AI and Data Science Journal
ISSN: 3078-7394
DOI: https://doi.org/10.61784/adsj3037

DATA-DRIVEN ASSESSMENT OF PAVEMENT PERFORMANCE AND RESILIENCE UNDER EXTREME WEATHER USING THE LONG-TERM PAVEMENT PERFORMANCE PROGRAM

ShuXin Zhang[1,2*], Lei Qiu[1,2]
[1]*University of California, CA 94720, Berkeley, USA.*
[2]*Ningbo University of Technology, Ningbo 315048, Zhejiang, China.*
[*]*Corresponding Author: ShuXin Zhang*

Abstract: Extreme weather driven by a changing climate is placing unprecedented stress on road infrastructure, yet the quantitative link between weather extremes and observed pavement deterioration—and the resilience of pavements to that stress—remains poorly characterized at scale. This paper presents an end-to-end, reproducible framework that couples the U.S. Federal Highway Administration Long-Term Pavement Performance (LTPP) database with interpretable machine learning and a novel pavement resilience formulation. From 376 asphalt-concrete test sections monitored between 1988 and 2016, we construct a panel of 2,656 section-year records enriched with engineered extreme-weather indicators (freeze–thaw cycles, extreme-heat and freezing days, intense-precipitation days, and thermal range). A variance decomposition shows that 48% of the variance in surface cracking is attributable to unobserved site-specific factors, explaining why absolute distress in unseen sections is intrinsically hard to predict from climate alone (transfer $R^2 \approx 0$). A linear mixed-effects model that controls for site and pavement age isolates intense-precipitation days (standardized $\beta = 1.10$, $p < 0.001$) and extreme-heat days ($\beta = 0.70$, $p = 0.043$) as the statistically significant weather accelerators of cracking, whereas a condition-informed forecasting model reaches $R^2 = 0.53$ with prior condition dominating the SHAP attribution. We introduce a Climate Stress Index (CSI) and a two-dimensional stress–impact positioning that yields a Demonstrated Resilience Score, separating climate-driven vulnerability from intrinsic fragility. The framework offers actionable screening for climate-adaptive pavement asset management.
Keywords: Pavement performance; Climate resilience; Extreme weather; Machine learning; Explainable AI; SHAP; LTPP; Freeze–thaw; Infrastructure asset management

1 INTRODUCTION

Road pavements are among the most extensive and capital-intensive assets that societies maintain, and their serviceability is tightly coupled to the climate in which they operate. Asphalt binders soften at high temperatures and stiffen at low temperatures; subgrades weaken as moisture infiltrates; and repeated freezing and thawing damages the pavement structure. As anthropogenic climate change intensifies the frequency and severity of extreme-weather events—heat waves, intense precipitation, and disrupted freeze cycles—transportation agencies increasingly recognize that engineering standards calibrated to a stationary climate may no longer be adequate [1-2]. Understanding how weather extremes translate into measurable pavement deterioration, and how resilient individual pavements are to that stress, is therefore central to climate-adaptive infrastructure planning.

The dominant line of research on climate and pavements has relied on mechanistic-empirical simulation, typically the Mechanistic-Empirical Pavement Design Guide (MEPDG) or AASHTOWare Pavement ME, driven by downscaled climate projections [3-6]. These studies have produced valuable forecasts—for example, that rutting will worsen with warming while low-temperature cracking will ease in previously cold regions [3,7], and that inappropriate binder selection under a stationary-climate assumption could add tens of billions of dollars to U.S. maintenance costs [1]. However, simulation-based results inherit the assumptions and calibration of the underlying models and are only indirectly validated against long-term field observation. In parallel, a rapidly growing body of work applies machine learning (ML) to the LTPP database to predict roughness, rutting, and cracking [8-15]. These models achieve strong in-sample accuracy, but they are frequently evaluated with random train–test splits that leak information across repeated measurements of the same section, and they rarely quantify resilience or interrogate which weather extremes matter and why.

Two gaps motivate this study. First, the observational relationship between engineered extreme-weather indicators and field-measured distress—properly separated from site-specific confounders and honestly evaluated for transfer to unseen pavements—has not been systematically established. Second, although quantitative resilience frameworks are mature in seismic and network contexts [16], a pavement-specific formulation that normalizes observed deterioration by climate-stress exposure is largely absent. This paper addresses both gaps using only real, publicly available data. Our contributions are:

1) a reproducible pipeline that transforms raw LTPP climate, distress, and traffic records into an extreme-weather-indexed pavement performance panel (376 sections, 2,656 section-year observations);
2) a rigorous statistical analysis—variance decomposition and a linear mixed-effects model—that quantifies how weather extremes relate to cracking while controlling for site heterogeneity, identifying moisture and heat extremes as the dominant significant drivers;

3) an interpretable, honestly benchmarked forecasting model (with group-aware cross-validation and SHAP), contrasting the near-impossibility of cross-site transfer with the practicality of condition-informed forecasting; and
4) a novel Climate Stress Index and two-dimensional stress–impact resilience framework that produces a Demonstrated Resilience Score and separates climate-driven vulnerability from intrinsic fragility.
Figure 1 summarizes the framework. All code and derived datasets are released to support reproducibility.

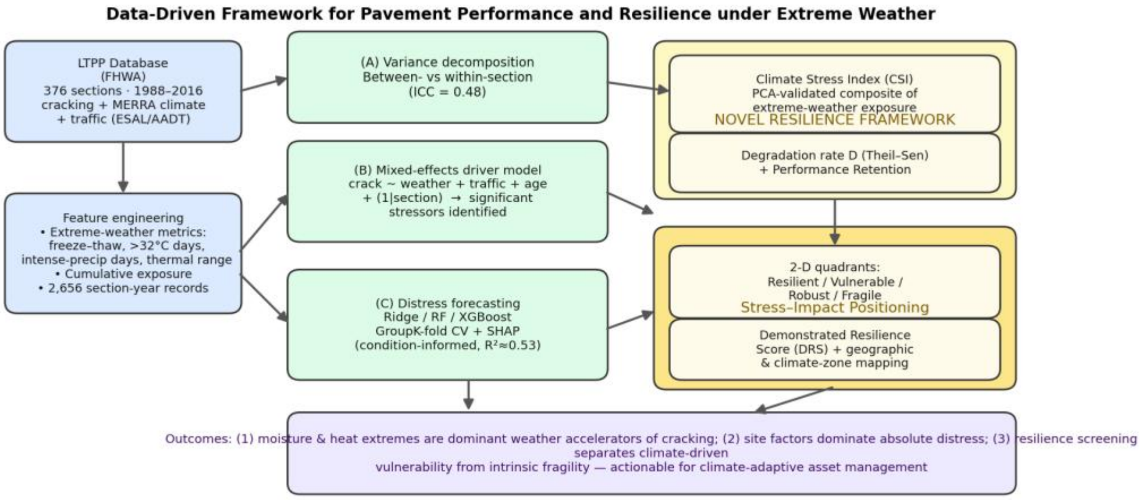

Figure 1 Overview of the Framework: Raw LTPP Climate, Distress, and Traffic Records are Transformed into an Extreme-Weather-Indexed Section-Year Panel, Analyzed for Drivers and Forecastability under Leakage-Free Evaluation, and Distilled into the Climate Stress Index and the two-dimensional Demonstrated Resilience Score

2 RELATED WORK

2.1 Climate Change and Pavement Performance

Early assessments established that pavement deterioration is sensitive to climatic loading. Mills et al. analyzed 17 southern-Canadian sites and simulated midcentury scenarios [3], concluding that shorter freeze seasons would reduce low-temperature cracking while higher in-service temperatures would raise rutting potential; Tighe et al. reached compatible conclusions for low-volume roads [6]. Meagher et al. and Qiao et al. [2,4,17],formalized methodologies for propagating climate projections through mechanistic models and evaluating life-cycle-cost consequences, and Qiao et al. later extended the economic analysis to 24 U.S. case studies [18]. Underwood et al. [1], using an ensemble of 19 global climate models, estimated that stationary-climate binder selection was already inappropriate for 35% of examined locations and could add up to US$26.3 billion in maintenance costs by 2040. Stoner et al. quantified lifetime reductions across the U.S. [7], and Lu et al. specifically examined the risk posed by future extreme-precipitation events—one of the few studies to foreground precipitation extremes [19], which our results reinforce. Collectively, this literature is simulation-centric and forward-looking; it rarely mines the historical field record directly for the empirical weather–distress relationship.

2.2 Machine Learning for Pavement Prediction

The LTPP program has become the standard testbed for data-driven pavement modeling [20]. Ker et al. developed fatigue-cracking models from LTPP as early as 2008 [14]. Random-forest regressions have been used to predict the International Roughness Index (IRI) and to probe mixture-property effects [8-9]; Georgiou et al. compared soft-computing roughness models [13]; Marcelino et al. proposed a generalizable ML workflow [12]; and gradient-boosting approaches, including XGBoost and gradient-boosted decision trees [11,21-22], now dominate reported accuracy. Damirchilo et al. applied XGBoost to 12,637 LTPP observations and identified traffic and material gradation as leading predictors [10]. Titus-Glover used ML with NASA MERRA-2 reanalysis to reassess pavement climate zones [15]. A recurring but under-examined issue is evaluation protocol: because LTPP contains many repeated measurements per section, random splitting inflates apparent accuracy. Studies that segment by climate zone report that the initial condition dominates prediction—an observation consistent with the SHAP attribution reported here.

2.3 Infrastructure Resilience

Quantitative resilience originates largely in the seismic and network-systems literature. Bruneau et al. defined resilience through robustness [16], redundancy, resourcefulness, and rapidity, and proposed measuring loss of resilience as the area between an ideal and a degraded performance trajectory—the "resilience triangle." While this integral-of-performance idea transfers naturally to pavement condition over time, existing pavement studies seldom normalize observed deterioration by the environmental stress a section actually experiences. Our framework adapts the performance-retention concept and augments it with an explicit climate-stress normalization, enabling comparison of pavements across dissimilar climates.

3 DATA AND METHODS

3.1 LTPP Data and Study Sections

We use processed records derived from the FHWA LTPP program [20,23], which has monitored in-service pavements across the United States and Canada since 1987. Distress data are the MEPDG-style asphalt-concrete (AC) cracking percentage recorded at successive surveys. Climate data originate from the LTPP MERRA-based climatic database and are provided as annual summaries per virtual weather station, including temperature statistics, freezing and freeze–thaw indices, precipitation, and snowfall. Traffic is represented by annual average daily traffic (AADT), truck AADT, and annual equivalent single-axle loads (ESAL, in thousands). We retain sections with at least three cracking observations and at least three years of matched annual climate, yielding 376 AC sections and 2,656 section-year records spanning 1988–2016. Section geolocation (latitude, longitude, elevation) is preserved for spatial analysis.

3.2 Extreme-Weather Feature Engineering

For each section-year we assemble a suite of indicators chosen to represent physically distinct weather stressors (Table 1). Thermal extremes are captured by the annual counts of days above 32°C and below 0°C, the freezing index, the annual freeze–thaw cycle count, and the annual thermal range (maximum minus minimum temperature). Moisture stressors comprise total annual precipitation, the count of intense-precipitation days, wet days, and total snowfall. To represent damage accumulation we additionally compute cumulative exposures from a section's first observation to each survey (e.g., cumulative freeze–thaw cycles and cumulative intense-precipitation days). Pavement age is defined as years elapsed since the first survey. Missing annual values are imputed by the section climatological mean before aggregation.

3.3 Climate-Zone Classification

We classify sections into the four canonical LTPP environmental zones defined by the intersection of a freeze/no-freeze axis and a wet/dry axis. A section is designated Freeze if its mean freezing index exceeds 50 °C-days or its mean minimum temperature falls below −5°C, and Wet if its mean annual precipitation exceeds 508 mm (20 in). The retained sections populate three zones: Wet-Freeze (272 sections), Dry-Freeze (76), and Wet-NoFreeze (28).

3.4 Variance Decomposition and Mixed-Effects Model

To quantify how much of the distress signal is intrinsically site-specific, we compute the between-section share of cracking variance (an intraclass-correlation analogue). We then fit a linear mixed-effects model with a random intercept per section, regressing cracking on standardized annual weather indicators, log-ESAL traffic, and pavement age. Standardizing predictors yields directly comparable coefficients, and the random intercept absorbs unobserved, time-invariant site heterogeneity so that the fixed effects reflect within-section associations. Variance-inflation factors are examined and collinear precipitation variables reduced (retaining intense-precipitation days as the extreme-weather indicator), keeping the maximum VIF below 2.8.

3.5 Distress Forecasting and Evaluation

We consider two prediction regimes that answer different practical questions. In the transfer regime, models predict the cracking level of pavements in held-out sections from weather, traffic, geometry, and age; evaluation uses five-fold GroupK-fold cross-validation with sections as groups, so no section appears in both training and test folds. In the condition-informed forecasting regime—closer to how pavement-management systems operate—the model predicts a section's next cracking measurement given its most recent measurement, the survey interval, and the weather and traffic accrued in between; the same group-aware protocol prevents optimistic leakage while still allowing a section's own history to inform its forecast. We compare a regularized linear model (ridge), random forest [24], gradient boosting [22], and XGBoost [21], and benchmark against a persistence baseline (next = previous). To isolate the marginal value of weather information, we repeat the forecasting experiment with all weather features removed. Model interpretation uses SHAP values via a tree explainer [25-26]; models are implemented in scikit-learn and XGBoost [27-29].

3.6 Climate Stress Index and Resilience Framework

For sections with at least four distinct observation ages we estimate a robust degradation rate D (percentage cracking per year) as the Theil–Sen slope of cracking against age, floored at zero because cracking is physically irreversible. We also compute a performance-retention metric as the time-average of a normalized condition $Q(t) = 1 - \text{cracking}(t)/C$, with a failure ceiling $C = 50\%$, following the resilience-triangle idea [16]. We define a Climate Stress Index (CSI) as the equal-weight mean of the z-scored extreme-weather exposures (freeze–thaw, extreme-heat days, freezing days, intense-precipitation days, thermal range, and snowfall), each oriented so that larger values denote greater stress; principal-component analysis validates the composite. Because a single scalar cannot rank sites of dissimilar climates fairly, we position each section in the two-dimensional CSI–degradation plane and label the four quadrants Resilient (high stress, low degradation), Vulnerable (high stress, high degradation), Fragile (low stress, high degradation), and Robust (low stress, low degradation). Finally we compute a Demonstrated Resilience Score, DRS = CSI-percentile × (1 − degradation-percentile) × 100, which rewards sustained performance under demonstrated high stress, and we verify its robustness against a retention-based variant.

Table 1 Engineered predictors used in the analysis

Group	Features
Thermal extremes	freeze–thaw cycles; days >32°C; days <0°C; freezing index; thermal range; max/min/mean temperature
Moisture	total precipitation; intense-precipitation days; wet days; snowfall
Cumulative	cumulative freeze–thaw, heat, freezing, and intense-precipitation days since first survey
Traffic	AADT; truck AADT; annual ESAL; cumulative ESAL
Structure/site	pavement age; elevation; latitude
Target	MEPDG AC cracking (%)

4 RESULTS

4.1 Descriptive Statistics and Variance Structure

Across the 2,656 records the AC cracking distribution is right-skewed and zero-inflated: 46% of observations show no cracking, while the upper decile exceeds 20%. Figure 2 plots cracking against pavement age by climate zone; all zones show the expected upward progression, but the trajectories are noisy and overlapping, foreshadowing weak univariate signal. Indeed, no single predictor correlates strongly with cracking—the largest Spearman magnitude at the observation level is 0.18 (cumulative intense-precipitation days). The variance decomposition is decisive: 48% of the total cracking variance lies between sections (intraclass correlation 0.48 from the mixed model). Nearly half of the distress signal is therefore attributable to unobserved, time-invariant site factors—construction quality, subgrade support, mixture design, and maintenance history—rather than to climate or traffic. This structural fact governs everything that follows.

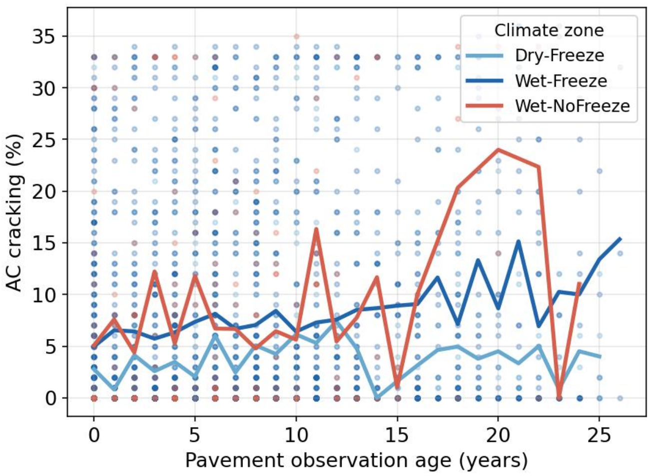

Figure 2 AC Cracking Versus Pavement Age, Colored by Climate Zone

Note: Trajectories rise with age but overlap substantially, foreshadowing weak univariate signal.

4.2 Drivers of Cracking: Mixed-Effects Model

Controlling for site via random intercepts and for pavement age, the mixed-effects model (Table 2) isolates the weather variables that are significantly associated with cracking. Pavement age is, unsurprisingly, the strongest term ($\beta = 1.74$, $p < 0.001$). Among weather stressors, intense-precipitation days is the dominant and highly significant accelerator ($\beta = 1.10$, $p < 0.001$), followed by extreme-heat days ($\beta = 0.70$, $p = 0.043$); snowfall is marginal ($\beta = 0.58$, $p = 0.07$), while freeze–thaw cycles and thermal range are not significant once site and age are controlled. This is a physically coherent result: moisture infiltration degrades structural support and, together with thermal softening under extreme heat, drives surface cracking, whereas freeze effects—often assumed dominant—are confounded with region-specific design and traffic in this sample. The finding echoes the specific concern raised by Lu et al. about extreme precipitation [19].

TABLE 2 Standardized Fixed-Effects (Mixed Model; Random Section Intercept, ICC = 0.48)

Predictor	β (std.)	SE	p
Pavement age	+1.74	0.16	<0.001
Intense-precip. days	+1.10	0.29	<0.001

Predictor	β (std.)	SE	p
Extreme-heat days (>32°C)	+0.70	0.35	0.043
Snowfall	+0.58	0.32	0.070
Thermal range	+0.35	0.36	0.328
Traffic (log ESAL)	−0.37	0.32	0.247
Freeze–thaw cycles	−0.21	0.40	0.590

4.3 Predictive Modeling: Two Regimes

The transfer regime confirms the implication of the variance structure. Predicting cracking in entirely unseen sections from weather, traffic, geometry, and age is essentially infeasible with these features: the best model (ridge) attains only $R^2 = 0.02$, and the tree ensembles do not improve on it (Table 3, upper block). This is not a modeling deficiency but a statement about the data—because half the variance is site-specific and site identity is withheld, there is little to transfer. Reporting such a result honestly, under leakage-free evaluation, is itself a contribution, given how often inflated random-split accuracies appear in the literature.

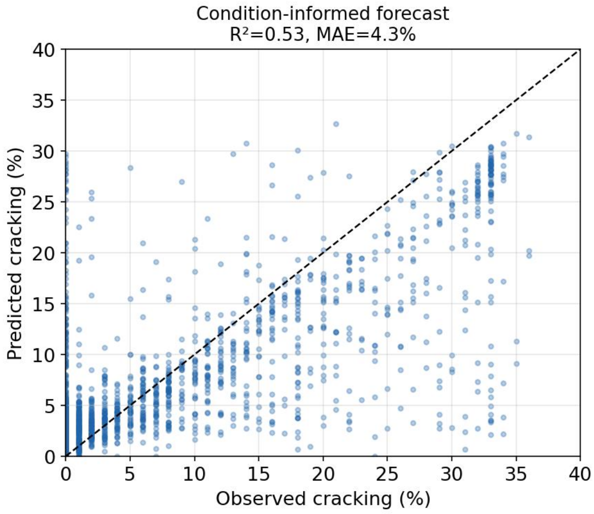

Figure 3 Condition-Informed Forecasting: Predicted Versus Observed Next-Survey Cracking (Ridge, Group-Aware CV), $R^2 = 0.53$.

The condition-informed forecasting regime is markedly more useful (Table 3, lower block; Figure 3). Once a section's most recent cracking measurement is available, ridge regression forecasts the next measurement with $R^2 = 0.53$ and a mean absolute error of 4.3%, comfortably beating the persistence baseline ($R^2 = 0.47$). Crucially, an ablation that removes all weather features leaves accuracy essentially unchanged ($R^2 = 0.51$ vs. 0.50 for XGBoost with weather). The interpretation is that, at the short horizon between consecutive surveys, the accumulated effect of weather is already encoded in the current condition; incremental weather adds little marginal predictive value. Extreme weather shapes the long-run level and rate of deterioration—as Sections B and E show—rather than the next incremental step given current condition.

Table 3 Cross-Validated Performance (GroupK-Fold by Section)

Model	R^2	RMSE	MAE
Transfer: ridge	0.02	9.54	7.40
Transfer: random forest	0.00	9.63	7.47
Transfer: gradient boosting	−0.06	9.93	7.59
Transfer: XGBoost	−0.08	10.02	7.63
Forecast: persistence	0.47	7.28	3.57
Forecast: ridge	0.53	6.84	4.34
Forecast: random forest	0.52	6.88	4.39
Forecast: XGBoost	0.50	7.04	4.53
Forecast: XGBoost, no weather	0.51	—	—

4.4 Model Interpretation with SHAP

SHAP attribution for the forecasting model (Figure 4) makes the condition-dominance explicit: the prior cracking measurement carries a mean absolute SHAP value of 5.5 percentage points, an order of magnitude larger than any other feature. The next tier is instructive, however—it is populated by weather and geography: the incremental freeze–thaw exposure, elevation, snowfall, incremental precipitation, latitude, and incremental heat days each contribute on the order of 0.3–0.5 points. Thus, after conditioning on current state, the residual predictable structure that remains is largely climatic. Figure 5 shows the SHAP dependence for cumulative intense-precipitation exposure: contributions rise with exposure and are modulated by the section's prior condition, consistent with moisture accelerating deterioration most in already-distressed pavements. These interpretations are associational and, given correlated predictors, should be read as descriptive of the fitted model rather than as causal effects.

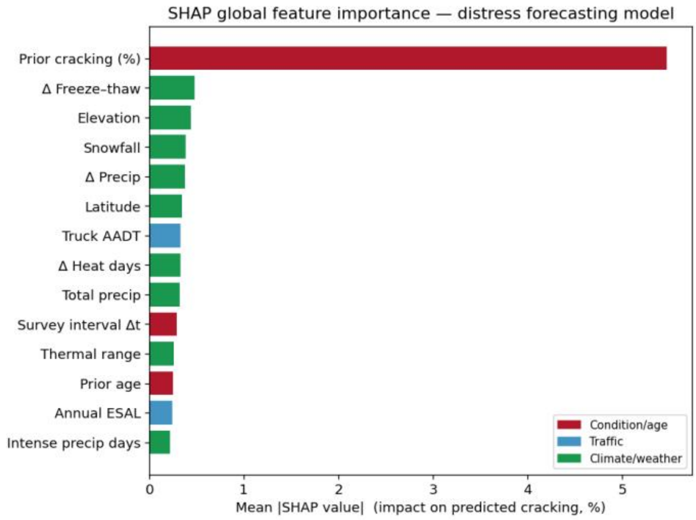

Figure 4 Global SHAP Importance for the Forecasting Model
Note: Prior condition dominates by an order of magnitude; the secondary tier is climatic and geographic.

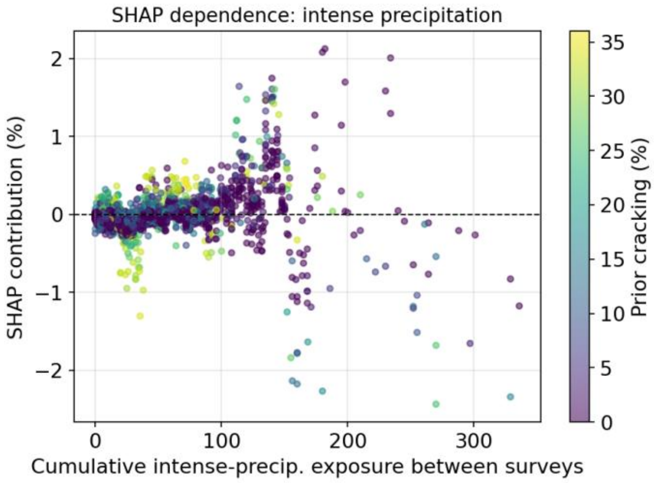

Figure 5 SHAP Dependence for Cumulative Intense-Precipitation Exposure, Modulated by the Section's Prior Condition

4.5 Climate Stress Index and Resilience

The Climate Stress Index is validated by principal-component analysis: the first component explains 67% of the variance in the six extreme-weather exposures and the second a further 17%. The loadings are physically interpretable—PC1 contrasts cold-continental stressors (freeze–thaw, freezing days, thermal range, snowfall) against warm-humid stressors (extreme heat, intense precipitation), i.e., a thermal-regime axis, while PC2 aligns with the moisture/heat gradient. The equal-weight CSI correlates 0.88 with PC1, confirming it captures the dominant stress gradient. Estimated degradation rates are modest for most sections (median 0.1%/yr) but heavy-tailed, reaching 5.3%/yr for the fastest-failing pavement. A central—and honest—finding is that CSI alone is only weakly associated with degradation rate (Spearman $\rho = -0.05$, $p = 0.38$). Because site factors dominate, harsher aggregate climate does not, by itself, predict faster deterioration in this sample. Rather than discard the resilience concept, we exploit this by positioning sections in the two-dimensional stress–impact plane (Figure 6). The four quadrants are populated in near-balance—Robust (87 sections), Vulnerable (84), Resilient (81), and Fragile (78)—and each carries a distinct management meaning. Resilient sections sustain low degradation despite high stress; Vulnerable sections deteriorate rapidly under high stress and are the priority for climate-adaptive intervention; and Fragile sections deteriorate rapidly despite mild climate, signaling intrinsic construction, material, or subgrade defects rather than weather. This separation of climate-driven vulnerability from intrinsic fragility is the framework's key diagnostic value.

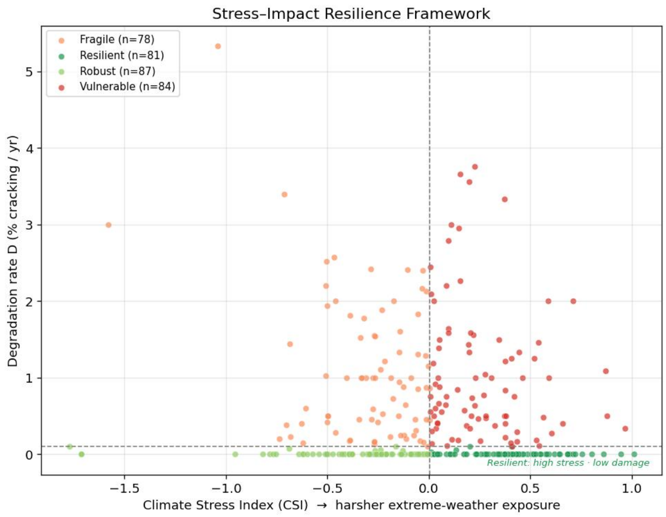

Figure 6 Two-Dimensional Stress–Impact Positioning

Note: The quadrants separate climate-driven vulnerability (Vulnerable) from intrinsic fragility (Fragile).

The Demonstrated Resilience Score condenses the positioning into a 0–100 ranking and is robust to definition: it correlates $\rho = 0.80$ with a retention-based variant. Aggregated by climate zone (Table 4; Figure 7), degradation differs significantly across zones (Kruskal–Wallis $H = 6.30$, $p = 0.043$). Dry-Freeze sections are the most resilient on average (mean DRS 38; 44% classified Resilient), whereas Wet-NoFreeze sections—by construction low-stress—cannot demonstrate resilience (0% Resilient) and contain several Fragile pavements failing under mild conditions. Figure 8 maps the resilience class geographically, revealing spatial clustering that can guide regional prioritization. The most vulnerable sections (degradation 3.3–5.3%/yr under moderate-to-high stress) are concentrated in Wet-Freeze environments, whereas the most resilient sustain near-zero cracking under some of the highest freeze–thaw loads in the network.

Table 4 Demonstrated Resilience by Climate Zone (n = 330 Sections)

Climate zone	n	Mean CSI	Med. D (%/yr)	Resil. (%)
Dry-Freeze	73	+0.23	0.00	43.8
Wet-Freeze	233	+0.02	0.16	21.0
Wet-NoFreeze	24	−0.84	0.17	0.0

5 DISCUSSION

5.1 Implications for Climate-Adaptive Asset Management

The two-dimensional stress–impact positioning translates directly into a screening protocol for agencies. Vulnerable sections—those deteriorating rapidly under demonstrably high climate stress—are the natural priority for climate-adaptive treatments such as stiffer or moisture-resistant binders, improved drainage, and shortened resurfacing cycles; in our sample these concentrate in Wet-Freeze environments with degradation rates of 3.3–5.3%/yr. Fragile sections are diagnostically distinct: they fail quickly despite mild climate, which points to intrinsic causes—construction quality, subgrade support, or mixture design—rather than weather, and therefore warrant forensic investigation rather than climate hardening. Resilient sections, sustaining near-zero cracking under some of the highest freeze–thaw loads in the network, are positive exemplars whose design and maintenance histories merit study and replication. Because the Demonstrated Resilience Score is a single 0–100 number that is robust to its precise definition ($\rho = 0.80$ against a retention-based variant), it can be embedded in a pavement-management dashboard for network-level triage.

5.2 The Dual Role of Weather

Our results reconcile two findings that at first appear contradictory: extreme weather is a statistically significant driver of cracking in the mixed-effects model, yet adding weather features barely improves short-horizon forecasts. The resolution is temporal. Over the multi-year service life, intense precipitation and extreme heat measurably accelerate the level and rate of deterioration—this is what the mixed model and the secondary SHAP tier capture. Over the short interval between consecutive surveys, however, the cumulative imprint of past weather is already embedded in the section's current measured condition, so the marginal weather increment carries little additional predictive value. Extreme weather thus shapes the long-run trajectory rather than the next incremental step given current state—an important distinction for how climate information should enter design (where it matters) versus short-term condition forecasting (where current state dominates).

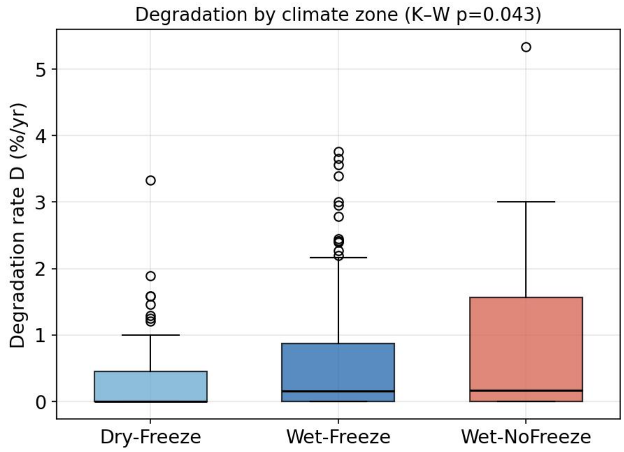

Figure 7 Distribution of Degradation Rate by Climate Zone (Kruskal–Wallis H = 6.30, p = 0.043)

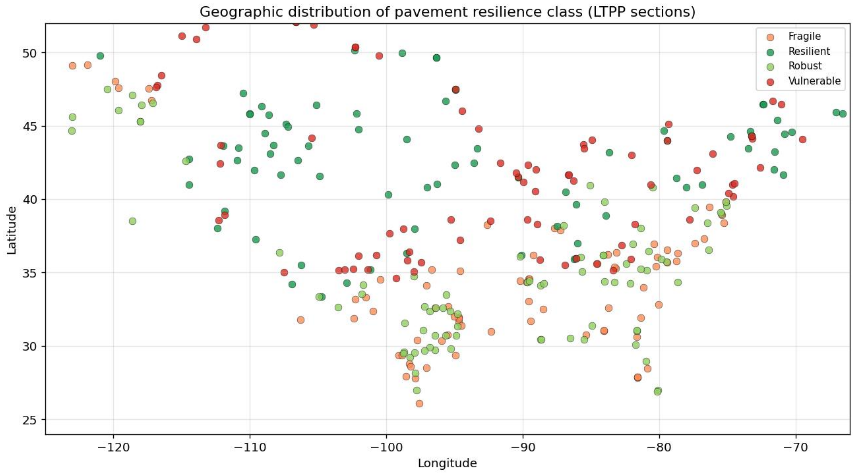

Figure 8 Geographic Distribution of Resilience Classes across the LTPP Sections
Note: Spatial clustering of Vulnerable and Resilient pavements can guide regional prioritization.

5.3 Why Absolute Cross-Site Prediction Is Hard

The variance decomposition—nearly half of cracking variance residing between sections—explains the near-zero transfer performance and cautions against the optimistic accuracies often reported under random train–test splits that leak repeated measurements of the same section. Predicting the absolute distress of an unseen pavement from climate, traffic, geometry, and age alone is intrinsically limited because the dominant signal is site identity: construction, materials, and subgrade. The practical corollary is that meaningful cross-site models require site-level structural and material covariates, not merely richer weather features. Reporting the honest leakage-free result is, we argue, more useful to practitioners than an inflated in-sample figure.

5.4 Limitations

Several limitations bound our conclusions. First, all associations are observational; correlated predictors and unobserved confounders mean the coefficients and SHAP attributions describe the fitted models rather than causal effects. Second, we model a single distress—MEPDG AC cracking; rutting, roughness (IRI), and thermal cracking may respond differently to weather extremes. Third, layer-thickness and air-void records were largely unavailable in the processed data, so we cannot directly attribute Fragile behavior to specific construction defects—only flag it for investigation. Fourth, the climate indicators derive from annual, MERRA-based summaries, which capture seasonal exposure but smooth over sub-daily extremes that may drive damage. Finally, the 376 retained sections skew toward freeze environments and populate only three of the four LTPP zones (no Dry-NoFreeze), and the degradation-rate analysis requires at least four survey ages (330 sections) and assumes an approximately linear Theil–Sen trend; both temper the generality of the zone-level resilience statistics.

5.5 Future Work

Natural extensions include a multi-distress resilience index that jointly considers cracking, rutting, and roughness; quasiexperimental or causal designs (e.g., matched sections or difference-in-differences around extreme events) to move from association toward attribution; enrichment with structural, material, and maintenance covariates to explain the large between-site variance and to formalize the Fragile diagnosis; and coupling the Climate Stress Index with downscaled climate projections to forecast forward-looking Demonstrated Resilience Scores under warming scenarios. Extending the analysis beyond North America would further test the transferability of the framework.

6 CONCLUSION

We presented a reproducible, end-to-end framework linking real Long-Term Pavement Performance observations to extreme-weather exposure through interpretable statistics, honestly benchmarked machine learning, and a novel resilience formulation. Four findings stand out. (i) Nearly half of surface-cracking variance is site-specific, which makes absolute cross-site prediction intrinsically hard and cautions against leakage-inflated accuracies. (ii) Controlling for site and age, intense-precipitation days and extreme-heat days emerge as the significant weather accelerators of cracking, foregrounding moisture and heat over the commonly emphasized freeze–thaw. (iii) Extreme weather governs the long-run level and rate of deterioration rather than the next incremental change, reconciling significant driver effects with the limited marginal value of weather in short-horizon forecasting. (iv) The Climate Stress Index and the two-dimensional stress–impact positioning yield a Demonstrated Resilience Score that separates climate-driven vulnerability from intrinsic fragility and supports network-level triage. By releasing all code and derived datasets, we aim to make climate-resilience screening of pavements both transparent and actionable.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

ACKNOWLEDGMENT

The authors gratefully acknowledge the U.S. Federal Highway Administration Long-Term Pavement Performance program for collecting and maintaining the field data used in this study, and the providers of the processed LTPP dataset for enabling reproducible analysis.

REFERENCES

[1] Underwood BS, Guido Z, Gudipudi P, et al. Increased costs to US pavement infrastructure from future temperature rise. Nature Climate Change, 2017, 7(10): 704-707.
[2] IPCC. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report. Cambridge, UK: Cambridge University Press, 2021.
[3] Mills B, Tighe SL, Andrey J, et al. Climate change implications for flexible pavement design and performance in southern Canada. Journal of Transportation Engineering, 2009, 135(10): 773-782.
[4] Meagher W, Daniel JS, Jacobs J, et al. Method for evaluating implications of climate change for design and performance of flexible pavements. Transportation Research Record, 2012, 2305: 111-120.
[5] Qiao Y, Dawson AR, Parry T, et al. Evaluating the effects of climate change on road maintenance intervention strategies and life-cycle costs. Transportation Research Part D: Transport and Environment, 2015, 41: 492-503.
[6] Tighe SL, Smith J, Mills B, et al. Evaluating climate change impact on low-volume roads in southern Canada. Transportation Research Record, 2008, 2053: 9-16.
[7] Stoner AMK, Daniel JS, Jacobs JM, et al. Quantifying the impact of climate change on flexible pavement performance and lifetime in the United States. Transportation Research Record, 2019, 2673(1): 110-120.
[8] Gong H, Sun Y, Shu X, et al. Use of random forests regression for predicting IRI of asphalt pavements. Construction and Building Materials, 2018, 189: 890-897.
[9] Gong H, Sun Y, Hu W, et al. Investigating impacts of asphalt mixture properties on pavement performance using LTPP data through random forests. Construction and Building Materials, 2019, 204: 203-212.
[10] Damirchilo F, Hosseini A, Mellat Parast M, et al. Machine learning approach to predict international roughness index using long-term pavement performance data. Journal of Transportation Engineering, Part B: Pavements, 2021, 147(4): 04021058.
[11] Guo R, Fu D, Sollazzo G. An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree. International Journal of Pavement Engineering, 2022, 23(10): 3633-3646.
[12] Marcelino P, de Lurdes Antunes M, Fortunato E, et al. Machine learning approach for pavement performance prediction. International Journal of Pavement Engineering, 2021, 22(3): 341-354.
[13] Georgiou P, Plati C, Loizos A. Soft computing models to predict pavement roughness: A comparative study. Advances in Civil Engineering, 2018, 2018: 5939806.
[14] Ker HW, Lee YH, Wu PH. Development of fatigue cracking prediction models using long-term pavement performance database. Journal of Transportation Engineering, 2008, 134(11): 477-482.
[15] Titus-Glover L. Reassessment of climate zones for high-level pavement analysis using machine learning algorithms and NASA MERRA-2 data. Advanced Engineering Informatics, 2021, 50: 101435.

[16] Bruneau M, EERI M, Chang SE, et al. A framework to quantitatively assess and enhance the seismic resilience of communities. Earthquake Spectra, 2003, 19(4): 733-752.
[17] Qiao Y, Flintsch GW, Dawson AR, et al. Examining effects of climatic factors on flexible pavement performance and service life. Transportation Research Record, 2013, 2349: 100-107.
[18] Qiao Y, Guo Y, Stoner AMK, et al. Impacts of future climate change on flexible road pavement economics: A life cycle costs analysis of 24 case studies across the United States. Sustainable Cities and Society, 2022, 80: 103773.
[19] Lu D, Tighe S, Xie WC. Pavement risk assessment for future extreme precipitation events under climate change. Transportation Research Record, 2018, 2672: 122-131.
[20] Federal Highway Administration. Long-Term Pavement Performance (LTPP) InfoPave. U.S. Department of Transportation. https://infopave.fhwa.dot.gov/.
[21] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785-794.
[22] Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics, 2001, 29(5): 1189-1232.
[23] Hajiha M, Liu X. Processed Long-Term Pavement Performance (LTPP) data. Department of Industrial Engineering, University of Arkansas, 2019. https://github.com/dnncode/LTPP-Data.
[24] Breiman L. Random forests. Machine Learning, 2001, 45(1): 5-32.
[25] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS), 2017, 30: 4765-4777.
[26] Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2020, 2(1): 56-67.
[27] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 2011, 12: 2825-2830.
[28] Mallick R, Radzicki M, Daniel JS, et al. Use of system dynamics to understand long-term impact of climate change on pavement performance and maintenance cost. Transportation Research Record, 2014, 2455: 1-9.
[29] Transportation Research Board. Potential Impacts of Climate Change on U.S. Transportation, Special Report 290. Washington, DC: TRB, 2008.