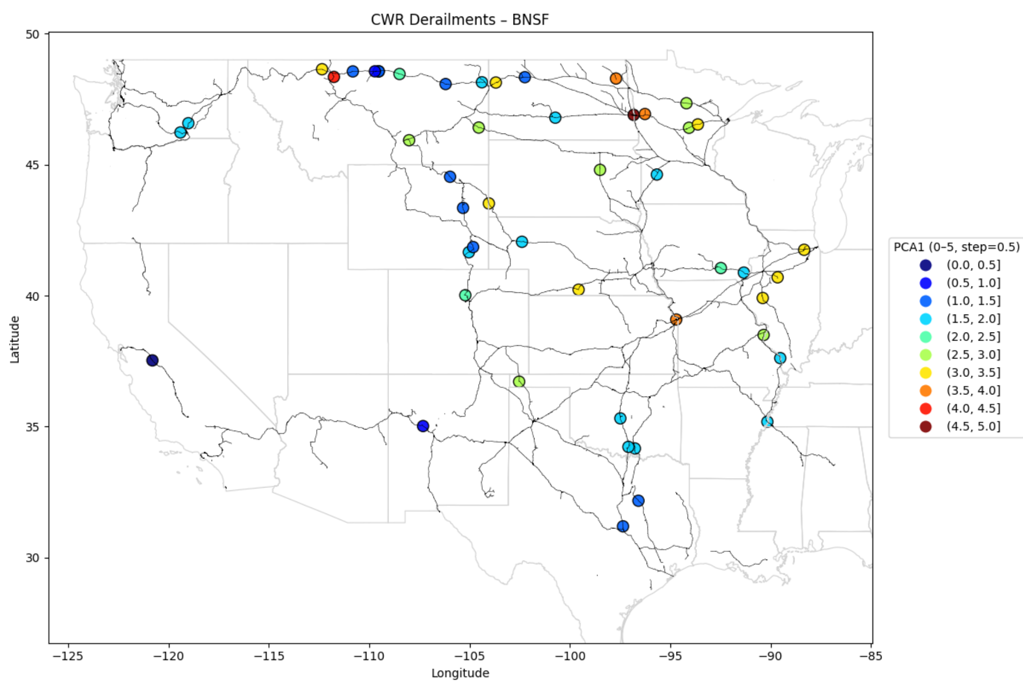

Figure 32. PCA results Burlington Northern Santa Fe

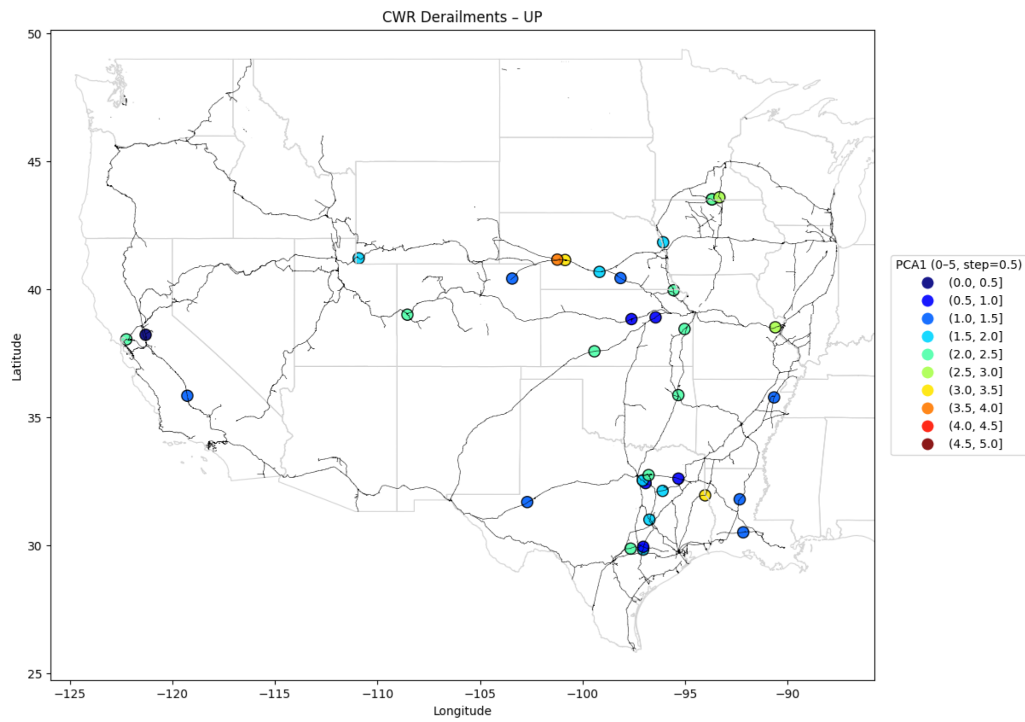

Figure 33. PCA Results Union Pacific

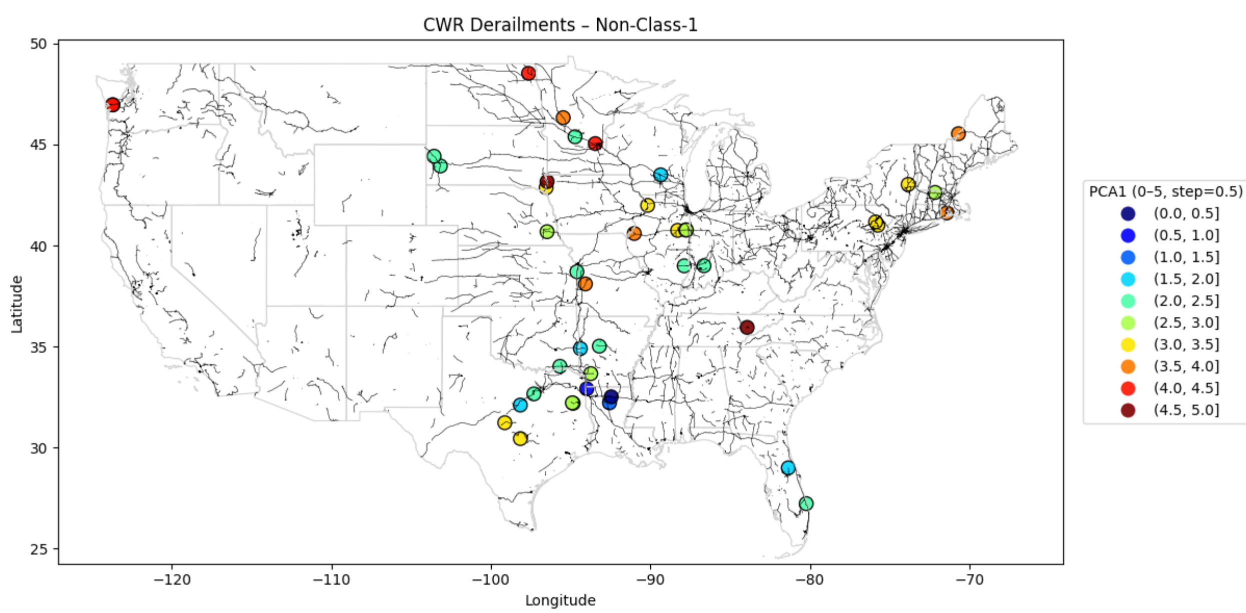

Figure 34. PCA Results Non-Class-1

4.2 Restoration / recovery-time curves by damage state

The FRA Derailment/Incident database reports the cost of many hazard-related events, but it does not record outage duration. It also does not track slow orders, so we cannot determine when a slow order begins, how long it lasts, or which segments it affects.

For buckles, washouts, and other forms of track damage, the database does not provide a consistent repair window or the time required to restore operations. As a result, we cannot quantify how hazard severity translates into outage duration at the segment level.

If outage-duration data were available, we could estimate relationships between hazard intensity and expected outage time for each segment and use those relationships to support resilience modeling and planning.

4.3 Tools & validation frameworks for resilience quantification

Outage duration and recovery time are required to build full resilience curves. Because the research team did not have consistent outage-duration data, they could not estimate the recovery portion of the curve. Even with this limitation, they developed the hazard–impact components that the available data support. In this section, they developed a fragility-style relationship for extreme heat that links event intensity to the likelihood of an outage-related outcome. For flooding, the research team produced resilience-relevant risk scores in Section 2.2.1 that summarize both severe single events and repeated exposure. They made the data and code for both analyses publicly available through the submitted GitHub repositories.

5. Conclusion

The research team developed new ways to quantify natural hazard risk across the U.S. rail network. We integrated datasets from multiple federal agencies with FRA records and used Form 54C incident and derailment data to validate key results. Across several tasks, they produced risk scores and analytic frameworks that support rail resilience assessment.

Section 1 identifies pathways to obtain U.S. rail operations data and describes the data needed to strengthen national-scale hazard assessments. It also explains how improved operational and maintenance records would enable better estimates of adaptive capacity and clarify how routine hazards affect daily rail performance.

Section 2 presents a data-driven framework to evaluate flash flood exposure across the U.S. rail system using high-resolution precipitation records, extreme value methods, and spatial mapping. The research team linked 53,990 NCEI-reported flash flood events with LCD-derived rainfall statistics from 1,489 stations to estimate flash flood intensity, duration, and frequency. They then translated these hydrometeorological measures into two segment-level risk scores: a maximum return-period score that captures extreme single-event exposure and a cumulative exposure score that captures repeated flooding. The maps show high-risk corridors concentrated in the Southeast, Midwest, and parts of the Northeast. Segments that score highly on both metrics represent priority locations for inspection, resilience planning, and targeted upgrades.

The research team also evaluated extreme heat by estimating rail-track temperature percentiles for derailment days and comparing them to the historical temperature distribution at each location. In Section 3, they developed a reduced-form fragility framework that combines rail-track temperature and train speed to estimate relative buckling risk based on past derailments. Together, these results provide an initial, data-driven view of how extreme heat and flooding affect U.S. rail infrastructure.

The report also identifies key needs for future work. Priorities include stronger link- and node-level criticality analysis using the NARN network and the integration of climate-model projections to evaluate future hazard scenarios at the segment level. These extensions require improved access to operational, maintenance, and outage-duration data.

Overall, this report supports the FRA mission of safe, reliable, and efficient movement of people and goods by providing datasets, methods, and a clear roadmap for expanding national-scale rail hazard and resilience assessment.

References

American Society of Civil Engineers (ASCE). (1982). *Guidelines for Evaluating Flood Frequency Analyses*. ASCE Manuals and Reports on Engineering Practice No. 28.

American Railway Engineering and Maintenance-of-Way Association (AREMA). Manual for Railway Engineering, Chapter 6: Roadbed and Ballast. 2021.

AMTRAK. Engineering Track Testing Weather & Rail Temperature Monitoring. 2024.

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.

Benson, M. A. (1963). *Distribution of frequency curves* (No. G-1). US Geological Survey.

Brasil, J. B., Guerreiro, M. S., Andrade, E. M. de, de Queiroz Palácio, H. A., Medeiros, P. H. A., & Ribeiro Filho, J. C. (2022). Minimum Rainfall Inter-Event Time to Separate Rainfall Events in a Low Latitude Semi-Arid Environment. *Sustainability*, *14*(3), Article 3. https://doi.org/10.3390/su14031721

BNSF Railway. Guidelines for Industry Track Projects: Drains for Steep Slopes (Sheet A–9). 2020. Sheet A–9 in BNSF Industry Track Standards.

Bubeck, P., Dillenardt, L., Alfieri, L., Feyen, L., Thieken, A. H., & Kellermann, P. (2019). Global warming to increase flood risk on European railways. *Climatic Change*, *155*(1), 19–36. https://doi.org/10.1007/s10584-019-02434-5

Bureau of Transportation Statistics. (2024). *National Transportation Atlas Database (NTAD)—Rail Lines*. https://geodata.bts.gov/search?q=Rail

Canadian National Railway (CN). *Culvert Crossings Specification*. 2020.

Canadian Pacific Kansas City (CPKC). Grade-Separation Design Guidelines. 2022

Canadian Pacific Railway (CP). Public Projects Manual. 2021.

Castillo, E., & Hadi, A. S. (1997). *Fitting the Generalized Pareto Distribution to Data*. Springer-Verlag.

Cheng, L., AghaKouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic Change*, *127*(2), 353–369.

Chung, S., Sardak, D., Kitsak, M., Jin, A., & Linkov, I. (2025). Contested logistics: Resilience of strategic highways and railways. *Transportation Research Interdisciplinary Perspectives*, *32*, 101507. https://doi.org/10.1016/j.trip.2025.101507

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer. https://doi.org/10.1007/978-1-4471-3675-0

Costabile, P., Costanzo, C., Ferraro, D., & Barca, P. (2021). Is HEC-RAS 2D accurate enough for storm-event hazard assessment? Lessons learnt from a benchmarking study based on rain-on-grid modelling. *Journal of Hydrology*, *603*, 126962. https://doi.org/10.1016/j.jhydrol.2021.126962

Council, N. E. R. (1975). *Flood Studies Report*. Institute of Hydrology, UK.

CSX Transportation. Pipeline Occupancy Specifications. 2021.

Data (IACWD), I. A. C. on W. (1982). *Guidelines for determining flood flow frequency*. Hydrology Subcommittee, Bulletin 17B.

Davison, A. C., & Smith, R. L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 393–442.

Douglas, E. M., Vogel, R. M., & Kroll, C. N. (2000). Regional frequency analysis using L-moments for hydrologic data in the United States. *Journal of Hydrology*, *241*(1–2), 43–63.

Federal Railroad Administration. (1997). *Notice of Safety Advisory 97-1: Damage to Tracks, Roadbed, and Bridges Resulting from Uncontrolled Flows of Water*. https://railroads.dot.gov/elibrary/flash-flooding

Federal Railroad Administration (FRA). Generic Plan for Continuous Welded Rail (CWR) Maintenance. U.S. Department of Transportation, 2021

Federal Railroad Administration (FRA). *Track Safety Standards: Continuous Welded Rail (49 CFR 213.118–119)*. *Federal Register*, 2019.

Ferreira, A., & Haan, L. de. (2013). On the block maxima method in extreme value theory: PWM estimators. *Annals of Statistics*, *43*, 276–298.

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, *24*, 180–190.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, *44*, 423–453.

Gomes, M. I., & Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review*, *83*(2), 263–292. https://doi.org/10.1111/insr.12058

Griffis, V. W., & Stedinger, J. R. (2007). Log Pearson Type 3 distribution and its application in hydrology. *Journal of Hydrologic Engineering*, *12*(5), 482–491.

Grimshaw, S. D. (1993). Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *Technometrics*, *35*(2), 185–191.

Ghoreishi, B., Staes, B.M., Liu, C. *et al.* Mapping Climate Extremes and Infrastructure Vulnerability: A New Database for U.S. Railroads. *Sci Data* **12**, 1932 (2025). https://doi.org/10.1038/s41597-025-06205-z

Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, *29*(3), 339–349.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Ilalokhoin, O., Pant, R., & Hall, J. W. (2025). Flood Risk Assessment Methodology for Complex Railway Networks: Application to the National Rail Network in Great Britain. *Natural Hazards Review*, *26*(1), 04024056. https://doi.org/10.1061/NHREFO.NHENG-1911

Jenkinson, A. F. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements. *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158–171. https://doi.org/10.1002/qj.49708134804

Jonkman, S. N., & Kelman, I. (2005). An analysis of the causes and circumstances of flood disaster deaths. *Disasters*, *29*(1), 75–97. https://doi.org/10.1111/j.0361-3666.2005.00275.x

Katz, R. W. (2013). Statistical Methods for Nonstationary Extremes. In A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, & S. Sorooshian (Eds.), *Extremes in a Changing Climate* (pp. 15–37). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4479-0_2

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8–12), 1287–1304.

Kellermann, P., Schöbel, A., Kundela, G., & Thieken, A. H. (2015). Estimating flood damage to railway infrastructure—The case study of the March River flood in 2006 at the Austrian Northern Railway. *Natural Hazards and Earth System Sciences*, *15*, 2485–2496. https://doi.org/10.5194/nhess-15-2485-2015

Kendall, M. (1975). *Rank correlation methods*. Griffin, London.

Kish, A., Samavedam, S.: Track buckling prevention: Theory, safety concepts, and applications. Technical report, John A. Volpe National Transportation Systems Center for FRA (2013)

Koks, E. E., Rozenberg, J., Zorn, C., Tariverdi, M., Vousdoukas, M., Fraser, S. A., Hall, J. W., & Hallegatte, S. (2019). A global multi-hazard risk analysis of road and railway infrastructure assets. *Nature Communications*, *10*(1), 2677. https://doi.org/10.1038/s41467-019-10442-3

Kotlyakov, V. M., Desinov, L. V., Dolgov, S. V., Koronkevich, N. I., Likhacheva, E. A., Makkaveev, A. N., Medvedev, A. A., & Rudakov, V. A. (2013). Flooding of July 6–7, 2012, in the town of Krymsk. *Regional Research of Russia*, *3*(1), 52–58. https://doi.org/10.1134/S2079970513010061

Lamb, R., Garside, P., Pant, R., & Hall, J. W. (2019). A Probabilistic Model of the Economic Risk to Britain's Railway Network from Bridge Scour During Floods. *Risk Analysis*, *39*, 2457–2478. https://doi.org/10.1111/risa.13361

Landwehr, J., Matalas, N., & Wallis, J. (1978). Some comparisons of flood frequency estimation methods. *Water Resources Research*, *14*(2), 329–340.

Landwehr, J., Matalas, N., & Wallis, J. (1980). Confidence intervals for design events from sample data. *Journal of the Hydraulics Division*, *106*(HY4), 623–642.

Liu, K., Wang, M., Cao, Y., Zhu, W., Wu, J., & Yan, X. (2018). A Comprehensive Risk Analysis of Transportation Networks Affected by Rainfall-Induced Multihazards. In *RISK ANALYSIS* (Vol. 38, Issue 8, pp. 1618–1633). WILEY. https://doi.org/10.1111/risa.12968

Liu, K., Wang, M., Cao, Y., Zhu, W., & Yang, G. (2018). Susceptibility of existing and planned Chinese railway system subjected to rainfall-induced multi-hazards. *Transportation Research Part A: Policy and Practice*, *117*, 214–226. https://doi.org/10.1016/j.tra.2018.08.014

Mattsson, L.-G., & Jenelius, E. (2015). Vulnerability and resilience of transport systems – A discussion of recent research. *Transportation Research Part A: Policy and Practice*, *81*, 16–34. https://doi.org/10.1016/j.tra.2015.06.002

Medina-Cobo, M. T., García-Marín, A. p., Estévez, J., & Ayuso-Muñoz, J. l. (2016). The identification of an appropriate Minimum Inter-event Time (MIT) based on multifractal characterization of rainfall data series. *Hydrological Processes*, *30*(19), 3507–3517. https://doi.org/10.1002/hyp.10875

Munich Re. (2017). *TOPICS GEO: Natural catastrophes 2016 – Analyses, assessments, positions*. Munich Reinsurance Company. https://www.munichre.com/

National Centers for Environmental Information. (2024). *National Geophysical Data Center*. https://www.ngdc.noaa.gov/

National Centers for Environmental Information (NCEI). (n.d.). *Storm Events Database (CSV Archive)*. NOAA National Centers for Environmental Information. https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/

National Oceanic and Atmospheric Administration (NOAA). (2024). *Precipitation Frequency Data Server (PFDS)*. https://hdsc.nws.noaa.gov/pfds/

Nocedal, J., & Wright, S. J. (2006). *Numerical Optimization* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-40065-5

Norfolk Southern Railway. Industrial-Track Specifications. 2022.

Norfolk Southern Railway. Geohazard Mapping and Slope-Stability Partnerships. 2023.

Ochsner, M., Palmqvist, C.-W., Olsson, N. O. E., & Hiselius, L. W. (2023). The effects of flooding on railway infrastructure: A literature review. *Transportation Research Procedia*, *72*, 1786–1791. https://doi.org/10.1016/j.trpro.2023.11.654

Organization, W. M. (2009). *Manual on estimation of probable maximum precipitation (PMP)* (No. 1045). WMO.

OptaSense. Fibre-Optic Slide and Rockfall Detection Systems. 2019.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, *3*, 119–131.

Reis, D. S., & Stedinger, J. R. (2005). Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology*, *313*(1–2), 97–116.

Sam, C., M. G. ,. Nwaogazie, I. L. ,. &. Ikebude. (2023). General Extreme Value Fitted Rainfall Non-Stationary Intensity-Duration-Frequency (NS-IDF) Modelling for Establishing Climate Change in Benin City. *Journal of Water Resource and Protection*, *15*, 194–214. https://doi.org/10.4236/jwarp.2023.155012

Schanze, J. (2018). Pluvial flood risk management: An evolving and specific field. *Journal of Flood Risk Management*, *11*, 227–229. https://doi.org/10.1111/jfr3.12295

Smith, R. L. (1984). Threshold Methods for Sample Extremes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *46*(3), 393–420.

Srivastava, K., Köpke, C., Walter, J., Faist, K., Berry, J. M., Porretti, C., & Stolz, A. (2023). Modelling and Simulation of Railway Networks for Resilience Analysis. In S. Katsikas, F. Cuppens, C. Kalloniatis, J. Mylopoulos, F. Pallas, J. Pohle, M. A. Sasse, H. Abie, S. Ranise, L. Verderame, E. Cambiaso, J. Maestre Vidal, M. A. Sotelo Monge, M. Albanese, B. Katt, S. Pirbhulal, & A. Shukla (Eds.), *Computer Security. ESORICS 2022 International Workshops* (pp. 308–320). Springer International Publishing. https://doi.org/10.1007/978-3-031-25460-4_17

Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency Analysis of Extreme Events. *Handbook of Hydrology*, 18.1-18.66.

Tu, A., Zeng, J., Liu, Z., Zheng, H., & Xie, S. (2023). Effect of minimum inter-event time for rainfall event separation on rainfall properties and rainfall erosivity in a humid area of southern China. *Geoderma*, *431*, 116332. https://doi.org/10.1016/j.geoderma.2023.116332

Union Pacific and BNSF Railway. *Joint Temporary Shoring Guidelines*. 2022.

U.S. Geological Survey. (2024). *Hydrologic Unit Maps*. https://nas.er.usgs.gov/hucs.aspx

Varra, G., Della Morte, R., Tartaglia, M., Fiduccia, A., Zammuto, A., Agostino, I., Booth, C. A., Quinn, N., Lamond, J. E., & Cozzolino, L. (2024). Flood Susceptibility Assessment for

Improving the Resilience Capacity of Railway Infrastructure Networks. *Water*, *16*(18), Article 18. https://doi.org/10.3390/w16182592

Wang, H., Staes, B. M., & Ghoreishi, B. (2025). Developing an integrated risk assessment framework to quantify the resilience of critical railway infrastructure: Characterization of risks and hazards to the U.S. rail network through empirical data and past events. Mineta Transportation Institute. https://doi.org/10.31979/mti.2025.2413

Wang, W., Yin, S., Xie, Y., Nearing, M. A., Wang, W., Yin, S., Xie, Y., & Nearing, M. A. (2019). Minimum Inter-Event Times for Rainfall in the Eastern Monsoon Region of China. *Transactions of the ASABE*, *62*(1), 9–18. https://doi.org/10.13031/trans.12878

Wasowski, J. and Bovenga, F. *Investigating landslides and unstable slopes with satellite multi-temporal interferometry: Current issues and future perspectives*. *Engineering Geology*, 174: 103–138, 2014. https://doi.org/10.1016/j.enggeo.2014.03.003

Wilby, R. L., Wasko, C., Mearns, L. O., & Newman, A. J. (2023). Uncertainty in Return Periods of Extreme Precipitation Using Large Ensemble Climate Simulations. *Frontiers in Climate*, *6*. https://doi.org/10.3389/fclim.2023.1343072

Yue, S., & Wang, C. (2004). The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. *Water Resources Management*, *18*, 201–218. https://doi.org/10.1023/B:WARM.0000043140.61082.60

Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, *23*(4), 550–560.

Zhang, Y.J., Clemenzi, J., Kesler, K. and Lee, S. Real Time Prediction of Rail Temperature. AREMA Annual Conference, September 9-12, 2007, Chicago, IL

Appendix A. State Based PCA results

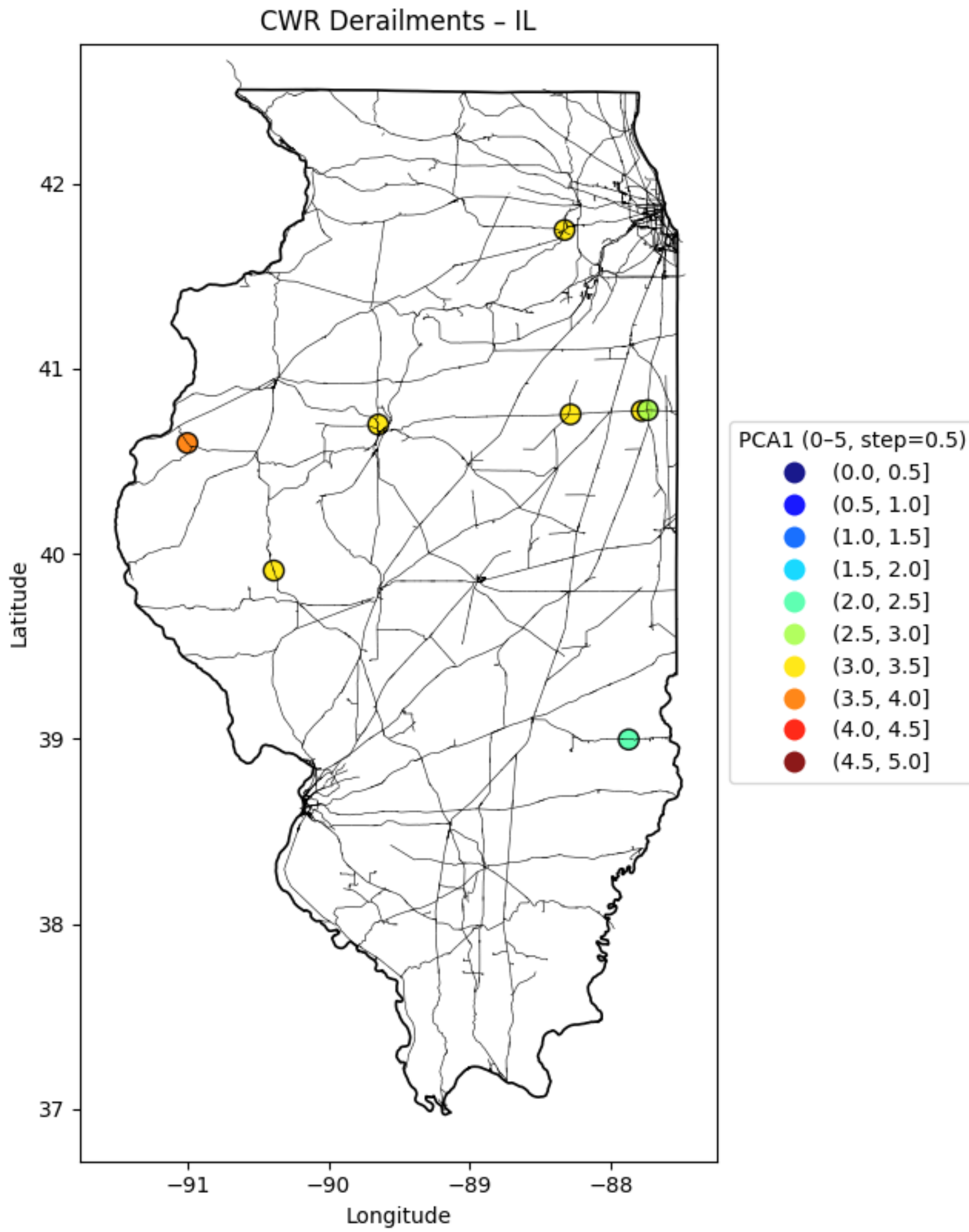

Figure 35. PCA results Illinois

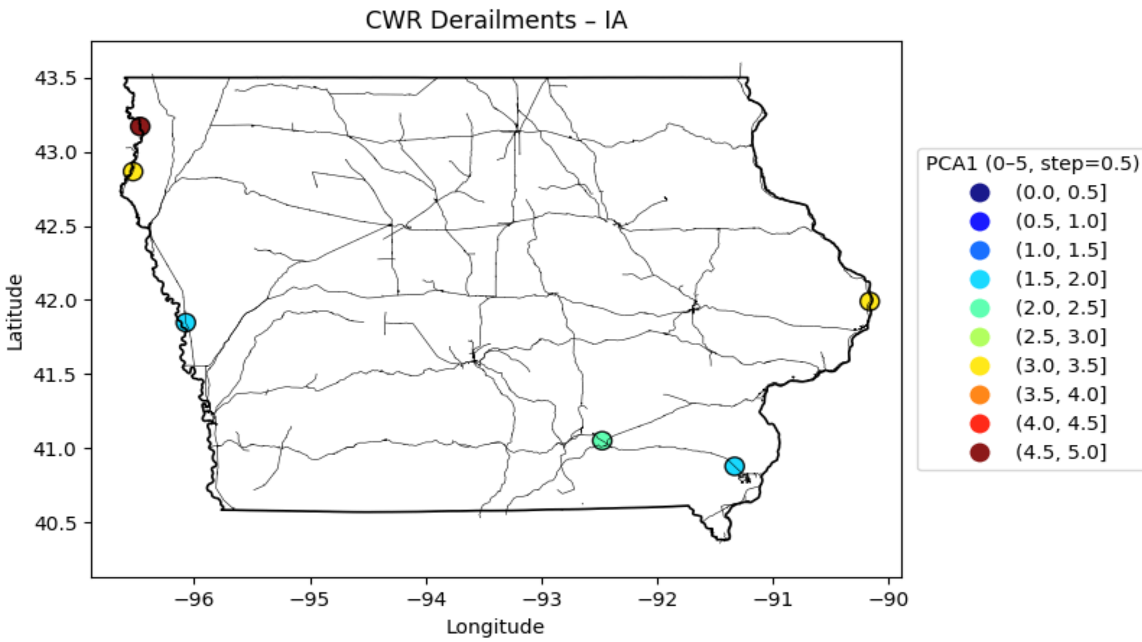

Figure 36. PCA results Iowa

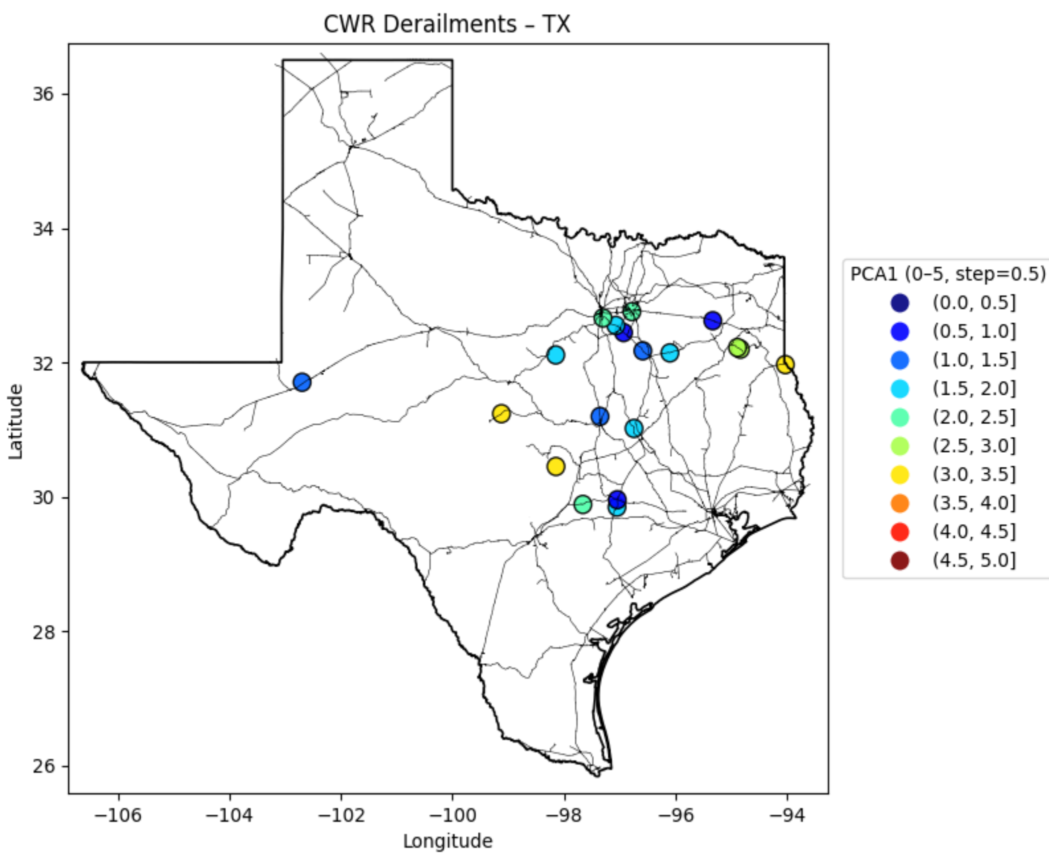

Figure 37. PCA results Texas

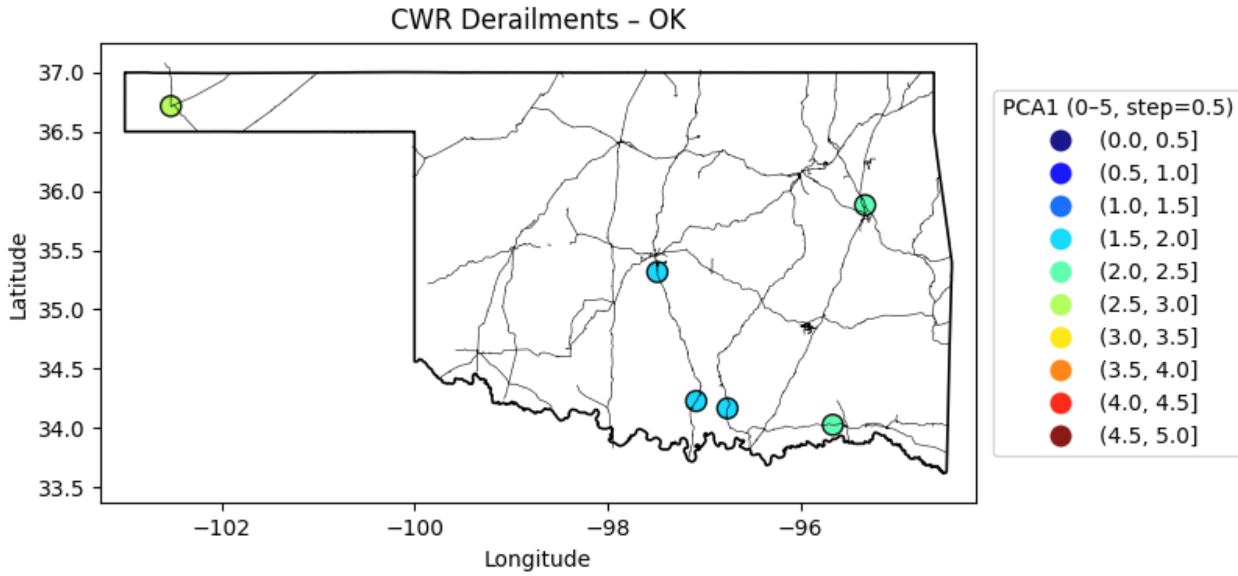

Figure 38. PCA results Oklahoma

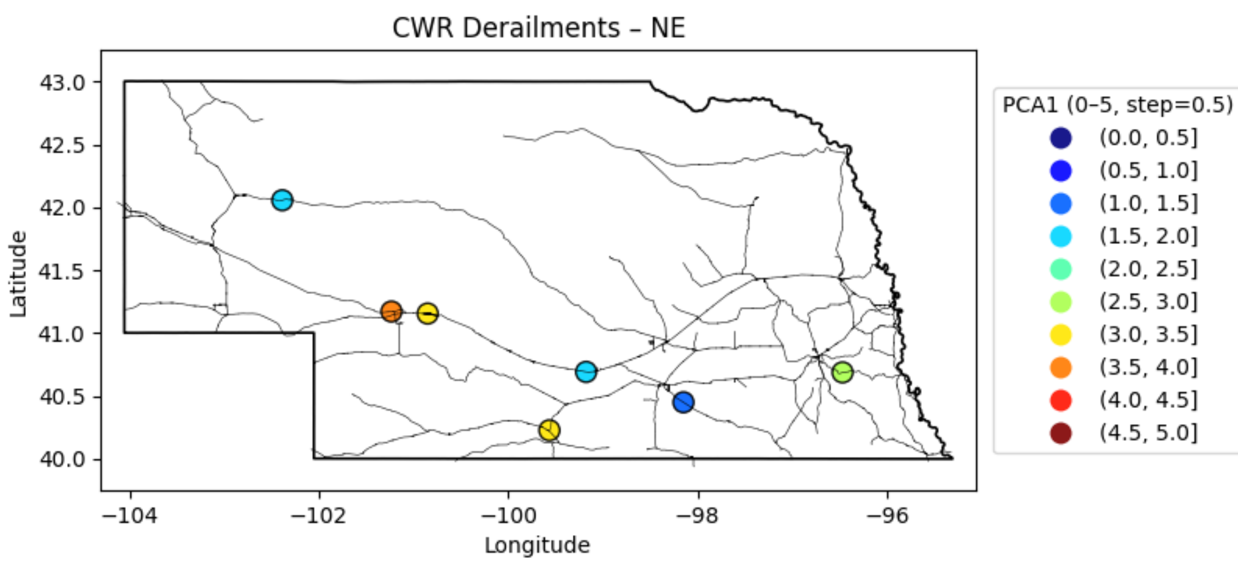

Figure 39. PCA results Nebraska

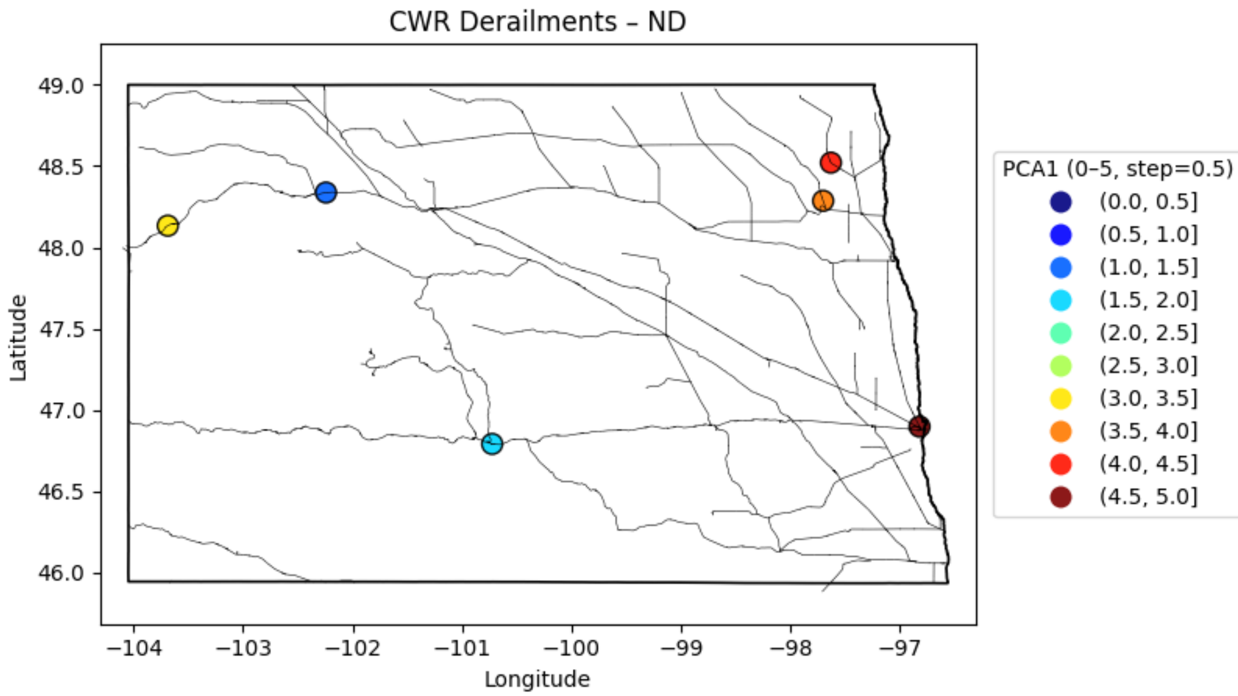

Figure 40. PCA results North Dakota

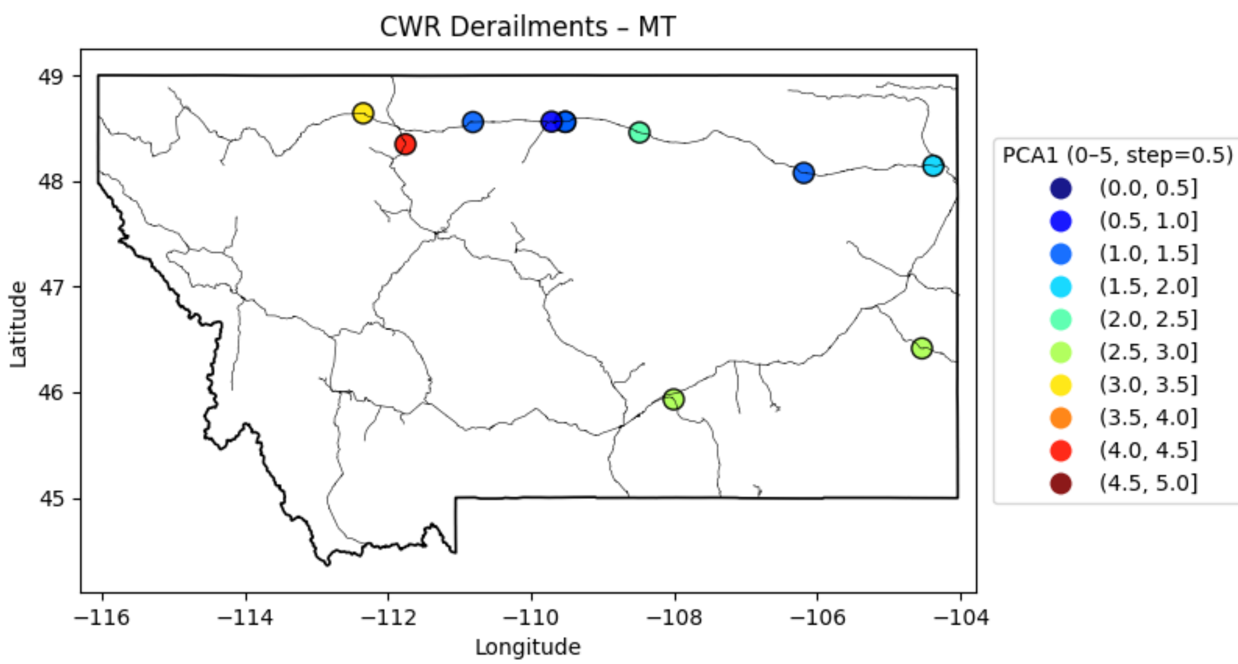

Figure 41. PCA results Montana

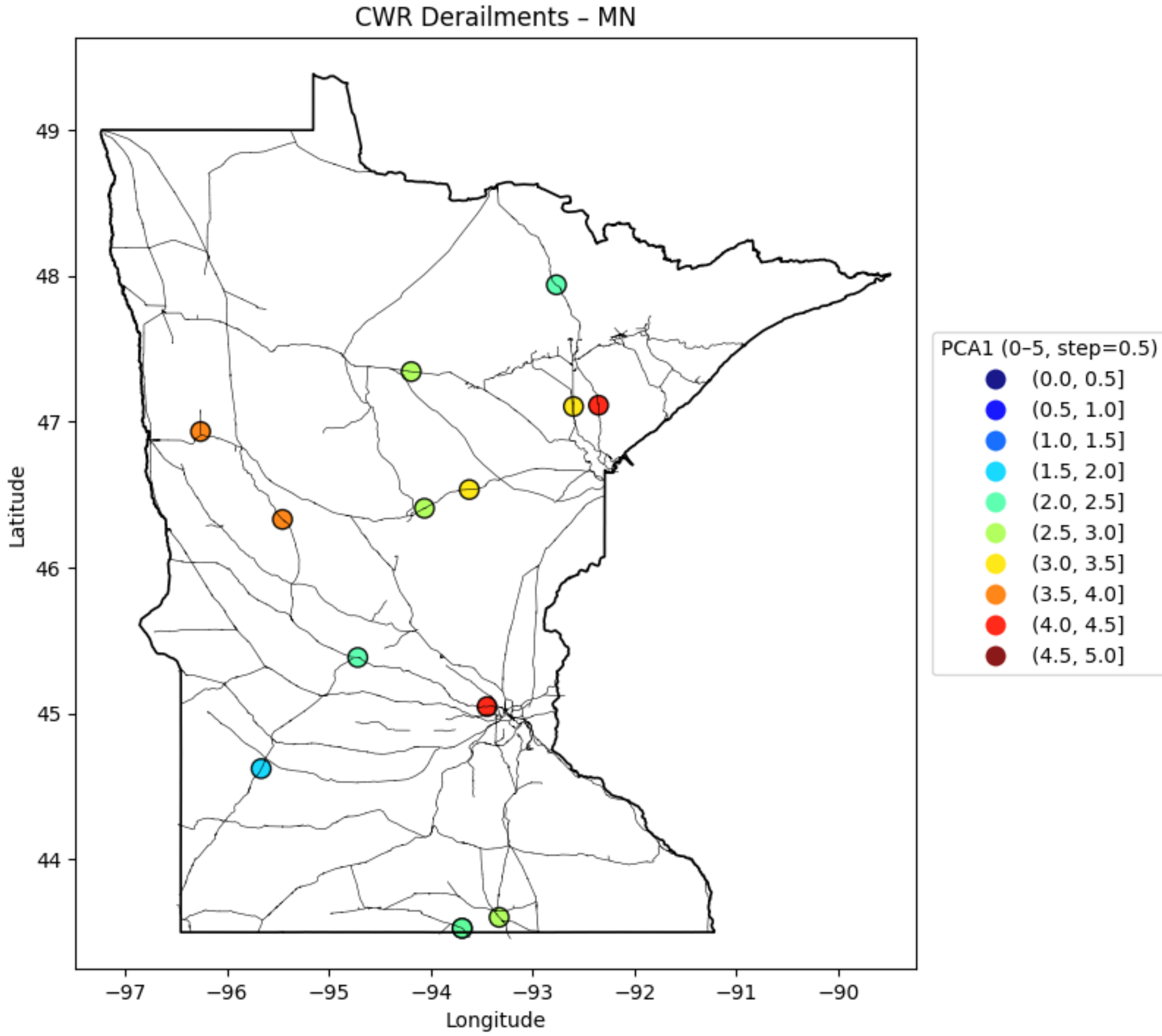

Figure 42. PCA results Minnesota

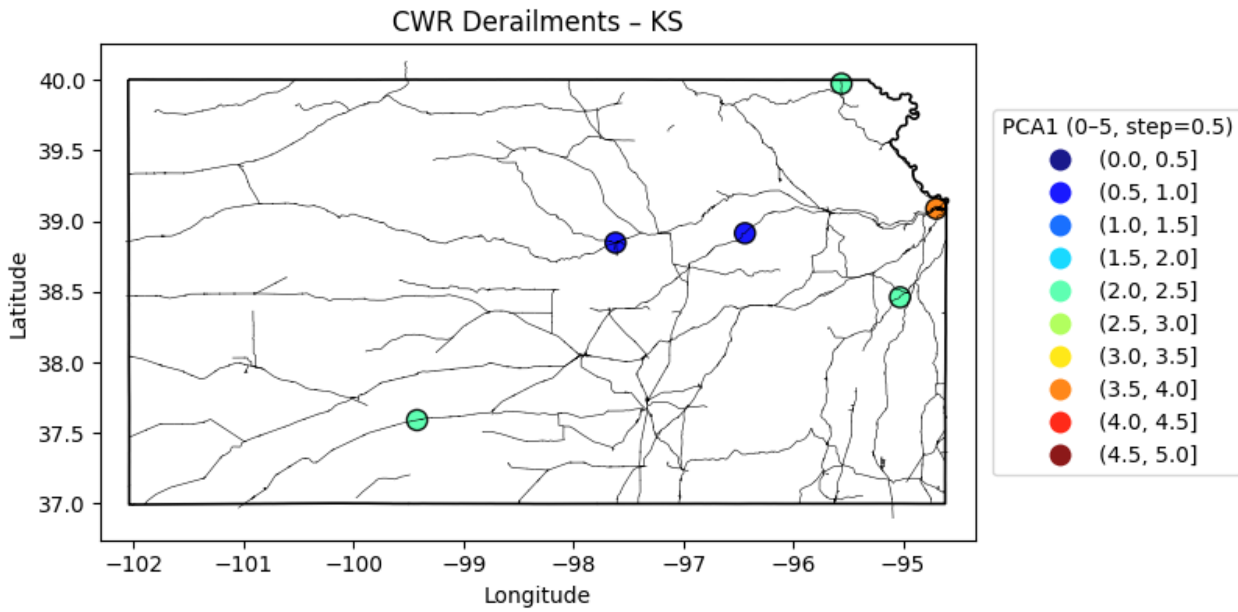

Figure 43. PCA results Kansas

Abbreviations and Acronyms

ACRONYM	DEFINITION
AREMA	American Railway Engineering and Maintenance-of-Way Association
BTS	Bureau of Transportation Statistics
BNSF	Burlington Northern Santa Fe
CN	Canadian National
CPKC	Canadian Pacific Kansas City
DOE	Department of Energy
FRA	Federal Railroad Administration
NCEI	National Centers for Environmental Information
NARN	North American Rail Network
NOAA	National Oceanic and Atmospheric Administration
NREL	National Renewable Energy Laboratory
NSRDB	National Solar Radiation Database
NWS	National Weather Service
USGS	United States Geological Survey

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors

MINETA TRANSPORTATION INSTITUTE