buildings

MDPI

Article

Evaluating the Shared Structure of Pavement Deterioration Across Multiple Pavement Indicators

Benjamin Famewo *, Mehdi Shokouhian, Zeinab Bandpey and David O. Owolabi

Department of Civil and Environmental Engineering, Center for the Built Environment and Infrastructure Studies (CBEIS), Morgan State University, Baltimore, MD 21251, USA; mehdi.shokouhian@morgan.edu (M.S.); zeinab.bandpey@morgan.edu (Z.B.); daowo1@morgan.edu (D.O.O.)
* Correspondence: befam1@morgan.edu

Abstract

Pavement management systems commonly evaluate network condition using a single indicator, most often the International Roughness Index (IRI), although pavement deterioration is expressed through multiple structural and surface distress mechanisms. This study develops a multi-indicator framework combining treatment-adjusted temporal analysis, cross-correlation analysis, and principal component analysis (PCA), applied to 165,372 section-year observations of IRI, rutting depth, structural cracking density, and functional cracking density on 9739 sections in Baltimore County, Maryland, from 2008 to 2024. After excluding observations adjacent to documented treatments, median IRI declined from 110.58 to 103.77 in/mi while rutting and both cracking indicators increased, and functional cracking worsened in 57.7% of directly observed between-treatment intervals. Most of the variation in IRI lay between sections, whereas most of the variation in cracking occurred within sections over time. Cross-correlation peaked at lag zero for every indicator pair in condition levels and after detrending and pre-whitening, and for all pairs except rutting and functional cracking after differencing; the only consistent lead–lag asymmetry, a small tendency for rutting to precede the other indicators, was much weaker than the contemporaneous association. The first principal component (PC1) explained 49.5% of total variance, to which all four indicators contribute substantially (20.7% to 29.4%); this structure was stable under treatment adjustment, across sub-periods, and when variation between and within sections was analyzed separately. Cracking contributed more strongly to the axis on Intermediate roads and IRI on High-Order roads, a difference consistent in direction across all sensitivity analyses. PC1 and an existing minimum-based remaining-service-life measure (RSL_4) agreed on fewer than half (45.2%) of the observations either flagged as high priority, with PC1 favoring multi-indicator distress profiles. Single-indicator and minimum-based approaches therefore capture different aspects of pavement condition, supporting class-sensitive, multi-indicator assessment for network-level pavement management.

Keywords: pavement deterioration; pavement management systems (PMS); multi-indicator analysis; network-level analysis time-series data; principal component analysis (PCA); functional class stratification

Check for updates

Academic Editor: An Cheng

Received: 14 September 2026
Revised: 16 October 2026
Accepted: 22 October 2026
Published: 27 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Transportation agencies monitor pavement condition to support maintenance planning and budget allocation. The International Roughness Index (IRI) is the most widely used condition indicator in practice. It is simple to measure, well-standardized, and easily

incorporated into pavement management systems. However, IRI only reflects ride quality. It does not measure cracking, permanent deformation, or degradation of the surface texture. Agencies that rely on IRI alone risk missing deterioration that has not yet affected roughness but has already advanced in other forms of distress [1,2].

Pavement deterioration is not a single process. It results from the interaction of traffic-induced fatigue, permanent deformation under repeated loading, environmental exposure, and material aging. Each mechanism affects a different condition indicator and progresses at a different rate [3,4]. Cracking often develops gradually before producing measurable changes in roughness. Rutting accumulates early in the pavement life and may stabilize before cracking appears. A section with significant structural cracking may still show acceptable IRI values if crack widths remain narrow and load transfer is maintained [5,6]. Because different indicators reflect different mechanisms, their values do not move together in a predictable way. Studies using the Long-Term Pavement Performance database have found that correlations between IRI, rutting, and cracking indices are often weak and inconsistent across pavement types and climatic zones [7,8]. Evaluating one indicator cannot reliably predict the state of another, and condition assessment based on a single measure will systematically miss certain types of deterioration.

Most existing pavement deterioration studies use cross-sectional data or focus on individual sections [9,10]. These approaches do not capture how indicators co-evolve across a full network over time. The temporal relationships between pavement condition indicators, including whether one distress type systematically leads or follows another over time, have received limited empirical attention [11]. This gap matters for maintenance planning. If cracking reliably precedes increases in roughness, cracking observations can serve as an early warning signal for future IRI deterioration. If the two move together without a stable sequence, monitoring programs must measure both concurrently. Data-driven approaches have expanded in pavement applications but most remain focused on predicting a single indicator. Multi-indicator frameworks that examine how indicators interact and co-evolve over time are uncommon in the literature. Fewer still account for how these relationships vary by roadway context [12]. Few network-level studies have jointly examined whether pavement distress indicators exhibit stable sequential or contemporaneous relationships, how the dominant multivariate condition structure varies across roadway functional classes [7], and how these findings change after accounting for rehabilitation activity. Addressing this gap is essential because maintenance resources continue to be allocated using either single-indicator thresholds, most commonly IRI, or, where multi-indicator data are already collected, minimum-based aggregation rules that flag a section only once its single most limiting indicator crosses a threshold, without evidence that either approach adequately represents multi-mechanism deterioration across a mixed road network. The novelty of this study lies in developing and applying a multi-indicator condition framework that quantifies, using 17 years of longitudinal network-level data comprising 165,372 section-year observations, the shared and class-dependent structure of pavement condition, and in demonstrating that this structure is not an artifact of rehabilitation timing.

This study develops a multi-indicator framework for network-level pavement condition assessment using longitudinal pavement data from the Maryland Department of Transportation. The dataset covers Baltimore County from 2008 to 2024 and comprises 165,372 section-year observations across four indicators: IRI, rutting depth, structural cracking density, and functional cracking density. The framework integrates treatment-adjusted temporal analysis, contemporaneous and lagged relationship analysis, and principal component analysis (PCA) to examine how pavement condition evolves, how indicators co-vary over time, and how the dominant multivariate condition structure differs across roadway

classes. Treatment history is incorporated to reduce the influence of rehabilitation-related condition changes when interpreting temporal patterns.

The study makes four contributions, each addressing a specific gap in the network-level pavement literature. First, it tests whether the four pavement indicators exhibit a stable temporal ordering or whether their strongest relationships are contemporaneous at annual survey resolution and finds that the strongest associations occur contemporaneously at that resolution, with only a small tendency for rutting to precede the other indicators Second, it tests whether the relative contribution of each indicator to the dominant shared condition axis varies across functional classes and shows that it does: IRI and rutting contribute more strongly on High-Order roads, while cracking indicators contribute more strongly on Intermediate roads. Third, it tests whether these temporal and PCA findings persist after treatment-adjacent observations are excluded and confirms that they do, indicating the findings are not primarily generated by rehabilitation-related condition resets. Fourth, it compares a covariance-based multi-indicator composite against an existing minimum-based remaining-service-life aggregation and shows that the aggregation rule itself materially changes the condition profiles entering the high-priority set, with the composite emphasizing concurrent multi-indicator distress and the minimum-based rule emphasizing single limiting indicators. Collectively, the framework shifts pavement condition assessment from a single-indicator perspective toward a multi-indicator representation of network condition, providing a more comprehensive basis for infrastructure monitoring and pavement management decision-making.

The remainder of this paper is organized as follows. Section 2 reviews relevant literature. Section 3 describes the dataset and methodology. Section 4 presents the results and discussion. Section 5 concludes the study and outlines future research direction.

2. Literature Review

The following sections review four bodies of literature that directly motivate the analytical framework developed in this study. Section 2.1 examines the evidence that pavement deterioration is a multi-dimensional process and discusses the limitations of single-indicator and composite-index representations. Section 2.2 addresses the time-dependent nature of deterioration, focusing on temporal offsets between indicators and the limitations of current approaches to characterizing lagged relationships. Section 2.3 reviews the literature on statistical relationships between indicators, identifying the gap in dynamic and time-lagged interaction modeling at the network level. Section 2.4 introduces principal component analysis as a tool for multi-indicator dimensionality reduction and discusses its relevance for multi-indicator pavement data, including its application across roadway subgroups to characterize indicator weight structures by functional context.

2.1. Multi-Dimensional Pavement Deterioration Behavior

Pavement performance is commonly evaluated using composite indices such as the Pavement Condition Index (PCI), which summarize multiple distress types into a single value. Originally developed by the U.S. Army Corps of Engineers and subsequently adopted by highway agencies worldwide, PCI aggregates observed extents and severities of distress on a scale of 0 to 100 intended to represent the general condition of the pavement [13,14]. Although such indices support efficient network-level decision-making by reducing inspection data to a manageable scalar, a growing body of literature has argued that this aggregation comes at a cost: it may obscure the distinct physical mechanisms associated with different types of pavement deterioration and suppress information that is critical for maintenance planning and performance diagnosis [15]. As Gharaibeh et al. [16] noted, composite indices can produce similar scores for sections in very different physical

states, making them insufficient as a basis for intervention selection. Pavement condition is inherently multi-dimensional, as the most widely monitored performance indicators—International Roughness Index (IRI), rutting depth, surface cracking, and skid resistance—each reflect a different aspect of functional and structural behavior. IRI measures ride quality as a function of the longitudinal road profile and is sensitive primarily to long-wavelength surface irregularities [17]. Rutting represents the accumulation of permanent vertical deformation in the bound asphalt layers and, in some cases, the unbound subgrade, and is governed by the creep resistance of the asphalt mix under repeated loading at elevated temperatures [18]. Cracking—whether manifesting as fatigue-induced alligator cracking, low-temperature transverse cracking, or top-down surface cracking—reflects damage to the asphalt binder and aggregate matrix and is influenced by load magnitude, structural capacity, material properties, and climate [5]. Skid resistance, by contrast, is a surface texture property related to road safety under wet conditions and deteriorates primarily through traffic polishing of aggregate rather than through structural loading [19]. These mechanisms are physically distinct, and there is no a priori reason to expect that they should progress at the same rate or in the same sequence under field conditions. Empirical evidence confirms that these indicators do not evolve uniformly or proportionally over time. Analyses based on the Long-Term Pavement Performance (LTPP) program database—the most extensive repository of longitudinal in-service pavement data available—have consistently demonstrated that correlations between IRI, rutting, and cracking indices are often weak or inconsistent across pavement types, climatic regions, and traffic levels [10,20]. For example, pavements with significant cracking may still exhibit acceptable ride quality if crack widths remain narrow and load transfer is maintained, while sections with high roughness may show limited visible structural distress if the roughness originates from surface texture loss rather than structural damage. The practical consequences of this divergence are visible in prior LTPP research: Dong and Huang [21] evaluated rehabilitation effectiveness using IRI as the sole performance indicator—not because roughness fully represents pavement condition, but because the available distress measures were highly variable across test roads and no standard method existed to aggregate multiple distress types into a composite index, and the authors explicitly identified the incorporation of distress condition into performance evaluation as a necessary future development. Comparable findings have been reported in European monitoring programs, where roughness and structural indicators have been found to diverge considerably over the service life, particularly in pavements subjected to thermal cracking [22]. These findings have important implications for pavement management practice. Several studies have demonstrated that the choice of performance indicator significantly affects the identification of maintenance priorities and the predicted timing of intervention [23–25]. A section ranked as high-priority under a roughness-based criterion may not appear critical under a structural cracking criterion, and vice versa. This difference is not merely a data artifact; it reflects genuine differences in the deterioration pathways active in different sections. Ferreira et al. [26] further demonstrated that maintenance and rehabilitation recommendations can vary substantially depending on the pavement performance representation adopted within a pavement management system. Using different deterioration models within the Portuguese PMS, the authors showed that the same pavement sections could receive different intervention timings and rehabilitation strategies under different condition representations. These findings suggest that the manner in which pavement condition is aggregated and modeled can influence network-level maintenance prioritization and intervention planning Taken together, the evidence in the literature points toward the need for multi-indicator representations that preserve the individual trajectories of different distress types rather than compressing them into a single scalar. Such an approach is necessary not only for more complete condition assessment, but also for un-

derstanding the relationships and temporal interactions between pavement indicators—an area that remains comparatively underexplored and forms a central motivation for the present study.

2.2. Temporal Behavior of Pavement Deterioration

The time-dependent nature of pavement deterioration has long been recognized in mechanistic–empirical pavement design and asset management frameworks. Pavement management systems (PMS), such as the Highway Development and Management Model (HDM-4) and the AASHTO Pavement ME Design procedure, rely on deterministic performance curves that represent condition indicators as smooth, monotonically decreasing functions of cumulative equivalent single-axle loads (ESALs) or time [27,28]. These models are typically calibrated using population-level regression relationships and are effective for network-level planning, where averaging reduces the influence of section-specific variability. At the level of individual pavement sections, however, observed deterioration trajectories frequently deviate from these idealized assumptions. Longitudinal monitoring data indicate that deterioration is characterized by distinct phases: an initial period of relatively stable performance, followed by the onset of distress, and subsequently a phase of accelerated decline that may include abrupt changes associated with material failure events such as pothole formation, delamination, or structural cracking [29]. These nonlinear behaviors are not fully captured by smooth polynomial or exponential decay functions and may lead to inaccuracies in remaining service life estimation. The challenge of accurate condition representation is further compounded by measurement uncertainty inherent in pavement inspection technologies. Madanat and Ben-Akiva [30] showed that pavement inspection measurements are subject to systematic and random errors and that maintenance and rehabilitation models that assume error-free condition observations can yield suboptimal decisions and higher lifecycle costs. Their Latent Markov Decision Process formulation treats the true condition state of a facility as observable only imperfectly through inspection and shows that inspection frequency should be optimized jointly with maintenance decisions to minimize total lifecycle costs. This finding has direct relevance for multi-indicator pavement condition frameworks: if individual indicators are each measured with error, condition assessment based on several indicators jointly is less dependent on the measurement characteristics of any one of them than assessment based on a single measured indicator. Lifecycle frameworks for structural systems have similarly emphasized that condition assessment must account for time-dependent performance trajectories rather than static snapshots, recognizing that deterioration evolves through distinct phases of gradual degradation and, in some cases, abrupt performance drops associated with material failure events [31]. Within such frameworks, multiple performance indicators—including reliability indices, risk measures, and condition state probabilities—have been shown to carry complementary information about system health that cannot be fully captured by any single scalar representation, reinforcing the case for multi-indicator approaches to condition assessment. A further complexity is that different pavement indicators do not share the same deterioration timeline. Rutting, for instance, often develops rapidly during early service life due to primary consolidation and densification under initial traffic loading, after which its progression may stabilize in the absence of significant structural deterioration. In contrast, fatigue cracking is governed by cumulative damage processes based on Miner's hypothesis and may remain latent for a substantial portion of the pavement life before becoming observable at the surface [32,33]. Because these distress mechanisms evolve under different physical processes and timescales, indicators may exhibit weak or inconsistent temporal alignment even when they originate from the same pavement section. Once initiated, cracking progression can contribute to deterioration in other indicators,

particularly roughness, as surface integrity degrades and load-transfer efficiency declines. These observations suggest that temporal offsets may exist between indicators, with changes in one distress type potentially preceding or following changes in another depending on the governing deterioration mechanisms. The existence of such offsets has direct implications for monitoring program design: if different distress types require different detection strategies because they evolve at different rates and through different physical processes, then no single indicator can reliably serve as a proxy for the overall condition state of a pavement section. Longitudinal pavement datasets provide an important framework for examining how condition indicators evolve and interact over time. Previous studies have used time-series pavement observations to analyze deterioration trends, distress progression, and temporal variation in pavement condition [29,34]. However, most existing approaches focus primarily on the temporal evolution of individual indicators rather than on the interactions or temporal dependencies between multiple distress measures. In pavement applications, however, these approaches have been used predominantly for univariate prediction, most commonly focusing on indicators such as IRI or PCI [34,35]. As a result, they capture temporal dependence within individual indicators but do not explicitly account for the co-evolution or interaction of multiple distress measures. Critically, the temporal relationships between pavement performance indicators have not been systematically quantified. While mechanistic theory supports the existence of lagged relationships between distress types, their empirical characterization using field data remains limited. Understanding these temporal offsets is important for identifying early indicators of deterioration, improving deterioration assessment, and informing maintenance timing. This gap motivates the analytical approach adopted in the present study, which examines the temporal dependencies and interactions between multiple pavement condition indicators.

2.3. Relationships Between Pavement Condition Indicators

Studies have shown that pavement performance indicators are not independent, as different distress types are influenced by shared loading conditions, material properties, and environmental factors [33,36]. As a result, statistical associations between indicators such as roughness, rutting, and cracking are commonly observed in pavement monitoring data. Several studies have investigated these relationships using cross-sectional or same-year analyses of pavement condition datasets [16]. Correlation analyses using datasets such as the Long-Term Pavement Performance (LTPP) program and state-level pavement management systems have examined pairwise relationships between indicators including IRI, rutting depth, and cracking measures. Reported findings are mixed. Several recent studies have identified moderate positive relationships between roughness and cracking, consistent with the physical mechanism by which progressive surface distress contributes to increased ride irregularity and pavement roughness [7,15]. In contrast, rutting and IRI often exhibit weak or inconsistent relationships, reflecting the fact that rutting primarily affects localized wheelpath deformation rather than the broader wavelength characteristics governing roughness response. Similarly, cracking and rutting frequently show limited correlation because they are governed by different deterioration mechanisms—fatigue damage and shear deformation, respectively—and may develop independently depending on pavement structure, material properties, and loading history [29,33]. A consistent limitation of these studies is their reliance on contemporaneous data, where correlations are computed using indicator values measured within the same inspection cycle. While this approach is suitable for identifying concurrent associations, it does not capture temporal dependencies between indicators. As a result, potential directional relationships—where one distress type precedes or influences another—remain unobserved. This limitation is important for both condition assessment and maintenance planning. For example, if cracking systematically

precedes increases in roughness, then cracking observations may provide early information about future IRI deterioration that is not captured through same-year correlation analysis. Cross-correlation analysis provides a direct approach for examining such temporal relationships by evaluating associations between time series at different lag intervals. Widely used in fields such as econometrics and signal processing, cross-correlation functions can identify leading and lagging relationships between variables [37]. In the context of pavement performance, a statistically significant correlation at a non-zero lag would indicate that one distress type tends to precede another, providing empirical evidence of sequential deterioration behavior. The concept of dynamic interaction between indicators extends beyond cross-correlation to multivariate time-series frameworks such as vector autoregressive (VAR) models, where the current state of one variable influences the future state of another. While such approaches have been applied in other infrastructure systems [38], their application to pavement performance indicators remains limited. More broadly, the dynamic coupling between indicators over time—including the direction, magnitude, and timing of interactions—has not been systematically characterized in the pavement literature. Taken together, the literature highlights three key limitations: the reduction of multi-dimensional pavement condition into single composite indices, the simplified representation of temporal deterioration behavior, and the limited examination of temporal interactions between indicators. These limitations define the research gap addressed in the present study. The analytical framework developed herein integrates multi-indicator representation, temporal dependency analysis, and dimensionality reduction to examine how the indicators co-vary, whether their relationships follow a temporal sequence, and how the dominant multivariate condition axis differs across roadway classes.

2.4. Principal Component Analysis as a Multi-Indicator Representation Tool

The limitations of single-indicator and composite-index approaches have motivated the use of data-driven techniques for representing complex systems described by multiple correlated variables. One of the most common approaches is principal component analysis (PCA), which transforms correlated variables into a smaller set of orthogonal components that retain most of the variation in the original dataset [39]. PCA is widely used when the goal is to reduce redundancy, identify dominant patterns, and construct a compact representation of a system from several observed indicators [40,41]. Although PCA has been applied in selected pavement engineering studies, the method has been applied extensively in other fields to achieve a similar purpose. In environmental analysis, PCA has been used to examine multivariate pollution data and identify dominant patterns in complex environmental datasets [42]. In finance, Mavungu et al. [43] applied PCA to portfolio return data and demonstrated that a reduced set of orthogonal principal components could capture most of the variance in correlated financial variables, illustrating how complex multivariate systems can be represented through a smaller number of dominant components. In structural health monitoring, PCA has been used to extract damage-sensitive features from multiple measurements and to distinguish normal structural behavior from deterioration or damage states [44]. More recent work has extended these data-driven approaches to deep learning and other artificial intelligence methods that extract features automatically from vision, vibration, and nondestructive testing data for damage identification across civil infrastructure [45,46]. PCA has also been applied in transportation decision-making problems to reduce multiple evaluation indicators into a smaller set of informative variables for project prioritization [47]. These applications show that PCA is an established tool for summarizing the dominant pattern of variation among several interrelated variables. These studies show that PCA is well suited to problems in which a system is described by several interrelated indicators but the researcher seeks a smaller number of latent dimensions

that capture the dominant structure in the data. This logic aligns closely with pavement condition assessment, where indicators such as IRI, rutting, cracking, and skid resistance represent different but related aspects of pavement performance. Rather than replacing individual indicators, PCA summarizes the dominant direction of covariation among them while retaining the contribution of each variable to that summary. In the present study, PCA is adopted in that broader sense, to summarize the dominant shared variation among multiple pavement condition indicators within a network-level time-series dataset. This extends the use of PCA beyond static cross-sectional reduction toward the representation of evolving pavement condition over time. The first principal component is interpreted as a dominant shared multivariate condition axis rather than as an estimate of an underlying latent construct.

3. Data and Methodology

3.1. Dataset and Variables

This study uses a network-level pavement condition dataset obtained from the Maryland Department of Transportation (MDOT), covering roadway sections in Baltimore County over the period 2008–2024. The dataset is structured at the section level, with each record representing a distinct roadway segment defined by route identifiers and milepoints, accompanied by condition measurements recorded for a given observation year. The longitudinal structure of the dataset enables consistent tracking of pavement performance across both space and time. Table 1 summarizes the variables used in this study, including their units, descriptions, and analytical roles.

Table 1. Dataset Variables and Descriptions.

Variable	Full Name	Description	Role in Study
IRI (in/mi)	International Roughness Index	Measures longitudinal surface irregularities along the wheel path; averaged across left and right wheel paths	Roughness indicator; temporal analysis, cross-correlation, and PCA
RUT (in)	Rutting Depth	Permanent vertical deformation in the asphalt layers under repeated traffic loading; averaged across left and right wheel paths	Deformation indicator; temporal analysis, cross-correlation, and PCA
SC_Density	Structural Cracking Density	Density of cracks within the wheel path; associated with load-induced fatigue damage and structural deficiency	Structural distress indicator; temporal analysis, cross-correlation, and PCA
FC_Density	Functional Cracking Density	Density of cracks outside the wheel path; associated with material aging, thermal cycling, and environmental exposure	Surface distress indicator; temporal analysis, cross-correlation, and PCA
SKID_NO_RECENT	Skid Number	Most recent skid resistance measurement under wet conditions; governed by aggregate polishing rather than structural loading	Excluded from PCA; different survey cycle and distinct deterioration mechanism
FC	Functional Class	FHWA roadway classification, grouped as High-Order (Interstate, Principal Arterial), Intermediate (Minor Arterial, Collector), and Low-Order (Minor Collector, Local)	Stratification variable; temporal and PCA analyses
GLOBAL_ROUTE_ID	Global Route Identifier	Unique numerical ID assigned to each route	Section tracking across observation years
INV_BMP/INV_EMP	Section Milepoints	Beginning and ending milepoints defining each pavement section; consistent across directional bounds	Spatial delineation of sections

Table 1. *Cont.*

Variable	Full Name	Description	Role in Study
YEAR	Reporting Year	Calendar year of the condition record; 2008 to 2024, reported annually	Temporal index for all analyses
Treatment History	Documented Treatment Record	Type and year of the most recent documented rehabilitation treatment for each segment, updated annually; segments may accumulate multiple treatments over 2008–2024	Defines the treatment-adjacent exclusion window used
IRI_RSL/RUT_RSL/ SCD_RSL/FCD_RSL	Component Remaining Service Life	Estimated remaining service life for each indicator, derived via MDOT condition-to-service-life breakpoint table	Component inputs to RSL_4; basis for comparison against PC1
RSL_4	Four-Indicator Remaining Service Life	Minimum of IRI_RSL, RUT_RSL, SCD_RSL, and FCD_RSL	Minimum-based aggregation rule; compared directly against PC1

Five pavement condition indicators are examined, each reflecting a different aspect of pavement performance. IRI measures longitudinal surface irregularities and serves as the most widely used condition metric in practice. RUT captures permanent vertical deformation along the wheel path under repeated traffic loading. FC_DENSITY and SC_DENSITY measure crack density outside and within the wheel path, respectively—the former associated with material aging and environmental exposure, the latter with load-induced fatigue damage and structural deficiency. Skid resistance, reported as a skid number, reflects surface friction under wet conditions; it is recorded in the dataset but excluded from the PCA and cross-correlation analyses (Section 3.4.3) and is used only as an external check on the specificity of the composite in Section 4.4. Used together, the four analyzed indicators describe surface condition and structural integrity, a more complete characterization of network condition than any single indicator allows.

The dataset additionally includes a remaining service life measure, RSL_4, used in Section 4.3.5 as a comparison point for the PCA composite. RSL_4 is constructed as the minimum of four component remaining-service-life estimates—IRI_RSL, RUT_RSL, SCD_RSL, and FCD_RSL—each derived from its corresponding raw indicator via the MDOT condition-to-service-life via the MDOT condition-to-service-life breakpoint table [48]. RSL_4 was selected over the five-indicator variant (RSL_5, which also incorporates skid resistance) to match the four-indicator scope used in PCA for the same reason skid resistance is excluded from PCA in Section 3.4.3: its distinct survey cycle. Because RSL_4 is built from the same four indicators used in PCA, comparing it against PC1 in Section 4.3.5 holds the underlying data fixed. The two measures still differ in both the scaling of the indicators (the MDOT breakpoint transform versus standardization) and the aggregation rule (minimum versus weighted sum); these two effects are separated in Section 4.3.5 by comparing rules that differ in only one of them.

3.2. Data Preparation

Several preprocessing steps were performed to produce a consistent and analysis-ready dataset. Engineering judgment thresholds were applied to remove physically implausible observations, including sections with IRI values exceeding 400 in/mile, and cracking density values exceeding 100. These thresholds were selected based on MDOT pavement condition guidance, Federal Highway Administration (FHWA) roughness interpretation ranges, and physical constraints associated with pavement distress measurements [49]. A total of 327 observations (0.20% of the dataset) were removed during this process and did not materially alter the overall network-level distributions.

A missingness diagnostic conducted prior to imputation showed that missing observations were not uniformly distributed across years. Missingness was concentrated in particular survey years, exceeding 30% of records in 2011–2012, 2020–2021, and 2024, reaching about 20% in 2023, and remaining below 11% in all other years; the 2020–2021 peak reflects disruptions in field data collection during the COVID-19 pandemic. Skid resistance exhibited consistently higher missingness because it is collected on a different survey cycle from the primary distress indicators. Overall, missingness was associated primarily with survey timing and data collection cycles, supporting the use of temporal interpolation methods. To test whether missingness was also related to pavement condition, the 7370 sections covered by the pre-imputation survey file were examined. For each indicator, the value recorded in the preceding year was compared between section-years in which the indicator was missing and those in which it was observed, and a logistic model of missingness on the standardized preceding value was fitted with survey-year effects and section-clustered standard errors. Results are summarized in Table 2. Interpolated values differed little from observed values (standardized mean differences of 0.17 or less in absolute value), with the largest difference for IRI. Within survey years, missingness of rutting, structural cracking, and functional cracking was unrelated to prior condition (odds ratios of 1.00 to 1.04 per standard deviation). Missingness of IRI was moderately associated with prior roughness (odds ratio 1.33 per standard deviation, 95% CI 1.29 to 1.36), so the mechanism cannot be treated as random for IRI. Missingness was not higher in years with a documented treatment, indicating that the association does not arise from sections going unsurveyed during rehabilitation. Because each missing value is interpolated from the same section's observed values on either side, an association with observed condition is largely accommodated by the imputation, and the principal analyses were also repeated without imputed values (Sections 4.1.2, 4.2.2 and 4.4).

Table 2. Diagnostic of the relationship between missingness and pavement condition (7370 sections covered by the pre-imputation survey file).

Indicator	Interp. (%)	Observed vs. Interpolated			Value in Preceding Year			OR per SD [95% CI]
		Obs. Median	Interp. Median	SMD	Next Missing	Next Observed	SMD	
IRI (in/mi)	16.5	96.28	105.07	+0.17	103.24	94.95	+0.16	1.33 [1.29, 1.36]
Rutting (in)	18.6	0.160	0.163	−0.05	0.170	0.160	+0.04	1.00 [0.98, 1.02]
SC_DENSITY	19.1	0.73	1.08	−0.02	0.75	0.74	+0.02	1.04 [1.02, 1.07]
FC_DENSITY	19.1	2.35	2.75	+0.01	2.30	2.34	0.00	1.01 [0.99, 1.03]

Note: SMD = standardized mean difference (difference in means divided by the standard deviation of observed values); OR = odds ratio of the indicator being missing in a given year per standard deviation of its directly observed value in the preceding year, from a logistic model with survey-year effects and section-clustered standard errors. Medians and means can differ in direction for the cracking indicators because of their strong right skew.

Missing values were filled independently within each pavement section using a three-stage temporal imputation procedure. Linear interpolation was first applied to internal gaps between observed years, followed by backward filling for leading gaps and forward filling for trailing gaps. The pre-imputation survey file covers 7370 of the 9580 sections; for these, 16.5% to 19.1% of records per indicator were filled by this procedure. The remaining 2210 sections, predominantly High-Order routes, are not included in that file, and interpolation rates are therefore reported for the covered sections only. While interpolation was necessary to preserve temporal continuity, the use of linear interpolation may locally reduce short-term variability and should therefore be considered when interpreting network-level correlation magnitudes and cross-correlation results. The sensitivity of the cross-correlation results to this interpolation is evaluated directly in Section 4.2.2, and the sensitivity of the PCA results is evaluated in the Limitations section (Section 4.4).

After imputation, the remaining gaps were section-years in which no value was available for one or more indicators. Prior to PCA, the four distress indicators (IRI, rutting, structural cracking density, and functional cracking density) were standardized to zero mean and unit variance to eliminate differences in measurement scale. Skid resistance was not included because it is a surface friction and safety measure rather than a structural or functional distress indicator; it is surveyed on a different schedule from the other indicators, and it is governed primarily by aggregate polishing under traffic rather than by the load-induced and environmental processes reflected in the four distress indicators. Of the 165,372 section-year observations, 162,684 had values for all four indicators and were used in the PCA. The remaining 2688 observations (1.6%) were excluded from the PCA but retained in the indicator-level trend analysis in Section 4.1.

The dataset comprises 9739 pavement sections, each defined by route identifier, beginning and ending milepoint, and direction. Of these, 9580 independent sections have complete records for all four indicators and form the basis of the PCA (5866 High-Order, 3638 Intermediate, and 76 Low-Order), observed over 17 annual reporting cycles. The record for each section is long and nearly complete: 97.8% of sections have a condition record in every year from 2008 to 2024, and all have at least six. Counting only directly observed surveys, for the sections that could be verified against the pre-imputation survey file, the median section was surveyed in 14 of the 17 years (interquartile range 12 to 15), and 93.8% were surveyed in at least 10 years. Five sections defined by zero-length milepoint references contain more than one record per year (221 records in total) and were treated as single sections in the cluster bootstrap.

3.3. Treatment History and Treatment-Adjusted Analysis

Documented rehabilitation activity was incorporated to distinguish deterioration from condition changes attributable to maintenance intervention. Treatment records were obtained from MDOT's construction and treatment history for Baltimore County and matched to condition sections by route identifier and milepoint range, yielding a treatment year for each documented intervention. Excluding all observations occurring after a section's first documented treatment would produce a survivorship-biased sample: sections never treated over the study period would be retained in full, while treated sections would be truncated at their first intervention, so the remaining sample would not be compositionally representative of the network as treatment activity accumulates over 17 years. To avoid this bias, the present analysis instead excludes only observations falling within a two-year window of any documented treatment event, regardless of treatment type, retaining all other observations regardless of a section's treatment history elsewhere in the study period. This treatment-free interval approach is applied in two ways: (i) network-level annual statistics are recomputed after excluding treatment-adjacent observations (Section 4.1.1), and (ii) section-level deterioration rates are computed only between consecutive observed surveys with no documented treatment falling between them (Section 4.1.2, Equation (1)).The two-year window was chosen because treatment dates are recorded by year and the timing of the annual condition survey relative to construction within that year is unknown, so the first survey reflecting a treatment may fall in the treatment year or the following year; excluding observations up to two years after a treatment removes both possibilities with one further year as a buffer against early post-treatment adjustment, while retaining the majority of observations. Because the choice is nonetheless a judgment, the principal results were repeated with windows of zero (treatment year only), one, and three years (Section 4.3.7), and the network IRI trend was also estimated by time since the last documented treatment (Section 4.1.1).

3.4. Analytical Framework

The analytical framework evaluates pavement deterioration as a multi-dimensional and time-dependent process through three analytical components, each addressing a specific aspect of the research objectives. Figure 1 presents an overview of the workflow.

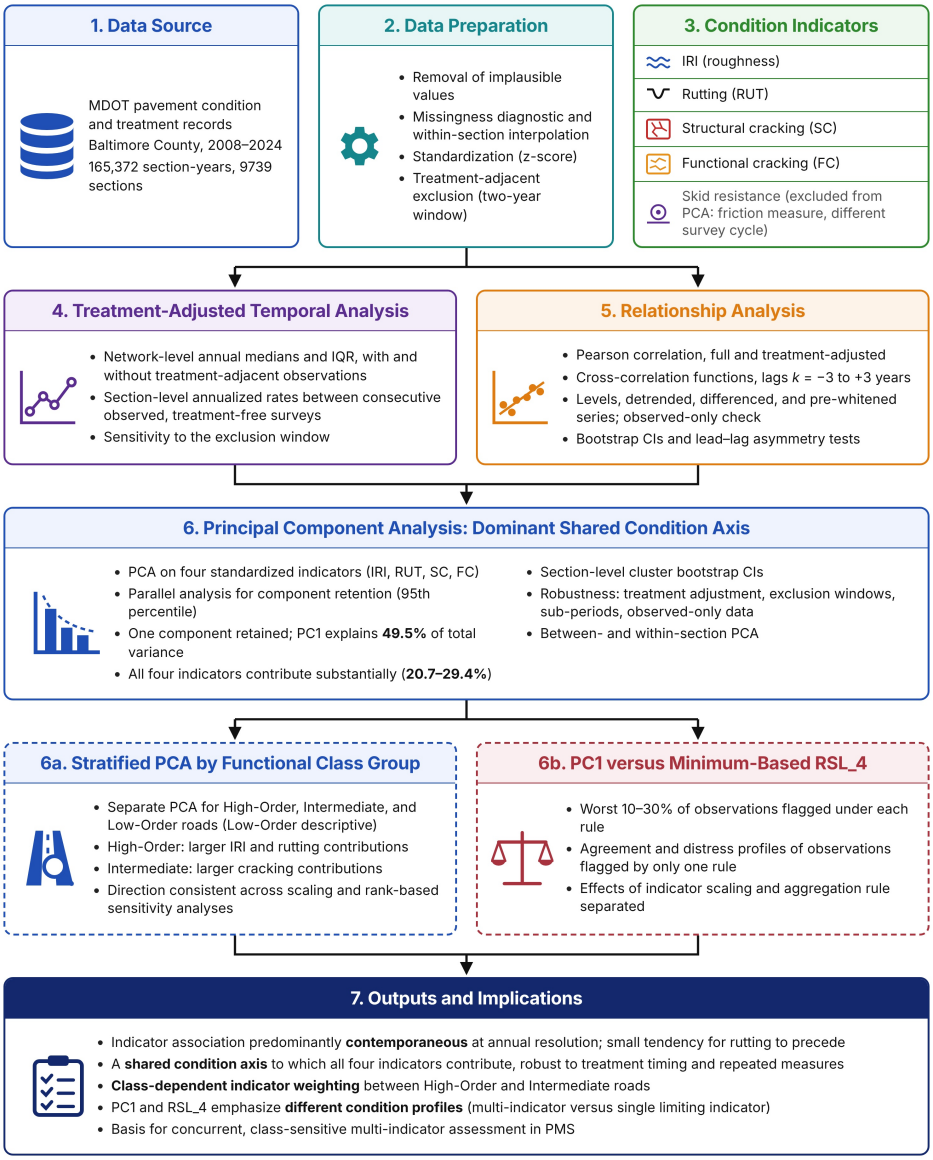


Figure 1. Framework Flowchart.

3.4.1. Temporal Analysis of Pavement Condition Indicators

Trends were analyzed at two levels of aggregation. At the network level, annual median, mean, and interquartile range (IQR, the range between the 25th and 75th percentiles) statistics were computed for each indicator to identify system-wide trends and changes in distributional variability over the study period. At the section level, a single ordinary least squares slope fit across a section's full 17-year history would conflate deterioration with intervening treatment events; deterioration rates are instead computed between consecutive observed surveys with no documented treatment falling within the interval itself (a stricter criterion than the two-year buffer window used for the network-level statistics above), excluding interpolated years, for each section. For the j-th qualifying observed interval of section i, spanning years t_{j-1} to t_j, the annualized rate is

$$r_{i,j} = \frac{Y_{i,t_j} - Y_{i,t_{j-1}}}{t_j - t_{j-1}} \quad (1)$$

where Y_i is the indicator value for section i. Because higher values of all four indicators represent poorer pavement condition, $r_{i,j} > 0$ denotes deterioration and $r_{i,j} < 0$ denotes improvement. Differences between functional classes are examined through the stratified PCA described in Section 3.4.3.

3.4.2. Relationship Analysis Between Indicators

Interdependencies between pavement condition indicators are evaluated using both contemporaneous and time-lagged analyses. Pearson correlation coefficients are computed to assess same-year linear relationships between all indicator pairs, quantifying the extent to which indicators vary together within the same observation cycle.

Cross-correlation analysis is further applied to examine temporal relationships in which changes in one indicator may precede or follow changes in another over time. For each indicator pair, cross-correlation functions are computed at lags from $k = -3$ to $k = +3$ years at the section level and aggregated using the median across sections. Lag values are interpreted in terms of potential leading and lagging behavior between indicators, providing empirical evidence on whether deterioration follows a consistent temporal sequence or whether indicator co-movement is primarily contemporaneous across the network. The primary analysis uses condition levels. Because cross-correlations between trending series can peak at $k = 0$ without dynamic coupling, the analysis was repeated on linearly detrended levels, on first differences between consecutive years excluding treatment-adjacent years (for all years and for directly observed years only), and on series pre-whitened with a first-order autoregressive filter estimated from pooled within-section data [37]. For each lag, the number of contributing sections and lagged pairs and the distribution of section-level CCFs are reported, both for all records and for directly observed values only; a section contributes at a given lag when it has at least five lagged pairs. Confidence intervals (95%) for the network median CCF at each lag were obtained by bootstrap resampling of Sections (1000 replicates), and differences between lags, including the asymmetry between lags of -1 and $+1$, were tested on paired section-level values using the bootstrap and the Wilcoxon signed-rank test. For pre-whitened series, each section-level CCF was also compared with its own approximate 95% bound, $\pm 1.96/\sqrt{n}$, where n is the number of lagged pairs.

3.4.3. Dimensionality Reduction Using PCA

Principal component analysis is applied to the standardized values of four condition indicators, namely IRI, rutting, functional cracking density, and structural cracking density, to summarize the shared variation among them. These four indicators are selected for PCA because they reflect the primary structural and surface distress mechanisms active in flexible pavement systems and are consistently measured across the full observation period. Skid resistance is excluded from this component of the analysis because it is surveyed on a different cycle from the distress indicators, resulting in substantially more missing observations, and because its governing mechanism, surface texture polishing under traffic, is physically distinct from the load-induced and environmental processes represented by the other four measures.

PCA decomposes the correlation matrix $\mathbf{R}$ of the indicators, which is the covariance matrix of the standardized indicators, into orthogonal components [39,41]. The j-th principal component is defined as

$$PC_j = \mathbf{a}_j^\top \mathbf{z} \quad (2)$$

where PC_j is the score of the j-th principal component (dimensionless); $\mathbf{z} = [z_{\mathrm{IRI}}, z_{\mathrm{RUT}}, z_{\mathrm{SC}}, z_{\mathrm{FC}}]^\top$ is the vector of standardized condition indicators, in which each element is obtained as $z_i = (x_i - \mu_i)/\sigma_i$, with x_i denoting the observed indicator value, μ_i the network-level mean, and σ_i the network-level standard deviation of indicator i; and $\mathbf{a}_j = [a_{j1}, a_{j2}, a_{j3}, a_{j4}]^\top$ is the vector of component coefficients satisfying $\mathbf{a}_j^\top \mathbf{a}_j = 1$. Because each indicator is standardized to zero mean and unit variance prior to analysis, all terms are dimensionless. The original indicators carry the following units before standardization: IRI in inches per mile (in/mi), rutting depth in inches (in), and structural and functional cracking densities as dimensionless density measures.

Three further quantities are derived from PC1 for each indicator j:

$$\ell_{1j} = a_{1j}\sqrt{\lambda_1}, \qquad c_{1j} = a_{1j}^2, \qquad \ell_{1j}^2 = \lambda_1 c_{1j} \tag{3}$$

where λ_1 is the first eigenvalue. The loading ℓ_{1j} is the correlation between indicator j and the PC1 score. The contribution c_{1j} is the proportion of the variance of PC1 attributable to indicator j; contributions sum to 100% across the four indicators and are used throughout this study to compare indicators across functional class groups and sub-periods [41]. The squared loading ℓ_{1j}^2 is the proportion of the variance of indicator j captured by PC1, and its sum over the indicators equals λ_1. These relations hold for indicators with unit variance; where this is not the case, as in the functional class analysis described below, loadings are computed directly as correlations.

PCA is used here as a descriptive dimension-reduction technique. It decomposes the total variance of the standardized indicators and does not separate variance shared among indicators from indicator-specific variance or measurement error, as a common-factor model would [39]. The first principal component is therefore interpreted as the dominant shared multivariate condition axis: the linear combination of the four indicators that accounts for the largest proportion of their total variance. Higher PC1 scores indicate concurrently poorer condition across the indicators, weighted by their coefficients.

The number of statistically meaningful components is determined using parallel analysis, which compares observed eigenvalues against those produced by randomly generated datasets of equivalent dimensions. A component is retained only when its observed eigenvalue exceeds the 95th percentile of the simulated eigenvalue distribution. Because the simulated thresholds approach 1 as the number of observations grows, the procedure is most informative for the smaller functional-class subsets and converges on the Kaiser criterion for the full network.

Functional class groups were formed from the functional classification code recorded for each section in the MDOT condition data, which follows the legacy rural and urban functional system coding defined in the MDOT data dictionary supplied with the dataset; codes absent from that dictionary were classified from the MDOT functional classification of the routes concerned [50]. The classes assign roads according to their role in the network, from mobility-oriented facilities carrying long-distance, high-volume traffic to access-oriented facilities serving adjacent land [51]. To obtain groups large enough for stable estimation, the classes were combined into three groups following this hierarchy: High-Order (Interstates, other freeways and expressways, and other principal arterials), Intermediate (minor arterials and major collectors), and Low-Order (minor collectors and local roads). The three groups are defined for this study, and the functional class codes assigned to each are listed in Table 3. The grouping is based on functional classification only; traffic volume and pavement structure were not available for the sections and were not used.

Table 3. Functional class codes in the MDOT condition data, their definitions, and their assignment to class groups.

Code	Functional Class (MDOT Definition)	Group	Section-Years
1	Rural Principal Arterial—Interstate	High-Order	6356
2	Rural Principal Arterial—Other	High-Order	1933
3	Principal Arterial (Other) [a]	High-Order	509
11	Urban Principal Arterial—Interstate	High-Order	42,345
12	Urban Principal Arterial—Other Freeways	High-Order	15,843
14	Urban Principal Arterial—Other	High-Order	32,640
6	Rural Minor Arterial	Intermediate	7512
7	Rural Major Collector	Intermediate	17,440
16	Urban Minor Arterial	Intermediate	30,306
17	Urban Collector	Intermediate	6524
9	Rural Local	Low-Order	34
18	Minor Collector [a]	Low-Order	270
19	Urban Local	Low-Order	972
	Total		**162,684**

Note: Definitions from the MDOT data dictionary supplied with the condition data. [a] Not defined in that dictionary; classified from the MDOT SHA functional classification of the routes concerned [50].

For the functional class analysis, PCA was applied within each group to the indicators standardized with the network-level mean and standard deviation, so that contributions and PC1 scores remain on a common scale across road classes. Under this scaling, the contribution of an indicator within a group reflects both its correlations with the other indicators and its within-group variance relative to the network. Because parallel analysis assumes unit-variance indicators, the number of components in each group was determined from the within-group correlation matrix. Two sensitivity analyses were conducted: PCA on indicators standardized within each group, which isolates the correlation structure, and PCA on rank-transformed indicators, which limits the influence of extreme values in the right-skewed cracking distributions.

Because each section contributes up to 17 annual observations, the section-year records are not independent, and the pooled analysis combines variation between sections with variation within sections over time. This repeated-measures structure was addressed in two ways. Confidence intervals for all PCA quantities were obtained from a cluster bootstrap with 1000 replicates that resamples whole pavement sections, preserving the dependence among repeated observations of the same section. In addition, the PCA was repeated at two levels, on section means (one observation per section) and on deviations of each observation from its section mean, following a two-level decomposition of the covariance structure [52], and on 200 samples each containing one randomly selected observation per section.

Within-section centering was also repeated on the treatment-adjusted subset, and PCA was applied to annualized increments between consecutive directly observed surveys with no documented treatment within the interval.

3.4.4. Framework Summary

The methodology integrates data preparation, treatment adjustment, temporal trend analysis, statistical relationship assessment, and dimensionality reduction into a unified analytical pipeline. Each component addresses a distinct aspect of multi-indicator pavement condition. Temporal analysis characterizes how individual indicators evolve at the network and section levels after excluding treatment-adjacent observations. Relationship analysis

quantifies the concurrent and lagged dependencies between indicators. PCA summarizes the dominant shared variation among the four indicators as a single condition axis, examines how the weighting of that axis differs across High-Order, Intermediate, and Low-Order roads, and provides a composite score that is compared against the minimum-based RSL_4 measure to assess how the aggregation rule affects section prioritization.

By treating pavement deterioration as a multidimensional and time-dependent process rather than a condition represented by a single metric, the framework provides an empirically grounded basis for evaluating the completeness of network-level pavement assessment. The framework is transparent and reproducible, and it is applicable to longitudinal multi-indicator datasets from other transportation agencies.

4. Results

This section presents and interprets the results of the multi-indicator pavement deterioration analysis across three analytical components. Section 4.1 examines the temporal evolution of condition indicators at the network and section levels. Section 4.2 evaluates concurrent and lagged relationships between indicators. Section 4.3 presents the PCA results, characterizing the dimensional structure of deterioration across the indicator set, including a stratified analysis by functional class group. Section 4.4 synthesizes the findings and discusses their implications for network-level pavement management.

4.1. Treatment-Adjusted Temporal Evolution of Pavement Condition Indicators

4.1.1. Rehabilitation Effects on Network-Level Condition Trends

The treatment-adjusted analysis retained 103,171 of the original 165,372 section-year observations (62.4%) after observations occurring within two years of a documented treatment event were excluded. Annual retention ranged from 51.5% to 81.6%. This filtering was used to reduce the influence of rehabilitation-related condition resets on the network-level temporal trends. Table 4 and Figure 2 present the resulting median condition trajectories. In Table 4, the IQR columns report the range between the 25th and 75th percentiles of each indicator across all sections in 2008 and in 2024, so that a change in the IQR indicates whether the spread of condition across the network widened or narrowed over the study period. For IRI, for example, the middle 50% of sections lay between 83.09 and 153.49 in/mi in 2008 and between 72.20 and 151.43 in/mi in 2024.

Removal of treatment-adjacent observations affected the four condition indicators differently. Treatment-adjusted IRI medians exceeded the corresponding unadjusted medians in 15 of 17 years, with an average difference of 3.8%. The effect was substantially larger for FC_DENSITY, for which adjusted medians were higher in 15 of 17 years and averaged 10.8% above the unadjusted values, exceeding 20% in five years (2014, 2016, 2017, 2018, and 2019) and reaching 48.2% in 2016. These results indicate that recently treated observations suppressed the apparent network-level magnitude of IRI and, more strongly, functional cracking. Rutting was less affected, with adjusted medians averaging 3.3% above the full-network values.

Structural cracking exhibited a different response. Treatment-adjusted SC_DENSITY medians exceeded unadjusted values in only 6 of 17 years and were 8.1% lower on average. This contrasting response shows that rehabilitation-related observations do not influence all pavement indicators uniformly. The effect of treatment filtering is therefore indicator-dependent, supporting separate evaluation of roughness, deformation, and cracking when interpreting longitudinal network condition.

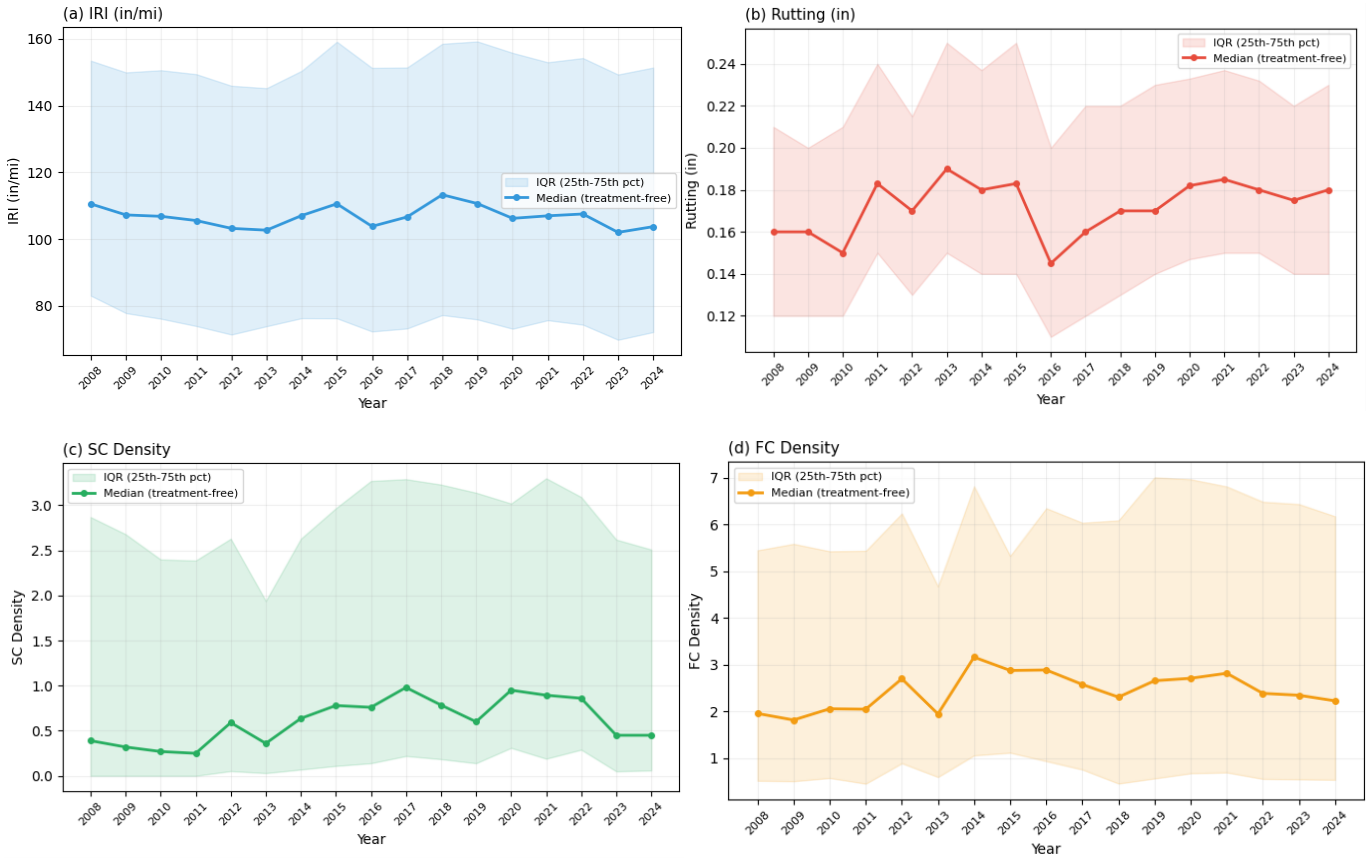


Figure 2. Network-level median trends with IQR after exclusion of treatment-adjacent observations, 2008–2024 (observations within 2 years of a treatment excluded): (**a**) IRI, (**b**) rutting, (**c**) structural cracking density, (**d**) functional cracking density.

Table 4. Network-level summary statistics for pavement condition indicators from 2008 to 2024 after exclusion of treatment-adjacent observations.

Indicator	2008 Median	2024 Median	Median Range	IQR 2008/2024	Main Observation
IRI (in/mi)	110.58	103.77	102.05–113.33	83.09–153.49/ 72.20–151.43	Median declines modestly after excluding treatment-adjacent observations; interquartile range remains wide, indicating persistent variability in ride quality across the network.
Rutting (in)	0.160	0.180	0.145–0.190	0.12–0.21/ 0.14–0.23	Values remain low across the study period, with a modest upward trend once treatment-adjacent observations are excluded.
SC Density	0.390	0.450	0.250–0.980	0.00–2.87/ 0.06–2.51	Median increases modestly while remaining low overall; damage remains concentrated in a subset of sections, as the lower IQR bound remains near zero.
FC Density	1.960	2.230	1.820–3.165	0.52–5.45/ 0.54–6.18	Median increases and interquartile range widens, indicating that functional cracking continues to spread across a larger share of the network.

More importantly, removal of treatment-adjacent observations did not eliminate the long-term divergence among the condition indicators. From 2008 to 2024, median IRI declined from 110.58 to 103.77 in/mi, while rutting increased from 0.160 to 0.180 in, SC_DENSITY increased from 0.390 to 0.450, and FC_DENSITY increased from 1.960 to 2.230. Func-

tional cracking therefore continued to worsen while network-level roughness did not increase, a difference reconciled with the section-level rates in Section 4.1.2. Moreover, because the treatment-adjusted FC_DENSITY median exceeded the unadjusted value by an average of 10.8%, inclusion of treatment-adjacent observations had previously understated the apparent magnitude of functional cracking for much of the study period. The divergence between ride quality and cracking condition therefore persists after controlling for the immediate influence of documented rehabilitation events and cannot be explained solely by treatment-related condition resets. The long-term direction of change was insensitive to the treatment-exclusion window. Treatment-year-only, one-year, and two-year exclusions all produced decreasing IRI and increasing rutting, SC_DENSITY, and FC_DENSITY between 2008 and 2024. The annual median trajectories were strongly correlated between the treatment-year-only and two-year analyses for rutting ($r = 0.950$), SC_DENSITY ($r = 0.927$), and FC_DENSITY ($r = 0.843$); IRI showed greater interannual sensitivity ($r = 0.573$), although its long-term direction remained unchanged.

Because resurfacing can keep roughness low for several years, a two-year window may still include recently treated sections, and their share grows as treatments accumulate. To test whether this explains the lower network IRI, IRI was compared across sections grouped by time since their last documented treatment (0–2, 3–5, 6–10, and 11 or more years, or no documented treatment), using a regression with standard errors clustered by section. Sections treated within the previous ten years had IRI 20.1 to 23.7 in/mi lower than untreated sections, while those treated 11 or more years earlier did not differ ($p = 0.27$), so the effect of treatment on roughness lasts longer than any practical exclusion window. Over the study period, the share of observations within five years of a treatment rose from 38.3% to 56.0%. When time since treatment was ignored, IRI decreased by 0.14 in/mi per year; when it was accounted for, IRI increased by 0.12 in/mi per year (95% CI 0.05 to 0.18). The lower network IRI therefore reflects the growing share of recently treated sections rather than improving ride quality.

4.1.2. Section-Level Deterioration Rates Between Treatments

Section-level deterioration rates were computed as defined in Section 3.4.1 (Equation (1)), yielding 72,669–75,608 qualifying observed intervals without a documented treatment per indicator. Figure 3 shows the median annualized rate and interquartile range by the calendar year in which each interval ends.

Of the 75,608 qualifying IRI intervals pooled across the study period, 55.1% showed a positive (worsening) rate, with a median annualized rate of 0.89 in/mi per year. Rutting intervals ($N = 73{,}633$) were positive in 51.3% of cases, with a median rate of 0.0067 in per year. SC_DENSITY intervals ($N = 72{,}670$) were positive in 51.0% of cases, with a median rate of 0.0100 per year. FC_DENSITY intervals ($N = 72{,}669$) showed the highest share of worsening, at 57.7%, with a median rate of 0.14 per year. Because each section contributes several intervals, the intervals are not independent observations, and a significance test treating them as such would overstate the precision of these proportions; they are therefore interpreted descriptively, through their deviation from an even split, which is +5.1 percentage points for IRI, +1.3 for rutting, +1.0 for SC_DENSITY, and +7.7 for FC_DENSITY. Because the qualifying intervals differ in length, the shares were also examined by interval length. The distribution of lengths was nearly identical across the four indicators: 88.9% to 89.6% of intervals spanned one year, about 8% two years, and fewer than 3% three years or more, with fewer than 0.2% exceeding three years. Dividing by interval length does not change the sign of a change, so the share of worsening intervals is unaffected by annualization; it can nonetheless depend on length, because short intervals contain less accumulated change relative to measurement variability. Shares rose

with interval length for every indicator, from 49.4% to 57.0% among one-year intervals to 73.7% to 81.1% among intervals of three years or more, so the pooled shares, which are dominated by one-year intervals, are conservative. The comparison between indicators was unchanged: among one-year intervals, functional cracking worsened in 57.0% of cases, IRI in 54.2%, rutting in 50.8%, and structural cracking in 49.4%, and weighting each interval by its length gave 58.7%, 56.2%, 52.4%, and 52.9%, respectively, with functional cracking and IRI again showing the clearest tendency to worsen and rutting and structural cracking remaining closest to an even split.

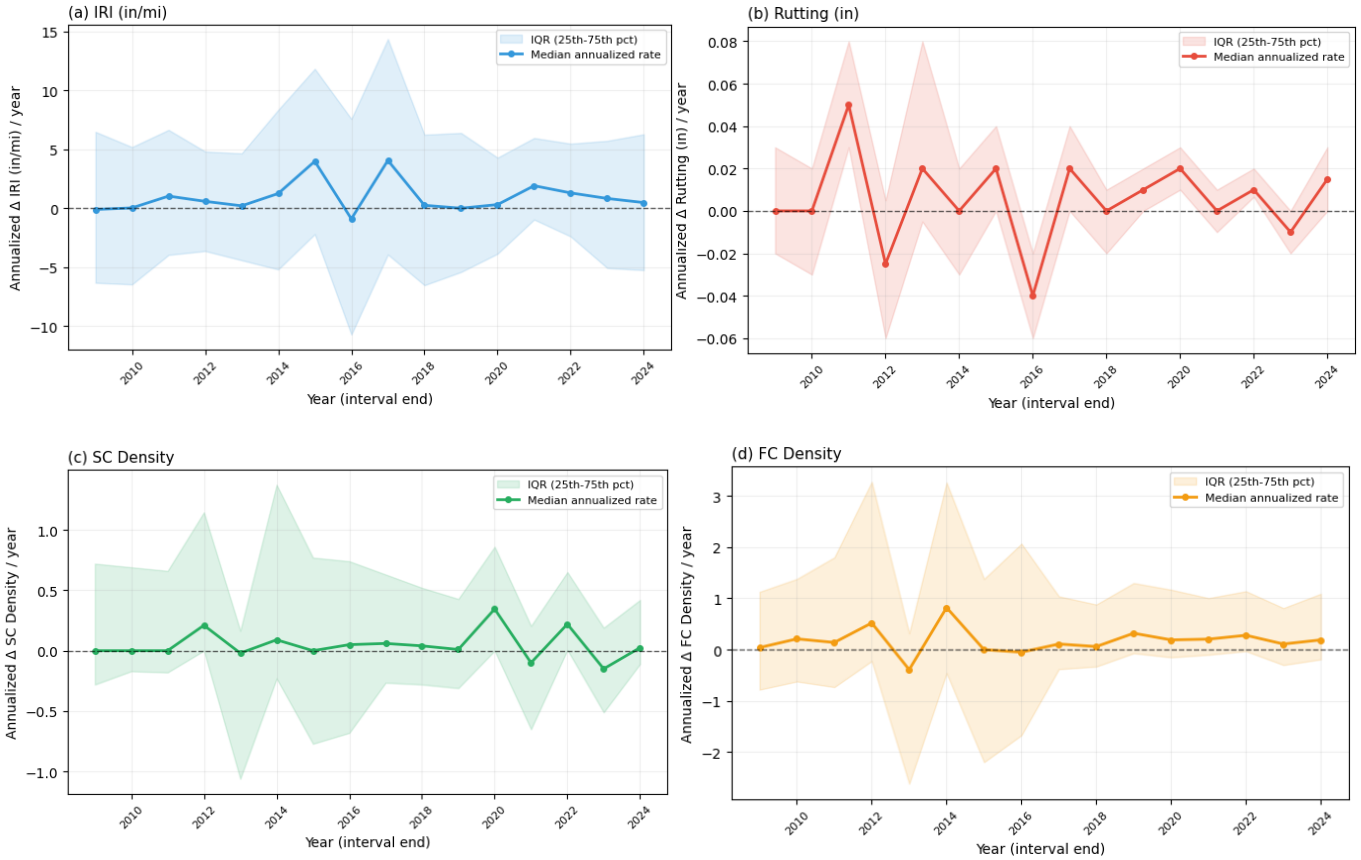


Figure 3. Annualized changes in pavement condition indicators between consecutive observed surveys during treatment-free intervals, 2008–2024. Intervals containing a documented treatment event were excluded. Positive values indicate worsening condition; the dashed line marks zero change. Shaded regions represent the interquartile range across intervals ending in each year. (**a**) IRI, (**b**) rutting, (**c**) structural cracking density, (**d**) functional cracking density.

These interval-level results show that deterioration between documented treatments is heterogeneous and tilted toward worsening, most clearly for functional cracking and roughness, while rutting and structural cracking density remain close to an even split. Because the units of the four indicators are not comparable (in/mi/year for IRI, in/year for rutting, density/year for the cracking indicators), the deviation from 50% and the direction of change are the meaningful basis for comparison across indicators here, not the raw magnitude of the rates. The interquartile ranges in Figure 3 frequently span zero, indicating substantial heterogeneity across pavement sections even where the median annualized rate is positive; the analysis therefore does not imply uniform monotonic deterioration at the section level but identifies a network-wide directional tendency superimposed on considerable section-specific variability. The interval and network-level results for IRI differ in sign and require reconciliation. The median treatment-free interval worsened by 0.89 in/mi per year, which, sustained over the 16 years of the study period, would corre-

spond to roughly 14 in/mi, whereas the treatment-adjusted network median declined by 6.8 in/mi between 2008 and 2024. The two quantities measure different things. The interval rate describes condition change between documented treatments only, and intervals containing a treatment, in which roughness is reset, are excluded by construction. The network median is a cross-sectional level: observations within two years of a treatment are excluded, but sections treated earlier re-enter the series carrying the lower roughness produced by that treatment, so the cumulative effect of treatment remains in the level, and its share grows as treatment activity accumulates over the study period. The most direct reconciliation is therefore that, in aggregate, the roughness removed by treatment exceeded the roughness accumulated between treatments, so that network ride quality was held near its 2008 level by maintenance rather than by an absence of deterioration. The comparison by time since treatment in Section 4.1.1 is consistent with this reconciliation: at a given time since treatment, IRI increased over the study period, and the network-level decline arose from the growing share of recently treated sections. For rutting and both cracking indicators, the network median and the interval rates both indicate worsening, implying that treatment did not offset between-treatment deterioration of these indicators to the same degree; this is consistent with treatment programs that respond primarily to roughness, a possibility the present data can suggest but not establish. Because the annual medians are computed over a changing set of observations as sections enter and leave the exclusion window, and the median of section-level changes need not equal the change in the network median, this comparison is one of direction rather than an exact accounting.

4.2. Indicator Relationships: Concurrent and Lagged Dependencies

4.2.1. Contemporaneous Correlations

The Pearson correlation matrix for the four condition indicators, computed across all section-year observations, is presented in Table 5. All six pairwise correlations are positive, ranging from 0.249 to 0.514, indicating that the indicators exhibit positive but varying degrees of contemporaneous association. The correlations are reported as descriptive measures of association; with more than 160,000 observations, conventional significance tests carry little information.

Table 5. Pearson correlation matrix for pavement condition indicators.

Indicator	IRI	Rutting	SC_Density	FC_Density
IRI	1.000	0.349	0.273	0.287
Rutting	0.349	1.000	0.249	0.277
SC_Density	0.273	0.249	1.000	0.514
FC_Density	0.287	0.277	0.514	1.000

Note: All correlations computed across 162,733–163,232 section-year observations depending on the indicator pair.

The strongest contemporaneous relationship is observed between SC_DENSITY and FC_DENSITY ($r = 0.514$, Table 5). Both indicators represent cracking-related pavement distress and may respond to overlapping traffic, environmental, material-aging, and moisture-related processes. Their moderate positive association therefore indicates substantial shared variation in cracking condition across the network, while the correlation remaining well below unity shows that the two measures retain distinct information. The contemporaneous correlation alone does not establish a directional progression from one cracking measure to the other; potential temporal ordering is examined separately through the cross-correlation analysis in Section 4.2.2.

IRI and rutting show the next strongest association (r = 0.349, Table 5), reflecting their shared sensitivity to traffic loading and structural deformation. The moderate rather than strong correlation is mechanistically expected: rutting quantifies localized permanent deformation within the wheel path, while IRI integrates longitudinal surface irregularities across a broader wavelength range that does not always align precisely with the rut trough geometry [7]. The two measures therefore capture overlapping but distinct aspects of structural deterioration, which is why sections with significant rutting can return acceptable IRI values and vice versa. Relying on either measure alone would consequently miss sections where the other has advanced beyond acceptable thresholds.

Relationships between roughness or deformation measures and cracking indicators are weaker—IRI versus SC_DENSITY yields r = 0.273 and IRI versus FC_DENSITY yields r = 0.287, while rutting versus SC_DENSITY and rutting versus FC_DENSITY yield r = 0.249 and r = 0.277, respectively (Table 5). These lower values indicate less shared same-year variation between roughness or deformation and cracking than within the cracking pair itself. The moderate correlation levels across all pairs indicate that although the indicators respond to a shared deterioration process, they do not provide equivalent information. Changes in one indicator do not consistently correspond to proportional changes in another—a statistical confirmation that each indicator captures deterioration information the others do not fully replicate, and the clearest quantitative argument for why single-indicator assessment is structurally insufficient for this network.

This correlation structure was tested for sensitivity to rehabilitation timing, since a treatment event can reset multiple indicators simultaneously and inflate their apparent association. The correlation matrix was recomputed on a treatment-adjusted subset excluding observations within two years of a documented treatment (103,171 of 165,372 observations, 62.4%). Table 6 compares the full-dataset and treatment-adjusted correlations for all six indicator pairs.

Table 6. Pearson correlation coefficients: full dataset versus treatment-adjusted observations.

Indicator Pair	Full Dataset	Treatment-Adjusted	Difference	N (Adjusted)
IRI – Rutting	0.349	0.329	−0.020	100,632
IRI – SC_DENSITY	0.273	0.274	+0.002	100,694
IRI – FC_DENSITY	0.287	0.268	−0.019	100,694
Rutting – SC_DENSITY	0.249	0.232	−0.016	100,923
Rutting – FC_DENSITY	0.277	0.254	−0.023	100,923
SC_DENSITY – FC_DENSITY	0.514	0.527	+0.013	101,034

Treatment adjustment produced only minor changes in the pairwise correlations, with absolute differences no greater than 0.023. The ordering of the relationships remained unchanged: SC_DENSITY–FC_DENSITY remained the strongest pair, IRI–rutting remained the next strongest, and the cross-family relationships between roughness/deformation and cracking remained weaker. The SC_DENSITY–FC_DENSITY correlation increased slightly from 0.514 to 0.527 after treatment adjustment, which is the opposite of what would be expected if simultaneous treatment-driven resets were inflating this relationship; the associations described above are therefore not primarily attributable to rehabilitation events resetting multiple indicators simultaneously.

4.2.2. Cross-Correlation and Lag Structure

Cross-correlation functions were computed for each section at lags from $k = -3$ to $k = +3$ years and summarized by the median across sections, for condition levels and for linearly detrended levels, first differences, and AR(1) pre-whitened series (Section 3.4.2); the

median lag-zero values under each transformation are reported in Table 7. Because linear interpolation smooths the series and could in principle shift lagged associations toward lag zero, the analysis was first repeated using only directly observed values. The peak remained at lag zero for all six pairs, and the median CCF at lag zero changed by no more than 0.029 relative to all records, increasing in five of six pairs; a complementary check restricted to 2008–2019, a window containing no COVID-19-era imputation, produced the same peak-lag structure. Interpolation therefore smoothed the lag profile slightly rather than generating the contemporaneous peak, and the results below use all records.

In condition levels, all six indicator pairs reach their maximum median cross-correlation at lag zero. The strongest relationship is between SC_DENSITY and FC_DENSITY (0.640), consistent with the strong contemporaneous association between the two cracking indicators reported in Section 4.2.1. Contemporaneous response is also mechanistically plausible where common stressors dominate, since environmental loading and traffic act on all distress types simultaneously rather than activating each mechanism in sequence [3,18].

Because condition levels trend over the study period, the lag-zero peak could in principle arise from shared trends rather than from contemporaneous coupling. Table 7 therefore also reports the median CCF at lag zero after detrending, differencing, and pre-whitening. Linear detrending left the lag structure essentially unchanged, with lag-zero values of 0.293 to 0.595 and the peak at lag zero for every pair. After first differencing or pre-whitening, correlations at non-zero lags were close to zero for all pairs (median CCF within ±0.15), and the peak remained at lag zero for every pair except rutting and functional cracking, for which no distinct peak emerges (Table 7, note). The contemporaneous coupling, however, separates into two parts. The two cracking indicators remain strongly coupled in year-to-year changes (lag-zero values of 0.395 to 0.522), whereas coupling between cracking, roughness, and rutting falls to 0.01 to 0.19. Much of the contemporaneous correlation between these indicator families in levels therefore reflects shared trends and persistent condition differences rather than same-year coupling of changes.

Table 7. Median cross-correlation at lag $k = 0$ for all six indicator pairs under alternative series transformations.

Indicator Pair	Levels	Detrended	Differences	Differences, Observed Only	Pre-Whitened
IRI & Rutting	0.364	0.323	0.111	0.108	0.188
IRI & SC Density	0.380	0.338	0.113	0.064	0.183
IRI & FC Density	0.426	0.378	0.094	0.052	0.190
Rutting & SC Density	0.343	0.293	0.075	0.077	0.169
Rutting & FC Density	0.381	0.326	0.052 [a]	0.011 [b]	0.168
SC Density & FC Density	0.640	0.595	0.459	0.395	0.522
Sections	~9560	~9550	~8370	~4570	~7190

Note: Values are median section-level CCFs at $k = 0$. Positive k indicates X leads Y. The peak occurs at $k = 0$ for every pair and transformation except where marked: [a] peak at $k = +1$ (0.060); [b] peak at $k = -3$ (0.071). Differenced and pre-whitened series exclude treatment-adjacent years; observed-only differences use consecutive directly observed years. Transformed series use sections defined by route, milepoints, and direction; under this definition the levels CCF reproduces the values shown within 0.02.

These patterns were assessed against sampling variation using bootstrap confidence intervals for the network median CCF at each lag and paired tests on section-level values (Figure 4). In levels, the confidence intervals are narrow (typically ±0.01), and the lag-zero CCF exceeds the CCF at both neighboring lags in every pair by 0.06 to 0.20 (95% confidence intervals excluding zero; Wilcoxon $p < 0.001$). After pre-whitening, the lag-zero CCF still exceeds both neighboring lags in every pair, by 0.03 to 0.25, except for rutting

and functional cracking, where the margin over a lag of one year is small (0.014; 95% CI 0.002 to 0.025). At the section level, where each pre-whitened CCF rests on at most 16 points, the lag-zero CCF exceeds the section's own 95% bound in 20% to 24% of sections for pairs of different indicator families and in 43% for the cracking pair, compared with 15% to 20% at lags of one year and 5% expected by chance. Individual series therefore carry limited information, and inference rests on the network-level distribution.

Small directional asymmetries are present in the three pairs involving rutting, and they point in the same direction: correlations are higher when rutting is taken one year earlier than the other indicator. In levels, the advantage of the lag at which rutting leads over the opposite lag is 0.041 for IRI, 0.039 for structural cracking, and 0.043 for functional cracking (95% confidence intervals excluding zero; Wilcoxon $p < 0.001$), and after pre-whitening it remains significant but smaller (0.017, 0.014, and 0.034). No material asymmetry was found for the pairs not involving rutting (median differences within ±0.007). Rutting therefore shows a small but consistent tendency to precede changes in roughness and cracking, plausibly reflecting the contribution of permanent deformation to subsequent roughness and wheel-path distress, but in every pair the correlation at lag zero remains larger than at either neighboring lag.

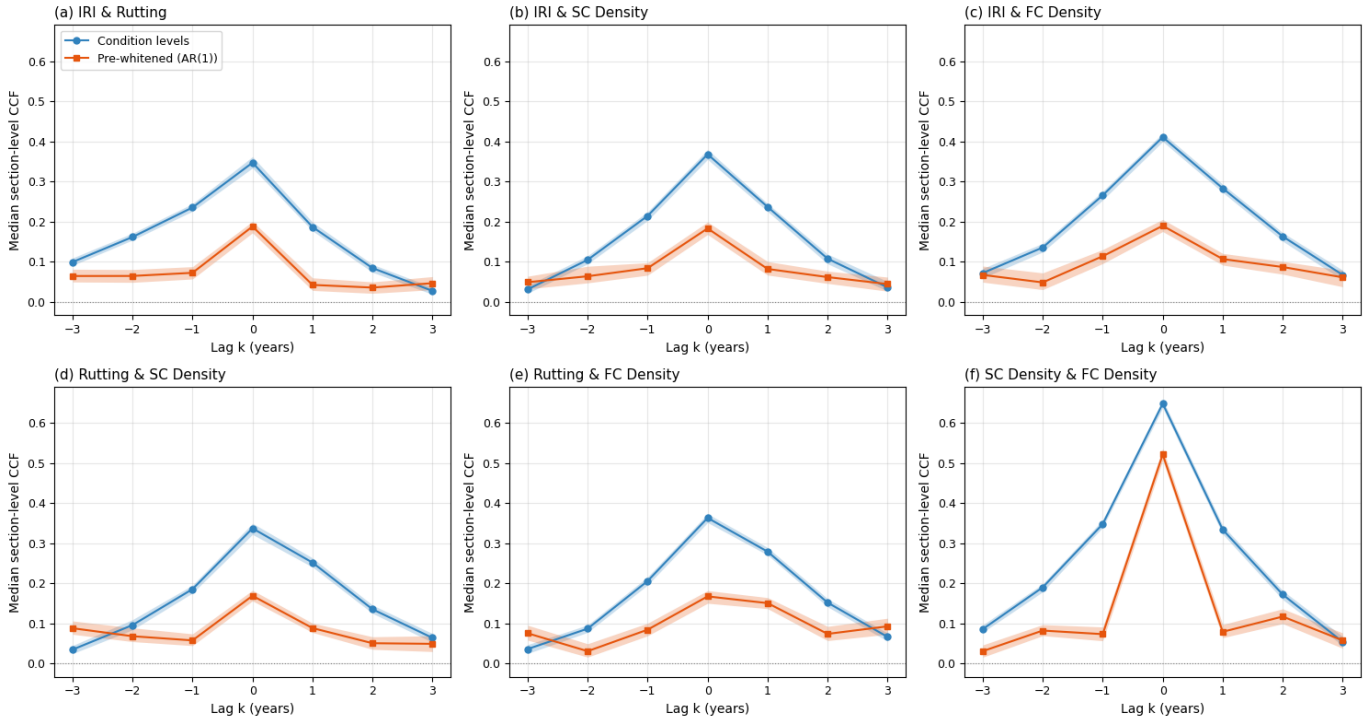


Figure 4. Network median section-level cross-correlation by lag for condition levels and AR(1) pre-whitened series. Shaded bands show 95% bootstrap confidence intervals (sections resampled, 1000 replicates), and the dotted line marks zero correlation. Positive lags indicate that the first-named indicator leads.

The distribution of section-level CCFs at each lag is shown in Figure 5. With all records, between 9519 and 9580 sections contribute at every lag, and the median number of lagged pairs per section falls by construction from 17 at lag zero to 14 at lags of three years. With directly observed values only, between 6992 and 7306 sections contribute at every lag, and the median number of pairs per section falls from 14 at lag zero to 11, 9, and 8 at lags of one, two, and three years, so the observed information at the longest lags is about two thirds of that at lag zero (approximately 66,000 against 98,000 lagged pairs). Section-level CCFs vary widely, as expected for series of 8 to 17 points: at lag zero the interquar-

tile range spans approximately 0.00 to 0.66 for pairs of different indicator families and 0.32 to 0.85 for the two cracking indicators. The whole distribution, however, shifts toward positive values at lag zero, not only its median. The share of sections with a positive CCF is 75% to 79% at lag zero for pairs of different indicator families and 90% for the cracking pair, falling to 53% to 60% at lags of three years. At the section level, the cross-correlation peaks at lag zero in 30% to 56% of sections, depending on the indicator pair. The remaining sections peak at other lags, partly reflecting noise in short series, but these section-level peaks do not produce a network-wide peak at any non-zero lag. Because heterogeneous section-level lags of opposite sign could cancel in the median, the section-level CCFs were also summarized by allowing each section its own preferred direction at each lag length, taking the larger of its correlations at $+k$ and $-k$. Since this takes the maximum of two noisy estimates, it is biased upward and gives an approximate upper bound on any lagged association that median aggregation could have cancelled. Even so, the folded correlation at a lag of one year (median 0.33 to 0.41 for pairs of different indicator families) only matched the lag-zero median (0.34 to 0.41), exceeding it by at most 0.02 in the pairs involving rutting, and it remained below the lag-zero median at two and three years (0.25 to 0.32); for the two cracking indicators it was well below at every lag (0.46 at one year against 0.65 at lag zero). The upper tail of the section-level distribution also declined with lag length, the 90th percentile falling from 0.75 to 0.94 at lag zero to 0.54 to 0.68 at three years. Lag zero was the most frequent section-level peak for every pair (30% to 56% of sections, against about 14% if peaks were spread evenly across the seven lags), and the remaining peaks were dispersed across lags, most often at three years, where each estimate rests on the fewest pairs. Where a section peaked away from lag zero, its peak exceeded its own lag-zero value by a median of about 0.2, within the sampling variability expected for correlations estimated from 8 to 17 points. The only directional excess was a higher share of sections peaking with rutting leading by one year in the three pairs involving rutting, consistent with the asymmetry reported above. Section-level lags therefore vary, as expected for short series, but no lag carries an association that, even allowing for cancellation, exceeds the contemporaneous one. The lag-zero peak is also insensitive to the length of the treatment-exclusion window (Section 4.3.7).

The interpretation of these results is bounded by the structure of the data. The annual survey cycle sets the temporal resolution of the analysis, and lead–lag relationships operating at monthly or seasonal timescales would be aliased into the same measurement cycle and appear contemporaneous [8,11]. Median aggregation across sections imposes a further constraint, as heterogeneous section-level lags can cancel when sections exhibiting positive lags offset those exhibiting negative lags for the same pair [9,12]. Maintenance activity introduces an additional confounding effect: treatment events such as crack sealing, resurfacing, and overlays reset multiple indicators within the same year, producing correlated condition changes driven by intervention rather than deterioration; this effect is addressed by excluding treatment-adjacent years from the differenced and pre-whitened analyses and is examined further in Sections 4.1.1 and 4.2.1. The analysis therefore shows that the association between indicators is predominantly contemporaneous at annual resolution, with only a small tendency for rutting to precede the other indicators. It does not exclude sequencing at finer timescales or at individual sections, which annual network-level data cannot resolve.

The operational consequence follows directly. Monitoring and programming decisions operate on the same annual cycle as the survey data, so even if sub-annual sequencing exists, it could not function as an early-warning signal within an annual monitoring program, and the small lead of rutting is too weak relative to the contemporaneous association to serve as a practical early-warning signal. Moreover, because year-to-year changes in

roughness and rutting are only weakly coupled to changes in cracking, a change observed in one indicator family carries little information about concurrent change in another. No single indicator can therefore be relied on as a proxy for developing deterioration in the others; condition assessment must measure the indicators concurrently, which motivates the dimensionality reduction analysis presented in Section 4.3.

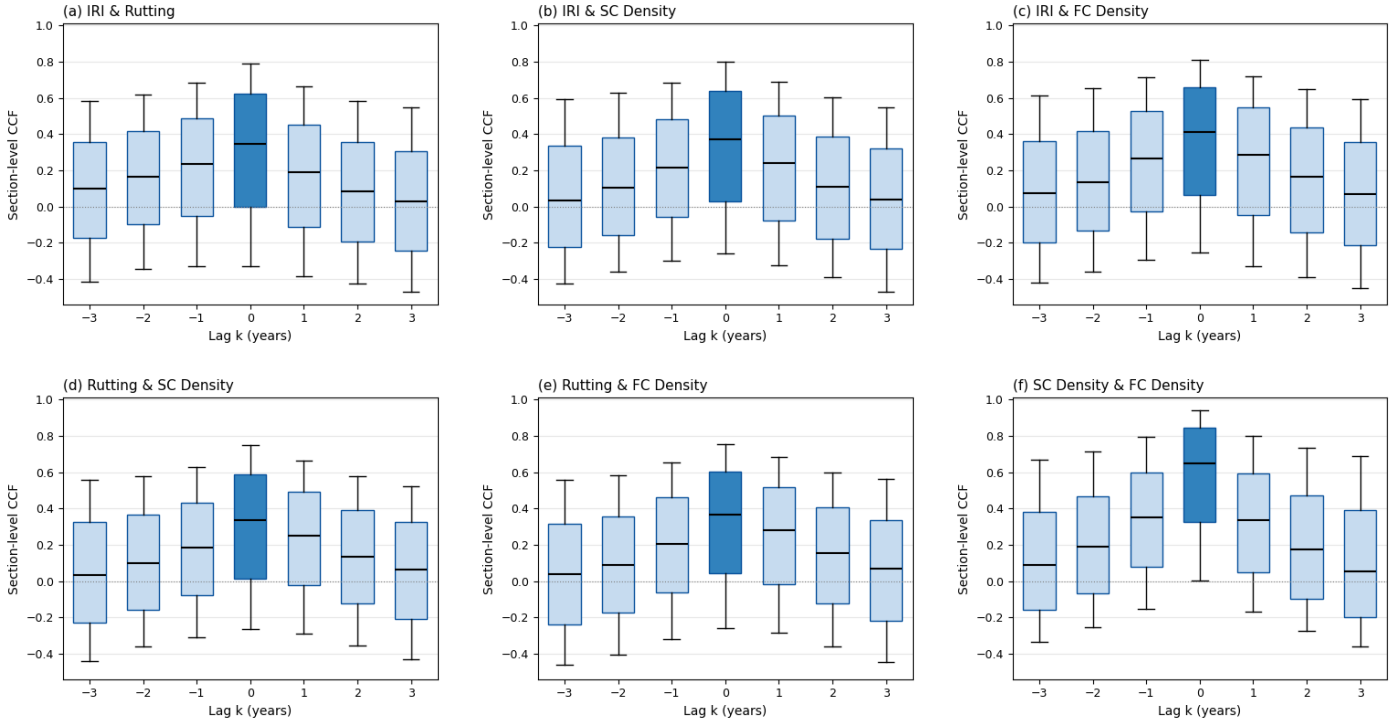


Figure 5. Distribution of section-level cross-correlations of condition levels by lag for the six indicator pairs (all records). Boxes show the interquartile range, whiskers the 10th to 90th percentiles, and the line the median; the lag-zero box is shaded darker, and the dotted line marks zero correlation. Between 9519 and 9580 sections contribute at every lag. Results for detrended, differenced, and pre-whitened series are summarized in Table 7.

4.3. Dimensionality of Pavement Deterioration: PCA Results

4.3.1. Component Retention via Parallel Analysis

Parallel analysis was applied to the standardized four-indicator dataset comprising 162,684 section-year observations to determine the number of statistically meaningful components underlying the shared variation among IRI, rutting, SC_DENSITY, and FC_DENSITY. The results are presented in Table 8.

Table 8. Parallel analysis results for PCA component retention.

Component	Observed Eigenvalue	Simulated 95th Percentile	Variance Explained (%)	Retained
PC1	1.982	1.011	49.5	Yes
PC2	0.883	1.006	22.1	No
PC3	0.650	1.002	16.3	No
PC4	0.485	0.997	12.1	No

Only PC1 has an eigenvalue above its simulated threshold (1.982 against 1.011). With 162,684 observations, the simulated 95th-percentile eigenvalues are close to 1, so at this sample size parallel analysis gives essentially the same result as the Kaiser criterion.

The size of PC1 is also largely determined by the overall level of correlation among the indicators: for four indicators with equal pairwise correlations at the observed mean of 0.325 (Table 5), the first eigenvalue would be 1.975 (49.4% of total variance), close to the observed 1.982 (49.5%). The single dominant component therefore primarily reflects the positive correlation among the indicators, and the more informative results are how the indicators contribute to PC1 and how that composition differs between road classes (Sections 4.3.2 and 4.3.6). PC1 accounts for 49.5% of total variance. The remaining 50.5% is distributed across PC2 to PC4, each of which is a contrast between at least two indicators rather than variance specific to a single indicator; PCA does not partition variance into common and indicator-specific parts, so this remainder is reported only as variance not captured by PC1. The size of this residual nonetheless shows that a single composite does not reproduce the individual indicators, which supports retaining the indicators alongside PC1 rather than replacing them. PC1 is used in the following sections as a composite condition reference.

4.3.2. Component Structure and Contributions

The coefficients, loadings, contributions, and squared loadings of PC1 for the full network are presented in Table 9.

Table 9. PC1 coefficients, loadings, contributions, and squared loadings for the full network.

Indicator	Coefficient a_{1j}	Loading ℓ_{1j}	Contribution c_{1j} (%)	Squared Loading ℓ_{1j}^2
FC_DENSITY	0.543	0.764	29.4	0.583
SC_DENSITY	0.529	0.745	28.0	0.555
IRI	0.468	0.659	21.9	0.434
RUT	0.455	0.640	20.7	0.409
Sum	–	–	100.0	1.982 ($=\lambda_1$)

Note: a_{1j} is the weight of indicator j in the PC1 score, with $\sum_j a_{1j}^2 = 1$; $\ell_{1j} = a_{1j}\sqrt{\lambda_1}$ is the correlation between indicator j and PC1; $c_{1j} = a_{1j}^2$ is the share of PC1 variance attributable to indicator j; ℓ_{1j}^2 is the share of the variance of indicator j captured by PC1. Contributions describe the composition of PC1 and are not measures of physical importance or causal priority.

All four indicators have positive coefficients on PC1, so higher PC1 scores correspond to concurrently poorer condition across the indicators. The loadings range from 0.640 for rutting to 0.764 for FC_DENSITY, so every indicator is strongly correlated with the PC1 score. Expressed as contributions to the variance of PC1, the two cracking indicators account for 57.5% (FC_DENSITY 29.4%, SC_DENSITY 28.0%) and IRI and rutting for 42.5% (21.9% and 20.7%). No indicator contributes less than one fifth or more than three tenths of the component, against an equal-contribution reference of 25%, so PC1 is not dominated by a single indicator, although it is moderately weighted toward cracking.

The squared loadings give the complementary view: PC1 accounts for 58.3% of the variance of FC_DENSITY but 40.9% of the variance of rutting, so the composite represents the cracking indicators more completely than it represents roughness and deformation. PC1 is therefore interpreted as a general condition axis to which all four indicators contribute substantially. These contributions are a statistical property of the standardized data and should not be read as measures of the engineering significance of each indicator. Standardization expresses every indicator in units of its own network standard deviation, so a contribution describes how strongly an indicator's variation co-moves with that of the others, not the consequence of that variation for safety, structural capacity, ride quality, or maintenance cost. A one-standard-deviation increase in IRI and in structural cracking

density carry equal weight on this scale but need not have equal engineering consequence, and the contributions themselves change with the scaling applied to the indicators, as shown in Section 4.3.6.

4.3.3. Distribution of PC1 Scores Across the Network

The distribution of PC1 scores across all 162,684 observations is shown in Figure 6. The distribution is right-skewed (skewness = 1.99), with a mean of 0.00 and a median of −0.362. The interquartile range spans −0.981 to 0.562, indicating that the central 50% of observations is concentrated within a 1.54-unit range of the PC1 scale, while a smaller number of sections extend into substantially higher values reaching a maximum of 15.49.

The right-skewed distribution implies that most sections fall within low to moderate levels of the composite condition measure, while a smaller subset exhibits significantly higher PC1 values. These high-PC1 sections correspond to cases where multiple indicators show elevated values simultaneously, reflecting more severe overall condition states. Deterioration is not uniformly distributed across the network. Sections with high PC1 values represent cases where multiple distress types co-occur, which may not be fully captured when indicators are evaluated individually. This provides further support for the multi-indicator framework examined in Section 4.4.

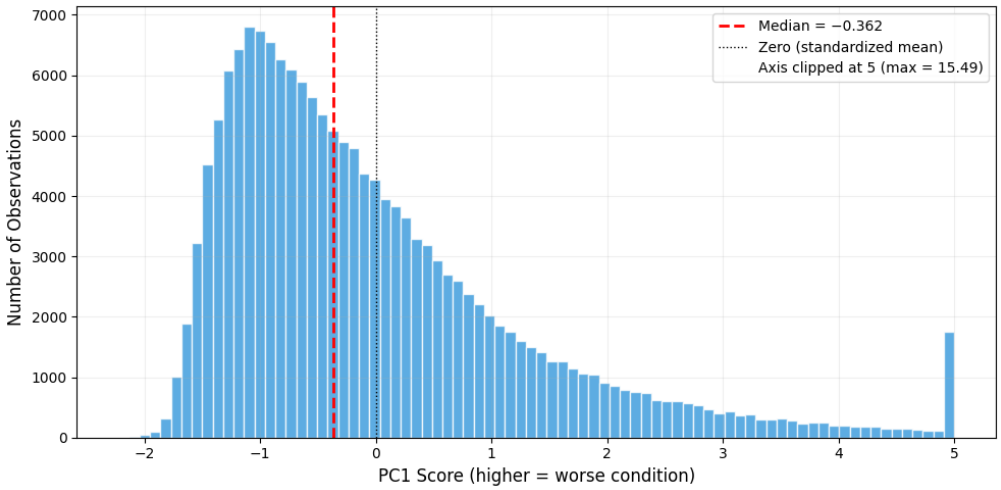


Figure 6. Distribution of PC1 scores across the network (n = 162,684). The horizontal axis is clipped at 5.

4.3.4. Indicator Behavior Across PC1 Quartiles

To confirm that PC1 represents meaningful differences in pavement condition, the distribution of each raw indicator was examined across PC1 quartiles (Q1 = best condition, Q4 = worst). All four indicators increase monotonically from Q1 to Q4 in both median and spread (Figure 7). For IRI, the median increases from approximately 67 in/mi in Q1 to 163 in/mi in Q4. Rutting increases from approximately 0.12 in to 0.24 in. Structural cracking increases from near zero in Q1 to approximately 5.4% in Q4, while functional cracking increases from approximately 0.34% to 7.86%. The monotonic progression confirms that PC1 provides an ordered, physically meaningful representation of pavement condition rather than a mathematical artifact. The quartile distributions, however, reveal a pattern that single-indicator summaries do not capture. IRI and rutting in Q4 retain wide interquartile ranges with substantial overlap against Q2 and Q3 values (Figure 7a,b). A significant share of Q4 sections fall below 200 in/mi in IRI—within acceptable bounds under standard roughness-based thresholds—and Q4 rutting medians do not approach typical intervention thresholds for permanent deformation. The cracking indicators tell a different

story: SC_DENSITY and FC_DENSITY show near-zero medians in Q1 and Q2 before rising sharply in Q3 and Q4 (Figure 7c,d). Sections that PC1 identifies as severely deteriorated are not uniformly severe across all indicators—they are sections where cracking has advanced significantly while roughness and deformation measures remain moderate [6,21]. This pattern is relevant to prioritization. A system based on IRI thresholds alone would rank Q4 sections in which cracking is the primary elevated distress below their composite condition. These are not borderline cases: they are sections where multiple cracking indicators have simultaneously reached elevated levels while roughness remains within acceptable bounds. Deferring intervention until IRI crosses a trigger threshold risks allowing cracking to progress from stages where preservation treatments are effective into stages requiring structural rehabilitation, which are reported to carry substantially higher unit costs [53,54]. The PC1 composite reflects this cracking directly in the condition ranking at empirically determined weights and therefore indicates the type of deterioration as well as its overall level, which roughness-only assessment does not capture. The Q4 profile, with elevated cracking, moderate IRI, and low rutting, suggests the hypothesis that surface treatments targeting ride quality alone may not be adequate for these sections; because no treatment-effectiveness data were analyzed, this is not tested here, and, as noted in Section 4.3.5, whether prioritization based on PC1 leads to better maintenance outcomes is beyond the scope of the present study.

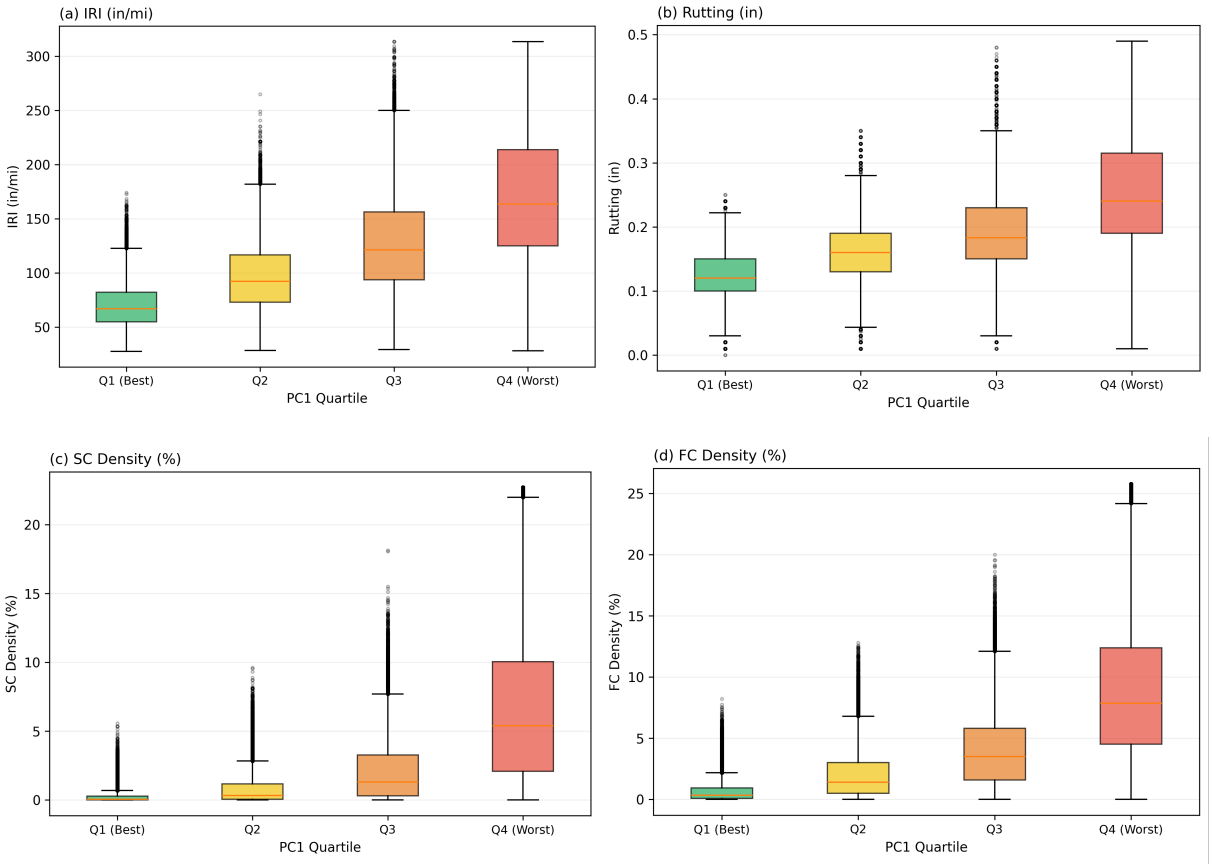


Figure 7. Indicator behavior across PC1 quartiles. Boxes show the interquartile range and the orange line the median; whiskers extend to 1.5 times the interquartile range, and points beyond them are individual observations: (**a**) IRI, (**b**) rutting, (**c**) SC density, (**d**) FC density.

4.3.5. Prioritization Under PC1 and Minimum-Based RSL Aggregation

The MDOT dataset includes a remaining service life measure, RSL_4, of the kind used in pavement management systems to prioritize sections. RSL_4 is constructed as the mini-

mum of four component service-life estimates (IRI_RSL, RUT_RSL, SCD_RSL, FCD_RSL), a relationship that holds exactly for all 162,684 observations, and each component is a near-deterministic monotonic transform of its own raw indicator via the MDOT breakpoint table (Spearman ρ between -0.93 and -0.997). RSL_4 and PC1 are therefore derived from the same four underlying indicators but represent them differently: PC1 combines standardized (z-scored) indicator values through a covariance-derived linear combination, whereas RSL_4 first converts each indicator into an indicator-specific service-life estimate and then takes the minimum across all four. Because each RSL component remains strongly monotonic with its corresponding raw indicator, the two measures capture closely related underlying condition information, but the comparison that follows reflects both the transformation and the aggregation architecture, not aggregation alone. The two measures are, as expected, strongly associated ($\rho = -0.818$); the question of practical interest is where and why their resulting priority classifications diverge.

A high-priority set was defined under each rule as the worst quartile of PC1 scores and the quartile of RSL_4 with the fewest remaining years. Because RSL_4 is reported in discrete service-life values, ties at the lower-quartile boundary (11 years) produced a high-priority set slightly larger than exactly 25% of observations (40,859 versus 40,671 for PC1). A fixed share was used rather than an operational trigger value because PC1, as a composite, has no agency-defined threshold; flagging the same share under each rule ensures that disagreement reflects which observations each rule ranks worst rather than differences in the size of the flagged sets and corresponds to budget-constrained programming, in which the worst-ranked sections are funded until the available budget is exhausted. Sensitivity to this share is reported below. The rules agree on 81.1% of section-year observations, but because most observations fall outside both high-priority sets, much of this agreement is expected from the marginal classification rates alone; Cohen's $\kappa = 0.497$ indicates that substantial disagreement remains after accounting for that baseline. The disagreement is nearly symmetric: 15,284 observations are flagged only by PC1 (37.6% of all PC1-flagged observations) and 15,472 only by RSL_4 (37.9% of all RSL_4-flagged observations). Of the 56,143 observations flagged by either rule, 30,756 (54.8%) are flagged by only one, and the two rules agree on 25,387 (45.2%).

The two rules disagree on systematically different kinds of observations. Counting, for each section-year, how many of the four indicators fall simultaneously in their own network-wide worst quartile (Figure 8), observations flagged only by RSL_4 are overwhelmingly single-indicator cases: 73.1% have exactly one elevated indicator and only 19.5% have two or more. Observations flagged only by PC1 are overwhelmingly multi-indicator cases: 91.8% have two or more elevated indicators, with 45.5% having exactly two and 36.4% exactly three. A small share of RSL_4-only observations (7.4%) have no indicator in the worst quartile, reflecting that the RSL breakpoint thresholds do not coincide exactly with the quartile cut-off used here.

The distress profiles of the two groups follow directly (Table 10; medians are reported because the cracking indicators are strongly right-skewed). Observations flagged only by PC1 carry elevated values on both cracking indicators (median SC_DENSITY 6.85, FC_DENSITY 8.16) with moderate roughness (137 in/mi). Observations flagged only by RSL_4 show the reverse: higher roughness (153 in/mi) with cracking at the network baseline (median SC_DENSITY 0.23, FC_DENSITY 1.36, comparable to the 0.32 and 1.26 of observations flagged by neither rule). Observations flagged by both approaches show elevated median values across all four indicators; the substantive disagreement therefore concerns observations prioritized by only one approach, not these unambiguous cases.

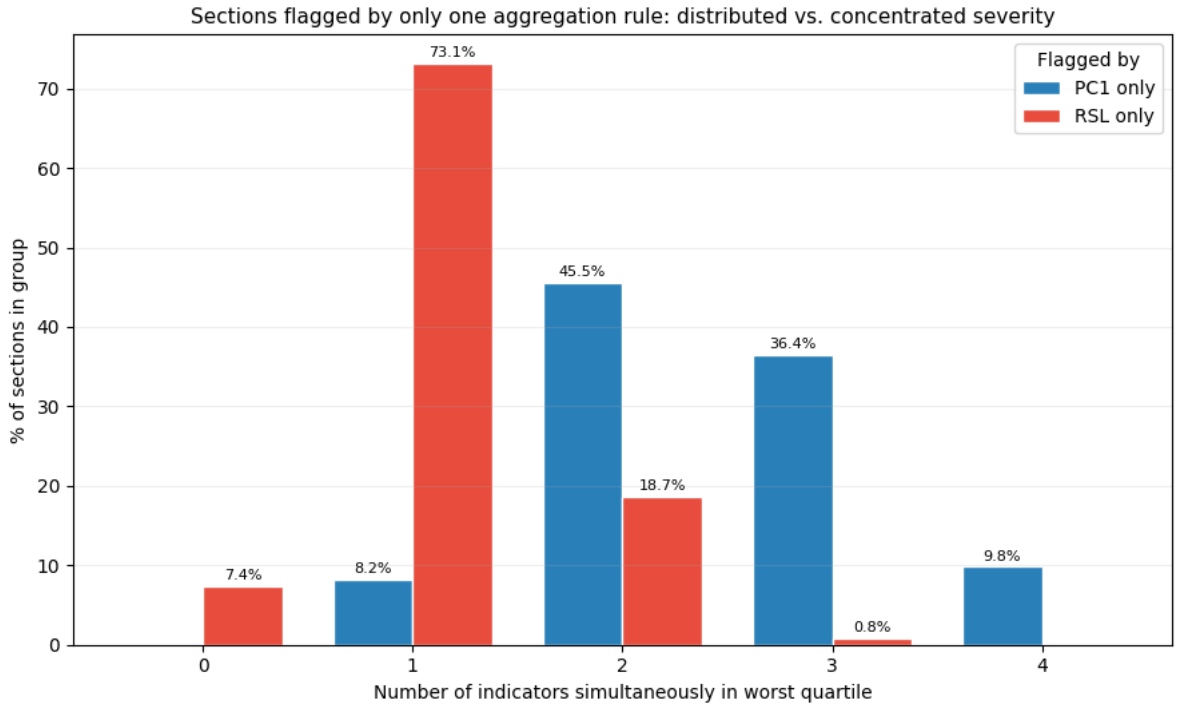


Figure 8. Number of indicators simultaneously in the network-wide worst quartile, for observations flagged as high-priority by only one aggregation rule.

Table 10. Median indicator values (interquartile range in parentheses) by agreement group between the PC1 composite and the minimum-based RSL_4 measure.

Group	IRI (in/mi)	Rutting (in)	SC_DENSITY	FC_DENSITY	N
Flagged by PC1 only	137.3 (113.1–164.9)	0.22 (0.18–0.26)	6.85 (4.02–10.02)	8.16 (5.38–11.43)	15,284
Flagged by RSL_4 only	153.3 (107.5–184.5)	0.18 (0.14–0.26)	0.23 (0.01–0.83)	1.36 (0.42–3.48)	15,472
Flagged by both	196.3 (140.9–241.4)	0.27 (0.20–0.36)	4.11 (1.37–10.09)	7.55 (3.93–13.45)	25,387
Flagged by neither	85.4 (65.9–110.9)	0.15 (0.12–0.18)	0.32 (0.04–1.46)	1.26 (0.32–3.38)	106,541

Which component determines the RSL_4 minimum confirms this reading (Table 11). Among RSL_4-only observations with a unique limiting component, IRI_RSL controls the minimum in 67.3% of cases and RUT_RSL in 31.0%; the two cracking components control fewer than 2% combined. The unique flags of RSL_4 in this network are therefore almost entirely roughness- or rutting-triggered, a pattern that is traced below largely to the MDOT breakpoint table rather than to the minimum rule itself. Among PC1-only observations with a unique limiting component, SCD_RSL (40.9%) or FCD_RSL (21.2%) is more often the controlling component, yet in these observations no component is restrictive enough to enter the RSL_4 high-priority quartile: cracking is elevated on both indicators, but neither crosses the threshold the minimum rule requires.

Table 11. Component controlling the RSL_4 minimum, by agreement group (percentage of observations with a unique limiting component; 90–100% of observations in each group).

Group	IRI_RSL	RUT_RSL	SCD_RSL	FCD_RSL
Flagged by RSL_4 only	67.3%	31.0%	1.0%	0.7%
Flagged by PC1 only	19.5%	18.4%	40.9%	21.2%
Flagged by both	49.5%	31.4%	9.5%	9.6%

The direction of the disagreement shown in Figure 8 and Table 10 follows from the mathematical form of the two rules rather than from the data. A minimum-based measure is determined by the component with the shortest estimated remaining service life; once that component becomes limiting, the values of the remaining three do not affect the aggregate score unless one of them becomes still more restrictive. PC1, in contrast, changes continuously with the standardized condition of all four indicators. A weighted sum therefore necessarily favors observations in which several indicators are elevated, and a minimum necessarily favors observations in which one indicator is extreme. This is confirmed on a common z-scale, where 94.5% of observations flagged only by the weighted sum have two or more indicators in their worst quartile, against 14.8% of those flagged only by the minimum. The empirical content of the comparison lies instead in the magnitude of the disagreement and in which indicators control the minimum in this network (Table 11).

Because PC1 and RSL_4 differ both in how the indicators are scaled, through standardization or the MDOT breakpoint table, and in how they are combined, through a weighted sum or a minimum, their disagreement reflects both choices. The two were separated by comparing rules that differ in only one of them (Table 12). Holding the scale fixed, the aggregation rule alone changes a substantial share of the high-priority set: at the worst 25%, a weighted sum and a minimum applied to the same z-scores disagree on 32.8% of the observations either flags, and the same two rules applied to the four RSL components disagree on 57.1%. Holding the rule fixed, the breakpoint transform alone produces disagreement of 25.5% under a weighted sum and 43.0% under a minimum. The disagreement between PC1 and RSL_4 therefore arises from both choices, with the aggregation rule contributing substantially on either scale and the breakpoint transform amplifying it under the minimum rule. The transform also determines which indicator limits the minimum: among the worst 25% under RSL_4, IRI or rutting is the limiting component in 88.4% of observations, whereas a minimum applied to z-scores is limited by IRI or rutting in 57.7% and by a cracking indicator in 42.3%. The predominance of roughness and rutting in RSL_4 thus largely reflects the MDOT breakpoint table rather than the minimum rule itself.

Table 12. Disagreement between prioritization rules at alternative high-priority thresholds: percentage of observations flagged by either rule that are flagged by only one.

Comparison	Scale	Rule	Worst Share Flagged			
			10%	20%	25%	30%
PC1 vs. RSL_4 (primary)	differs	differs	73.0	59.8	54.8	48.3
Weighted sum vs. minimum, z-scores	same	differs	43.1	35.9	32.8	30.2
Weighted sum vs. minimum, RSL components	same	differs	74.1	61.4	57.1	50.3
z-scores vs. RSL components, weighted sum	differs	same	36.5	28.1	25.5	23.1
z-scores vs. RSL components, minimum	differs	same	59.3	46.8	43.0	37.3

Note: Each rule flags its worst 10%, 20%, 25%, or 30% of the 162,684 observations, with ties at the threshold included. Weighted sum on z-scores = PC1; weighted sum on RSL components = first principal component of the four component RSL values; minimum on z-scores = largest of the four standardized indicators; minimum on RSL components = RSL_4.

The disagreement is not specific to the quartile cut. It persists at every threshold examined and is largest at the most selective: between PC1 and RSL_4 it falls from 73.0% of the observations flagged by either rule at the worst 10% to 48.3% at the worst 30%, and the aggregation rule alone on z-scores follows the same pattern (43.1% to 30.2%).

Neither representation is inherently preferable for every management objective. The minimum rule suits a decision philosophy in which any single limiting condition can trigger intervention, whereas PC1 is more responsive to the accumulation of distress across indicators. The distinction is relevant to prioritization because the preceding analy-

ses show that deterioration in this network is not expressed uniformly across indicators. Functional cracking exhibited the strongest directional worsening in the treatment-adjusted interval analysis, and the composite's most deteriorated quartile was characterized by advanced cracking without comparably extreme roughness. The PC1-only group identified here exhibits the same profile, with median SC_DENSITY and FC_DENSITY several times those of the RSL_4-only and neither-priority groups. Cracking of this kind is generally reported to progress from stages treatable by preservation into stages requiring structural rehabilitation at substantially higher unit cost [53,54], and under RSL_4 these observations are not flagged until a single component, whether cracking, roughness, or rutting, reaches the high-priority threshold on its own. For a management objective that explicitly seeks to identify concurrent deterioration across several condition dimensions, PC1 is therefore better aligned by construction than RSL_4, which does not represent that profile directly. Conversely, RSL_4 identifies roughness- and rutting-dominated observations that PC1 may rank below its highest-priority quartile, so the two measures should be read as emphasizing different condition profiles rather than as interchangeable measures of priority. Establishing which rule produces better maintenance outcomes would require subsequent treatment, performance, and cost data and is beyond the scope of the present analysis.

4.3.6. Stratified PCA by Functional Class Group

To assess whether the weighting of PC1 is consistent across road types, PCA was applied separately within three functional class groups defined by FHWA hierarchy and pavement use context. High-Order roads comprise Interstate and Principal Arterial roads (FC 1, 2, 3, 11, 12, 14; n = 99,626), which typically carry the highest traffic volumes and are designed to the most stringent structural standards. Intermediate roads comprise Minor Arterial and Collector roads (FC 6, 7, 16, 17; n = 61,782), which serve moderate traffic and connect arterials to the local network. Low-Order roads comprise Minor Collector and Local roads (FC 9, 18, 19; n = 1276), which typically carry the lightest traffic and serve primarily access functions. The codes, their definitions in the MDOT data dictionary supplied with the condition data, and their group assignments are listed in Table 3. Codes 3 and 18 occur in the condition records but are not defined in that dictionary; they were classified from the MDOT SHA functional classification of the routes concerned [50], which identifies the single route coded 3 as a principal arterial and the three routes coded 18 as minor collectors. The resulting PC1 contributions are compared with the full-network result in Figure 9 and Table 13.

Parallel analysis was applied independently within each group to the within-group correlation matrix, with simulated datasets matching each group's sample size (500 simulations per group), so that the retention threshold adapts to group size. A single component was retained in all three groups, consistent with the full-network result. Applied directly to indicators in network units, parallel analysis would retain two components for Low-Order roads, because the four indicators have within-group variances of 1.36 to 2.40 on that scale, which inflate the eigenvalues relative to the unit-variance reference distribution; on the within-group correlation matrix, the second eigenvalue (0.91) falls below its threshold (1.07). The within-group correlation matrix is therefore used for retention in all groups. In network units, the composition of PC1 differs markedly between High-Order and Intermediate roads (Table 13, Figure 9). On High-Order roads, IRI and rutting contribute 33.3% and 27.8% and the two cracking indicators 38.9% combined; on Intermediate roads, the cracking indicators contribute 79.3% and IRI 8.6%. The section-level cluster bootstrap intervals for the two groups do not overlap for any indicator. PC1 also explains a larger share of variance on Intermediate roads (56.5%) than on High-Order roads (48.6%), indicating that the four indicators covary more strongly on Intermediate roads.

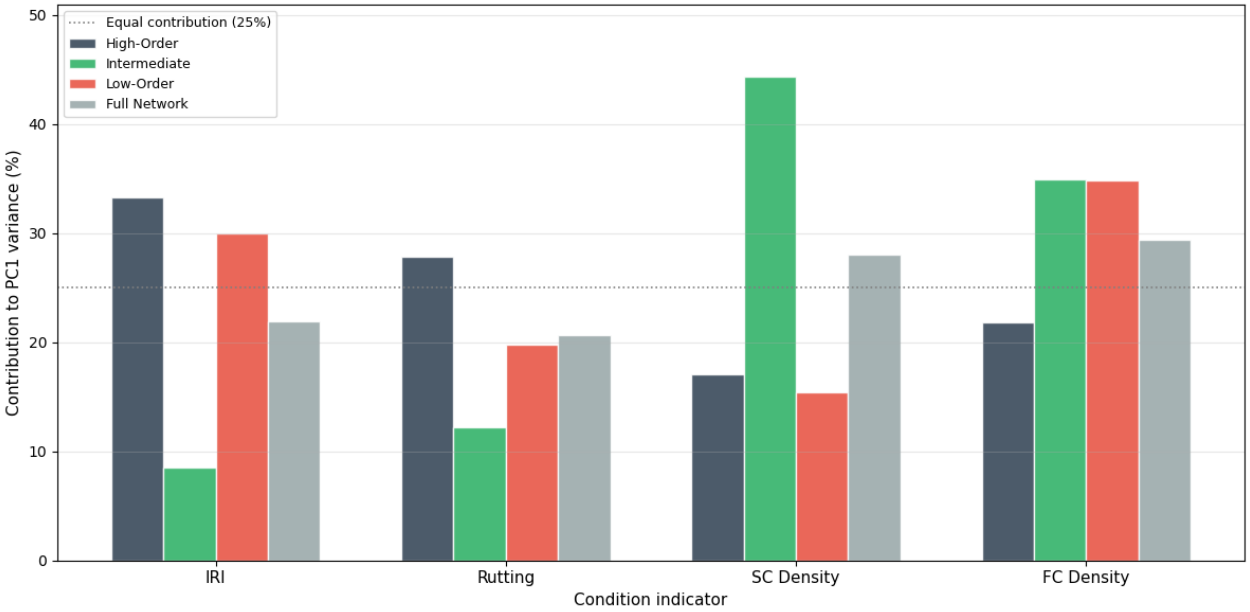


Figure 9. PC1 contributions (%) by functional class group.

Table 13. Stratified PCA: PC1 contributions (%) in network units, with section-level cluster bootstrap 95% confidence intervals, by functional class group.

Indicator	High-Order	Intermediate	Low-Order	Full Network
IRI	33.3 [30.1–36.4]	8.6 [7.4–9.7]	29.9 [11.0–55.0]	21.9 [19.9–24.0]
Rutting	27.8 [24.5–31.3]	12.2 [10.7–13.6]	19.8 [7.9–29.0]	20.7 [18.6–22.9]
SC_DENSITY	17.1 [14.6–20.3]	44.3 [40.8–47.4]	15.4 [7.4–28.1]	28.0 [25.6–30.8]
FC_DENSITY	21.8 [18.6–25.4]	35.0 [31.4–39.4]	34.9 [15.5–54.2]	29.4 [26.7–32.1]
Cracking total	38.9 [34.0–44.3]	79.3 [77.1–81.6]	50.3 [23.5–79.8]	57.5 [53.8–60.8]
PC1 variance explained	48.6%	56.5%	61.7%	49.5%
Components retained	1	1	1	1
Section-years (sections)	99,626 (5866)	61,782 (3638)	1,276 (76)	162,684 (9580)

Note: Indicators standardized with network-level mean and standard deviation. Components retained determined by parallel analysis on the within-group correlation matrix. Confidence intervals from 1000 cluster bootstrap replicates resampling whole sections.

These differences combine two effects. Because the indicators are expressed in network units, an indicator that varies more within a group contributes more to that group's PC1. On High-Order roads, IRI and rutting vary more than across the network as a whole (variance ratios of 1.16 and 1.12) and the cracking indicators less (0.82 and 0.96), whereas on Intermediate roads the pattern is reversed (0.69 and 0.74 against 1.21 and 1.03). The correlation structure also differs: the two cracking indicators are more strongly correlated on Intermediate roads than on High-Order roads ($r = 0.59$ versus 0.47), and rutting is more closely associated with cracking ($r = 0.35$ versus 0.24).

The magnitude of the class difference depends on how the indicators are scaled (Table 14). When indicators are standardized within each group, which removes the effect of within-group spread, cracking contributes 55.0% of PC1 on High-Order roads and 60.2% on Intermediate roads. On rank-transformed indicators in network units the corresponding values are 54.3% and 66.6%, and on ranks standardized within group, 59.7% and 63.0%. The direction of the difference is consistent across all four specifications: cracking contributes more, and IRI less, on Intermediate roads than on High-Order roads. The large contrast in the primary solution, however, depends partly on extreme values, and in three of the four specifications the cracking indicators account for more than half of PC1 even on

High-Order roads. The difference between road classes is therefore one of degree in the relative weight of each indicator, not a change in which indicator family dominates.

Table 14. Sensitivity of PC1 composition to scaling and rank transformation: cracking total and IRI contribution (%) by functional class group.

Specification	Cracking Total (%)		IRI (%)	
	High-Order	Intermediate	High-Order	Intermediate
Raw values, network units (primary)	38.9	79.3	33.3	8.6
Raw values, within-group	55.0	60.2	23.9	18.1
Ranks, network units	54.3	66.6	29.1	11.9
Ranks, within-group	59.7	63.0	22.3	18.3

Note: Network units = indicators standardized with the network-level mean and standard deviation; within-group = standardized separately within each functional class group; ranks = network-wide ranks of each indicator, then standardized. Cracking total = SC_DENSITY + FC_DENSITY, computed before rounding.

This pattern is consistent with the structural design hierarchy of these road classes. Although structural properties and traffic loading were not measured in this study, High-Order roads are typically built to higher structural numbers, with thicker asphalt layers and stronger base courses, and repeated heavy loading on these facilities acts strongly on roughness and permanent deformation [18,55]. Intermediate roads are typically designed to lower structural numbers and lighter traffic, and their thinner structures reach fatigue thresholds at lower cumulative axle loads, so that cracking can emerge early and progress faster than roughness [6,56,57]. The greater spread and stronger coupling of cracking on Intermediate roads, and the larger contribution of IRI on High-Order roads, are consistent with these expectations, but the mechanisms are offered as interpretation rather than as findings of the analysis.

For Low-Order roads, comprising 1276 section-years from 76 sections, the cracking indicators contribute 50.3% of PC1 in network units. The cluster bootstrap interval for the cracking total is wide (23.5% to 79.8%), and results for this group are therefore reported descriptively only.

The full-network contributions (20.7% to 29.4%) therefore represent an average over road classes whose indicator weighting differs, rather than a structure that applies to any individual class. A single network-wide composite does not represent all functional class groups equally, while cracking contributes substantially to the condition axis in every class; applications of the framework to specific road classes would benefit from class-specific component solutions.

4.3.7. Treatment Robustness of the PC1 Structure

The covariation summarized by PC1 could in principle reflect rehabilitation events that reset several indicators within the same year rather than covariation arising from deterioration. To test this, the full-network and stratified PCA solutions were recomputed on the treatment-adjusted subset, excluding observations within two years of a documented treatment. The treatment-adjusted temporal dataset contained 103,171 observations, of which 100,583 had complete values across all four indicators and were included here. The same network-level standardization parameters were used for the full and treatment-adjusted solutions, so that any change in the contributions reflects the removal of treatment-adjacent observations rather than a rescaling of the indicators. Tucker's congruence coefficient ϕ was computed between each treatment-adjusted component and its full-dataset counterpart, with values above 0.95 indicating high congruence. The results are summarized in Table 15.

Table 15. Treatment robustness of PC1: contributions (%) and variance explained for the full dataset and the treatment-adjusted subset, by functional class group.

Indicator	Full Network Full → Adjusted	High-Order Full → Adjusted	Intermediate Full → Adjusted	Low-Order Full → Adjusted
IRI	21.9 → 22.4	33.3 → 32.2	8.6 → 7.2	29.9 → 30.4
Rutting	20.7 → 18.2	27.8 → 25.0	12.2 → 9.2	19.8 → 17.9
SC_DENSITY	28.0 → 25.9	17.1 → 18.2	44.3 → 37.7	15.4 → 16.2
FC_DENSITY	29.4 → 33.6	21.8 → 24.6	35.0 → 45.9	34.9 → 35.5
Cracking total	57.5 → 59.5	38.9 → 42.8	79.3 → 83.6	50.3 → 51.7
PC1 variance explained (%)	49.5 → 48.9	48.6 → 48.3	56.5 → 55.3	61.7 → 62.7
Components retained	1 → 1	1 → 1	1 → 1	1 → 1
Tucker's ϕ (PC1)	0.9987	0.9990	0.9936	0.9997
N observations	162,684 → 100,583	99,626 → 69,211	61,782 → 30,349	1276 → 1023

Note: Contributions in network units. Components retained determined by parallel analysis on the within-group correlation matrix.

Treatment adjustment left the full-network solution essentially unchanged in form. Parallel analysis again retained a single component, PC1 variance explained changed from 49.5% to 48.9%, and Tucker's congruence between the full and treatment-adjusted PC1 was $\phi = 0.9987$. Because contributions are squared coefficients, they respond more visibly than the coefficients themselves: the largest change was an increase for FC_DENSITY from 29.4% to 33.6%, offset mainly by rutting and SC_DENSITY, while all four indicators continued to contribute between 18.2% and 33.6%.

The class-specific solutions were similarly stable, with ϕ above 0.99 for PC1 in every functional class group and a single component retained in each. On High-Order roads, IRI and rutting remained the two largest individual contributors at 32.2% and 25.0%. On Intermediate roads, the cracking total increased from 79.3% to 83.6%, with contribution moving from SC_DENSITY (44.3% to 37.7%) to FC_DENSITY (35.0% to 45.9%), the largest change of any indicator in any group. Intermediate roads also lost the largest share of observations under treatment adjustment, retaining 30,349 of 61,782 (49.1%) compared with 69.5% for High-Order roads. The increase in FC_DENSITY after these observations were removed is consistent with Section 4.1.1, where treatment-adjacent observations were found to suppress the apparent magnitude of functional cracking. The Low-Order solution was also stable ($\phi = 0.9997$), although its small adjusted sample (1023 observations from 76 sections) warrants the caution noted in Section 4.3.6.

Exclusion of treatment-adjacent observations therefore left the number of retained components, the variance explained, and the direction of the class differences in PC1 composition unchanged. The PC1 structure is not primarily driven by rehabilitation events resetting several indicators simultaneously, although, as noted in the Limitations, treatment adjustment does not remove all sources of confounding.

The two-year exclusion window was also varied to test whether the PCA and cross-correlation results depend on it. With no exclusion and with windows of zero (treatment year only), one, two, and three years, retaining between 54% and 100% of observations, PC1 explained 48.9% to 49.6% of total variance and reproduced the all-records component closely, with Tucker's $\phi \geq 0.997$ for the full network and $\phi \geq [0.990]$ for High-Order and Intermediate roads. The median section-level cross-correlation peaked at $k = 0$ for every pair in every window, with exceptions only for rutting and functional cracking at $k = +1$: in condition levels under the two-year window, where the value at $k = +1$ exceeded that at $k = 0$ by 0.005, and in first differences, where this pair shows no distinct peak (Table 7, note). Both correspond to the small tendency for rutting to precede the other indicators

noted in Section 4.2.2. Neither the PC1 structure nor the contemporaneous peak therefore depends on the choice of exclusion window.

4.3.8. Temporal Stability of the PC1 Structure

To assess whether the PC1 structure identified in the full-network PCA is stable across time, PCA was applied separately to three sub-periods: 2008–2012, 2013–2017, and 2018–2024. Each sub-period was standardized independently and evaluated via parallel analysis using 500 simulations. Independent standardization was used here, unlike the treatment robustness test in Section 4.3.7, because the aim is to compare the correlation structure within each period rather than to isolate the effect of removing observations from a fixed sample. Tucker's congruence coefficient ϕ was computed between each period's PC1 coefficients and the full-network coefficients to quantify structural similarity; values exceeding 0.95 indicate high congruence between component solutions. The results are summarized in Table 16.

Table 16. Temporal stability of PC1 contributions and variance explained across sub-periods.

Indicator	Full Network	2008–2012	2013–2017	2018–2024
IRI (%)	21.9	24.1	22.0	19.5
Rutting (%)	20.7	15.9	23.5	22.1
SC_DENSITY (%)	28.0	28.3	27.2	29.0
FC_DENSITY (%)	29.4	31.7	27.3	29.4
Cracking total (%)	57.5	60.0	54.5	58.4
PC1 variance explained (%)	49.5	49.8	48.7	51.4
Components retained	1	1	1	1
Tucker's ϕ vs. full network	–	0.9979	0.9993	0.9995
N observations	162,684	47,855	47,845	66,984

Note: Each sub-period standardized independently. Tucker's $\phi > 0.95$ indicates high congruence between period and full-network PC1.

Parallel analysis retained a single component in all three periods, consistent with the full-network result. PC1 variance explained was 49.8% for 2008–2012, 48.7% for 2013–2017, and 51.4% for 2018–2024, against the full-network value of 49.5%, a range of less than 3 percentage points across the 17-year study period. The congruence coefficients were $\phi = 0.9979$, 0.9993, and 0.9995 for the three periods respectively, all substantially exceeding the 0.95 threshold. The cracking indicators contributed between 54.5% and 60.0% of PC1 in every period, and SC_DENSITY and FC_DENSITY remained the two largest contributors throughout. The main variation was a lower contribution from rutting in 2008–2012 (15.9%, against 22.1% to 23.5% in the later periods), accompanied by a gradual decline in the contribution of IRI from 24.1% to 19.5%. No period produced a change in which indicators dominate the component.

The 2018–2024 sub-period contains 19,484 observations (29.1% of that window) corresponding to COVID-19-affected years estimated through linear interpolation. To evaluate whether this interpolation influenced the loading structure, the 2018–2024 PCA was repeated excluding 2020 and 2021. The resulting contributions differed by no more than 0.24 percentage points from the full-period solution, PC1 variance explained was identical at 51.4%, and Tucker's ϕ between the two solutions was 1.0000. The PC1 structure in the 2018–2024 period is therefore not sensitive to the inclusion or exclusion of COVID-19-interpolated observations.

Collectively, these results show that the single retained component, the predominance of the cracking indicators within it, and the PC1 variance explained are stable properties of the Baltimore County pavement network rather than artifacts of any particular phase of the

study period or of the COVID-19 interpolation window. Sensitivity to interpolation across the full study period is evaluated separately in the Limitations.

4.3.9. Sampling and Repeated-Measures Stability of the PC1 Structure

Sensitivity of the PC1 structure to sampling variation was evaluated with a cluster bootstrap of 1000 replicates in which whole pavement sections were resampled with replacement from the globally standardized dataset (9580 sections for the full network; 5866, 3638, and 76 for the High-Order, Intermediate, and Low-Order groups). Each section contributes up to 17 annual observations, and resampling sections rather than individual section-years preserves this dependence, so that the intervals reflect the effective amount of independent information in the data. PCA was applied to each replicate with sign alignment to the reference solution, and 95% confidence intervals were taken as the 2.5th and 97.5th percentiles of the bootstrap distributions. Intervals for the contributions are reported in Table 13.

For the full network, PC1 variance explained was 49.5% (95% CI 48.9% to 50.2%), and no contribution interval was wider than 5.4 percentage points; the cracking total was 57.5% (53.8% to 60.8%). Between road classes, the intervals for High-Order and Intermediate roads did not overlap for any indicator or for the cracking total, which was 38.9% (34.0% to 44.3%) on High-Order roads and 79.3% (77.1% to 81.6%) on Intermediate roads. PC1 variance explained also differed beyond sampling variation, at 48.6% (47.7% to 49.6%) and 56.5% (55.3% to 57.7%), respectively. The difference in PC1 composition between these groups is therefore not attributable to sampling variation, although, as shown in Section 4.3.6, its magnitude depends on how the indicators are scaled. For the Low-Order group, with 76 sections, the intervals are wide (cracking total 23.5% to 79.8%; variance explained 54.5% to 70.1%), consistent with its descriptive treatment.

The repeated-measures structure was examined directly by separating variation between sections from variation within sections over time (Table 17). The share of each indicator's variance lying between sections differs markedly: 78.5% for IRI and 51.1% for rutting, but 38.3% and 38.1% for structural and functional cracking. Roughness therefore largely distinguishes persistently rough sections from smoother ones, whereas most of the variation in cracking occurs within sections over time. Despite this difference, PCA on section means and on within-section deviations each retained a single component and reproduced the pooled solution closely ($\phi = 0.9992$ and 0.9998). Because PCA on section means contains no variation over time, the between-section result shows that the PC1 structure is present in cross-sectional co-variation alone and is not generated by network-wide time trends; the network trends would in any case act against it, since median IRI declined while cracking increased over the study period, which would induce a negative rather than a positive association between roughness and cracking. When each sample contained one randomly selected observation per section, removing repeated measures entirely, a single component was retained in all 200 samples, PC1 explained a median of 49.6% of variance (95% range 48.7% to 50.5%), and congruence with the pooled solution was at least 0.9997. The PC1 structure is therefore not an artifact of repeated observation of the same sections. Within-section centering on the treatment-adjusted subset reproduced the pooled solution ($\phi = 0.9986$), and PCA on annualized increments between consecutive directly observed surveys with no documented treatment produced a first component congruent with the pooled solution ($\phi = 0.9911$), although it explained less variance (40.1%) and parallel analysis retained two components, as expected for year-to-year changes that carry a larger share of measurement variability. The second component of the increments explains 25.5% of their variance and contrasts the two cracking indicators (loadings 0.36 and 0.40) with IRI and rutting (-0.50 and -0.68), with the same pattern when only one-year intervals are used. Year-to-year deterioration therefore separates into a cracking dimension and a rough-

ness and rutting dimension, consistent with Section 4.2.2, where the cracking indicators remained strongly coupled in first differences while their coupling with roughness and rutting fell to 0.01 to 0.19.

At the functional class level, the direction of the difference between High-Order and Intermediate roads holds at both levels. Between sections, the cracking indicators contribute 18.5% of PC1 on High-Order roads and 51.1% on Intermediate roads, and IRI 60.0% and 35.4%; within sections, the cracking indicators contribute 74.5% and 86.1%. The larger contribution of IRI on High-Order roads thus reflects differences between sections, whereas year-to-year variation within sections is dominated by cracking in both classes. Because the within-section deviations were computed on the full dataset, they also include condition changes associated with treatment.

Table 17. PC1 structure at the pooled, between-section, and within-section levels, with one observation per section, and for observed deterioration increments.

Analysis	N	PC1 Var. (%)	Comp.	Tucker's ϕ	IRI (%)	Rutting (%)	SC (%)	FC (%)	Cracking Total (%)
Pooled (primary)	162,684	49.5	1	–	21.9	20.7	28.0	29.4	57.5
Between sections	9580	51.1	1	0.9992	25.1	19.4	26.6	28.9	55.4
Within sections	162,684	49.4	1	0.9998	20.3	20.9	29.1	29.8	58.9
Within sections, treatment-adjusted	100,561	47.0	1	0.9986	19.1	18.4	30.9	31.6	62.5
One observation per section	9580	49.6	1	≥0.9997	21.9	20.7	28.0	29.4	57.4
Observed increments, treatment-free	70,044	40.1	2	0.9911	18.3	12.4	35.4	33.9	69.2

Note: Between sections = PCA on section means; within sections = PCA on deviations from section means; one observation per section = median of 200 samples, each with one randomly selected year per section; observed increments = annualized changes between consecutive directly observed surveys with no documented treatment within the interval. Comp. = components retained by parallel analysis. Tucker's ϕ for PC1 relative to the pooled solution.

4.4. Discussion

Most operational pavement management systems rest on two assumptions that this study tests directly. The first is that pavement deterioration can be adequately represented by a single indicator, most commonly IRI, on the grounds that roughness accumulates as a downstream consequence of structural distress and therefore serves as a reliable proxy for overall network condition [58,59]. The second is that condition thresholds and maintenance triggers can be applied uniformly across a mixed road network, treating all functional classes as though they deteriorate through the same mechanisms at comparable rates [60,61]. Classical deterioration theory reinforces both assumptions by treating distress mechanisms as largely separable, with individual performance models developed for each distress type independently [62,63]. The implicit premise is that monitoring one indicator provides sufficient information to characterize the overall condition state of a pavement section and that this logic holds regardless of road class. The results of this study challenge both assumptions empirically. The four indicators are positively correlated, and a single dominant component, PC1, accounts for 49.5% of their total variance. All four indicators contribute substantially to it (20.7% to 29.4%) and are strongly correlated with it (loadings of 0.640 to 0.764), although the component is moderately weighted toward cracking, which accounts for 57.5% of its variance. Roughness, rutting, and cracking therefore tend to be elevated together, consistent with the expectation that deterioration mechanisms are not fully independent, and no single indicator dominates the shared axis. Half of the total variance nonetheless lies outside PC1, distributed across components that each contrast two or more indicators, so it cannot be attributed to distinct deterioration mechanisms or to any single indicator. The moderate pairwise correlations (r = 0.249 to 0.514) do establish, however,

that the indicators are not interchangeable: no one indicator, and no single composite, reproduces the condition information carried by the others. This shared structure persists after treatment adjustment: recomputing PC1 on a subset excluding observations within two years of a documented treatment reproduces the full-network component almost exactly (Tucker's $\phi = 0.9987$), and the direction of the class differences described below is unchanged ($\phi > 0.99$ for every functional class group). The covariance structure summarized by PC1 is therefore unlikely to be primarily generated by simultaneous rehabilitation-related changes in the condition indicators.

A closely comparable structure has been reported for physical transport infrastructure indicators: Saba et al. [64] applied PCA to three national-level transport infrastructure measures across 102 countries and retained a single dominant component explaining 56.9% of total variance from pairwise correlations of 0.237–0.495, a correlation range and variance share closely matching those observed here ($r = 0.249$–0.514; 49.5%). Because a first component of this size follows largely from the level of the pairwise correlations (Section 4.3.1), this agreement reflects similar correlation strength rather than a common underlying structure, and the single-component result is not, in itself, specific to pavement deterioration. This data-derived weighting differs in purpose from the fixed, expert-assigned deduct weights embedded in conventional composite indices such as PCI [13,14]. Deduct values encode judgments about the engineering consequence of each distress type and severity level, whereas PC1 contributions describe how the indicators covary in the observed data. The composite therefore complements rather than replaces severity-based weighting. It does, however, address the concern raised by Gharaibeh et al. [16] that composite scores can mask very different physical states. This concern is examined directly by comparing PC1 against RSL_4, a remaining-service-life measure already present in the dataset and constructed as the minimum of four component service-life estimates built from the same four indicators. Because the two measures share inputs, their disagreement reflects how the indicators are scaled and combined rather than differences in the underlying data. When the two effects are separated, the aggregation rule alone changes about one third of the high-priority set with the scale held fixed (32.8% on z-scores), while the MDOT breakpoint transform both amplifies this disagreement and makes the minimum predominantly roughness- and rutting-driven. The two rules disagree on 54.8% of the observations either flags as high priority; 37.6% of the observations flagged by PC1 and 37.9% of those flagged by RSL_4 are not flagged by the other rule, and the disagreement is systematic. Observations the minimum rule uniquely flags are overwhelmingly single-indicator cases (73.1% have exactly one indicator in its network-wide worst quartile, and IRI_RSL or RUT_RSL is the limiting component in 98.3% of those with a unique limiting component), while observations PC1 uniquely flags are overwhelmingly multi-indicator cases (91.8% have two or more indicators simultaneously elevated, most often led by cracking). By construction, a minimum-based rule ranks an observation with one severe problem above one with several moderately elevated indicators, whereas a weighted composite can rank the latter as equally or more deteriorated. As a further check on the composite's specificity, skid resistance, excluded from PCA because it reflects aggregate polishing rather than the load-induced and environmental processes represented by the other four indicators, shows only weak association with PC1 ($\rho = -0.155$), comparable to IRI alone and weaker than rutting alone ($\rho = -0.208$), consistent with skid resistance being governed by a distinct mechanism rather than the shared condition axis PC1 captures.

The cross-correlation analysis reinforces this from a temporal perspective: all six indicator pairs peak at $k = 0$ in levels, detrended, and pre-whitened series, as do all pairs except rutting and functional cracking in differenced series, and the only consistent asymmetry, a small tendency for rutting to precede the other indicators by about one year,

is much weaker than the contemporaneous association. The absence of a strong lead–lag structure warrants interpretation against the mechanistic literature on deterioration sequencing. Roughness progression models represent IRI growth as the downstream outcome of accumulated damage, a framing under which surface distress would be expected to precede measurable roughness change [6], and lagged effects at the distress level have been reported on timescales of days to weeks [11]. The network-level evidence here does not contradict such fine-timescale dynamics; rather, it indicates that at annual survey resolution and network aggregation, sequencing of this kind is not observable as a strong lag between condition indicators, consistent with sub-annual dependencies being aliased into contemporaneous co-movement. This distinction matters for practice: mechanistic sequence at fine timescales does not translate into an exploitable early-warning lag at the temporal resolution at which agencies actually collect condition data. The differenced and pre-whitened analyses add a second practical point. Once shared trends are removed, year-to-year changes remain strongly coupled between the two cracking indicators but only weakly between cracking and roughness or rutting, so the indicator families do not simply rise and fall together from one survey to the next. This is consistent with the divergence observed in the temporal record, where median IRI declined while cracking increased, and it means that a change in roughness carries little information about concurrent change in cracking. The treatment-adjusted temporal analysis sharpens rather than weakens the case for concurrent, multi-indicator monitoring. Once observations within two years of a documented treatment are excluded, median IRI still declines over the study period while rutting, structural cracking, and functional cracking all increase; because the unadjusted data understated functional cracking by an average of 10.8% across the study period, and by as much as 48.2% in individual years, the treatment-adjusted divergence between roughness and cracking is at least as pronounced as in the unadjusted analysis. Section-level evidence corroborates this at finer grain: among genuinely observed, treatment-free survey intervals, functional cracking density worsened in 57.7% of cases and IRI in 55.1%, compared with closer to an even split for rutting and structural cracking (51.3% and 51.0%, respectively), indicating that network-wide deterioration in cracking and roughness reflects real, if heterogeneous, within-section change between interventions, not merely compositional drift in which sections are observed. The divergence between median FC_DENSITY rising from 1.96 to 2.23 while median IRI declined from 110.6 to 103.8 in/mi over the same period makes the cost of ignoring this concrete: a roughness-only monitoring program would have read this network as improving even as surface deterioration continued to spread across a growing share of sections [6,7]. The magnitude of the contemporaneous correlations observed here is consistent with prior network-level evidence that pavement indicators do not evolve uniformly and should not be used interchangeably [7]. The weak rutting–cracking correlations ($r = 0.249$–0.277) align with the mechanistic expectation that shear deformation and fatigue damage develop through largely independent pathways [33]. Where earlier LTPP-based analyses relied on a single indicator because distress data could not be systematically combined (Section 2.1), the PCA composite provides one empirically derived aggregation of the four indicators, and the moderate correlations ($r = 0.249$–0.514) quantify why no single indicator, roughness included, can stand in for the others over a 17-year record.

The stratified PCA is where the assumption of uniform network-wide thresholds is most directly challenged. In network units, the composition of PC1 differs systematically between road classes: the cracking indicators contribute 79.3% of PC1 on Intermediate roads against 38.9% on High-Order roads, and IRI 8.6% against 33.3%. The cluster bootstrap intervals for the two groups do not overlap, and the difference persists after treatment adjustment, under which the cracking total on Intermediate roads rises to 83.6%. Its magnitude, however,

depends on how the indicators are scaled. When within-group spread or the influence of extreme values is removed, the cracking share on High-Order roads rises to between 54.3% and 59.7%, and the gap between the two classes narrows to between 3.3 and 12.3 percentage points. The robust finding is therefore one of direction rather than dominance: cracking carries more weight and IRI less on Intermediate roads than on High-Order roads in every specification, while in three of the four specifications cracking accounts for more than half of PC1 even on High-Order roads. Roughness-based monitoring therefore does not capture the full condition signal on either class, and its shortfall is largest on Intermediate roads.

The direction of this difference is consistent with the structural design hierarchy of the road classes. Intermediate roads are typically built to lower structural standards than High-Order facilities, and thinner structures reach fatigue thresholds at lower cumulative axle loads, so that cracking can emerge early and progress faster than roughness [6,56,57], whereas heavier traffic on High-Order roads acts more strongly on roughness and permanent deformation [18,55]. Layer thickness, structural number, traffic, and subgrade data were not included in the dataset, so these mechanisms are offered as interpretation; the difference in PC1 composition itself is an empirical property of the condition data and does not depend on them. The pattern is consistent with studies reporting that cracking governs deterioration on thinner, lower-order pavements and that the relationship between roughness and distress varies by road type and functional class (Section 2.1) [7,56,65,66]. The stratified PCA extends these observations from pairwise relationships to the full multi-indicator composite and quantifies the class-dependent weighting directly from network condition data.

A uniform IRI threshold applied across a mixed network is therefore likely to underrepresent cracking-dominated deterioration on every road class, but most strongly on Intermediate roads, where cracking can advance without producing proportional roughness change. The shortfall would be directional rather than random, falling most heavily on the road classes that serve the connectivity function between arterials and local roads [67]. Because no consistent annual lead–lag relationship was observed, roughness cannot be relied upon as an annual early-warning substitute for cracking. These findings support concurrent multi-indicator monitoring, particularly on classes where cracking contributes more strongly to the condition axis. These findings are relevant to how PMS frameworks treat mixed networks. Rather than applying a single network-wide IRI threshold, the stratified PCA results suggest that class-specific composites may be appropriate, in which indicator contributions are estimated separately for each functional class group (Table 13). Because the class-specific solutions are expressed in network units, their scores remain on a common scale and can be ranked together in a single network-wide prioritization list. These weights are derived directly from observed network data rather than assigned through expert judgment or design guide defaults, making the approach transparent, auditable, and updatable as new condition data accumulate. Because these contributions reflect statistical covariation rather than engineering consequence, operational use would combine them with agency criteria such as intervention thresholds, safety requirements, and treatment costs, rather than treat them as measures of indicator importance in their own right. Section-level PC1 scores within each class-specific solution could serve as a complementary input to existing threshold-based prioritization, although their effect on maintenance outcomes was not evaluated in this study. The comparison with RSL_4, the minimum-based measure already available in the dataset, illustrates this complementary role: PC1 identifies a distinct set of sections characterized by concurrent multi-indicator distress that a minimum rule does not flag until a single component reaches its threshold on its own. Neither measure is inherently preferable for every management objective, since a minimum rule reflects the reasonable premise that a pavement fails at its weakest point,

but for an objective that explicitly seeks to identify concurrent deterioration across several indicators, PC1 provides information a minimum-based measure does not. The PC1 quartile analysis further suggests that this reranking has diagnostic implications, not just prioritization implications: Q4 sections show elevated cracking with moderate roughness, indicating that prioritization based only on roughness could miss sections where cracking warrants separate engineering evaluation [53,54,66].

The PC1 composite has properties relevant to emerging digital twin and AI-assisted PMS frameworks, most of which currently rely on single-indicator state definitions, typically IRI, that this study has shown to be incomplete for mixed networks [68,69]. The composite is low-dimensional, empirically grounded, physically interpretable, and derived from the observed covariance structure of multi-indicator data, properties consistent with what such state variables require. This is a structural argument, not a demonstrated one: the composite was developed and evaluated exclusively on annual network-level survey data, and its behavior under real-time or high-frequency sensor-driven monitoring has not been tested. Whether the PC1 structure identified here is stable at sub-annual sampling rates, or whether sensor noise characteristics would require re-estimating the component, remains open. The class-specific solutions extend the same argument, offering class-conditional state representations that reflect the indicator weighting of each functional class group, again without sensor-level validation. For AI-assisted condition prediction, graph neural network frameworks and sequence-based predictive models depend on the richness of the condition representation supplied to them, and the present framework offers one structured, replicable method for constructing such a representation from standard network-level survey data, without requiring specialized sensors or instrumented test sections [70,71]. Its extension to sensor-based digital twin platforms is a direction for future work rather than a claim of this study.

As transportation agencies invest in data-driven infrastructure management, the analytical pipeline presented here may serve as a transparent link between conventional PMS data collection and the richer condition representations that next-generation decision-support systems are expected to require.

Limitations

Several limitations bound the conclusions of this analysis. Treatment history was incorporated throughout the temporal and PCA analyses (Sections 4.1, 4.2.1 and 4.3.7), substantially reducing the direct influence of rehabilitation-related condition resets. However, treatment adjustment does not remove all confounding. Traffic loading, climatic exposure and structural design variables were not included in the dataset and may contribute to the observed section-level and functional-class heterogeneity. The analysis also covers a single network, and the PC1 structure, including the indicator contributions and the proportion of variance explained, may not generalize to networks with different pavement compositions, traffic profiles, or climatic environments. The magnitude of the class differences in PC1 composition also depends on how the indicators are scaled and on the treatment of extreme values in the right-skewed cracking distributions (Section 4.3.6); the direction of these differences is robust, but class-specific contributions should be re-estimated rather than transferred when the framework is applied to other networks. The 2024 observation count is lower than earlier years, introducing potential sample bias in terminal-year statistics. The section-level deterioration-rate analysis in Section 4.1.2 is not affected by interpolation because interpolated observations are explicitly excluded; a rate computed between two observations is nonetheless a linear approximation of that interval and cannot capture non-monotonic condition changes occurring between the two surveys, including abrupt deterioration or short-term measurement variability. For the 7370 sections covered by the raw

pre-imputation survey file, 16.5% to 19.1% of records per indicator were filled by interpolation, and missingness of IRI was moderately associated with prior roughness (Section 3.2). The sensitivity of the PCA and cross-correlation results to this interpolation was therefore evaluated by recomputing each analysis using only directly observed values. Restricting PCA to section-years with all four indicators directly observed (N = 98,079 of 162,684) left the PC1 structure materially unchanged (Tucker's $\phi = 0.9987$, single component retained). This count is lower than the interpolation rates in Section 3.2 would suggest for two reasons: the 37,494 section-years of the 2210 sections absent from the pre-imputation survey file cannot be verified and are therefore excluded by default, and a section-year is counted only if all four indicators were directly observed, whereas Section 3.2 reports rates for each indicator separately. Of the 125,190 section-years in covered sections, 78.3% have all four indicators directly observed. The cross-correlation analysis was likewise repeated using only directly observed lagged pairs across the full 2008–2024 period: all six indicator pairs continued to peak at $k = 0$, with median CCF values at $k = 0$ shifting by no more than 0.029 and, for five of the six pairs, shifting slightly higher rather than lower. The PC1 structure and co-movement pattern are therefore not attributable to the imputation procedure. Finally, the Low-Order functional class group contains only 1276 observations from 76 sections. Although its PC1 solution was stable under treatment adjustment ($\phi = 0.9997$), the cluster bootstrap intervals for its contributions are wide, and results for this group are therefore reported descriptively only.

5. Conclusions and Future Work

5.1. Conclusions

This study developed and applied a multi-indicator analytical framework for network-level pavement condition assessment using longitudinal data from the MDOT Baltimore County network spanning 2008 to 2024. By integrating temporal trend analysis, cross-correlation analysis, and principal component analysis across four condition indicators, namely IRI, rutting depth, structural cracking density, and functional cracking density, the framework characterizes pavement deterioration as a multidimensional process. The results of this analysis support the following conclusions:

1. Pavement deterioration in this network is multidimensional, and single-indicator evaluation cannot adequately represent it. The four indicators are positively correlated, and their first principal component, PC1, accounts for 49.5% of total variance, a share that follows largely from the level of their pairwise correlations; all four indicators contribute substantially to it (20.7% to 29.4%), with the cracking indicators together accounting for 57.5%. The remaining 50.5% of the variance is not captured by this axis, and the moderate pairwise correlations show that the indicators are not interchangeable. This structure persists after accounting for rehabilitation timing: recomputing PC1 on a treatment-adjusted subset reproduces the full-network component almost exactly (Tucker's $\phi = 0.9987$). The practical consequence is visible in the temporal record: after excluding treatment-adjacent observations, median IRI declined from 110.6 to 103.8 in/mi over the study period while functional cracking density increased from 1.96 to 2.23. Because the unadjusted network data understated this cracking increase by an average of 10.8% across the study period, reaching 48.2% in individual years, the treatment-adjusted divergence between improving ride quality and worsening cracking is at least as pronounced as in the unadjusted analysis. A roughness-only monitoring program would therefore have characterized this network as stable or improving while surface cracking continued to spread.
2. The association between indicators is predominantly contemporaneous at the resolution of annual condition surveys. All six indicator pairs reach maximum median

cross-correlation at lag zero in condition levels and after detrending and pre-whitening, as do all pairs except rutting and functional cracking after differencing, and the lag-zero correlation exceeds the correlation at neighboring lags beyond sampling variation. The only consistent lead–lag asymmetry is a small tendency for rutting to precede roughness and cracking by one year, too weak relative to the contemporaneous association to serve as a practical early-warning signal. Sequencing may still occur at sub-annual timescales or at individual sections, but it cannot be detected or exploited with annual network-level data. Year-to-year changes are strongly coupled between the two cracking indicators but only weakly between cracking, roughness, and rutting, so each must be observed directly, and concurrent multi-indicator monitoring is therefore necessary for network-level assessment.

3. Indicator weighting is road-class-dependent. Cracking contributes more strongly to PC1 on Intermediate roads and IRI more strongly on High-Order roads, a difference that is robust under a section-level cluster bootstrap and consistent in direction across scaling and rank-based sensitivity analyses, although its magnitude depends on how extreme values are treated. Cracking accounts for a substantial share of PC1 in every road class, so no class is adequately represented by roughness alone, and a uniform IRI threshold is likely to under-represent cracking-dominated deterioration most strongly on Intermediate roads. These results suggest that condition assessment may benefit from stratification by functional class group.
4. The PC1 composite is an interpretable descriptive measure of multi-indicator condition. All four indicators increase monotonically across PC1 quartiles, confirming that the composite provides an ordered representation of condition rather than a mathematical artifact, and the right-skewed score distribution isolates a subset of sections in which multiple distress types have advanced simultaneously. Contrasting PC1 against RSL_4, a minimum-based remaining-service-life measure built from the same four indicators, shows that the two measures agree on fewer than half (45.2%) of the observations either identifies as high priority, a disagreement that persists at thresholds from the worst 10% to the worst 30%. With the indicator scale held fixed, the aggregation rule alone changes about one third of the high-priority set. RSL_4's unique flags are overwhelmingly driven by a single limiting indicator, almost always roughness or rutting, a pattern that largely reflects the MDOT breakpoint table rather than the minimum rule itself, while PC1's unique flags are overwhelmingly observations with multiple indicators simultaneously elevated, most often led by cracking. Because Q4 sections combine elevated cracking with moderate roughness, the composite indicates the type of deterioration as well as its overall level, which a roughness-triggered or minimum-based measure does not represent directly. Whether prioritization based on PC1 improves maintenance outcomes was not evaluated and is left to future work.
5. The framework is transparent, auditable, and transferable. Its weights are derived from the observed covariance structure of standard network-level survey data rather than assigned through expert judgment, can be recalibrated as new data accumulate, and require no specialized instrumentation. The resulting section-level composite scores can complement existing threshold-based prioritization within current PMS infrastructure.

Beyond these specific conclusions, the study makes a broader conceptual contribution: pavement deterioration should be interpreted as a multidimensional process rather than as a set of independent distress manifestations or a single-indicator decline. The dominant shared axis, the contemporaneous co-movement of indicators, and the class-dependent shift in indicator weights together indicate that the condition indicators co-vary within a common structure whose composition depends on road class, a perspective demonstrated here at

network scale using a multi-decade condition record. This carries implications for how pavement management systems are architected and how monitoring programs are designed across mixed road networks. It also points to a possible role for the framework in emerging digital twin and AI-assisted asset management systems, which require physically grounded, low-dimensional condition representations of the kind constructed here; validating that role under real-time or sensor-driven monitoring is a task for future work, not a result established in this study.

5.2. Future Research Directions

1. Extend the PC1 composite toward sequence-based predictive modeling using LSTM and Transformer architectures for remaining service life estimation across functional class groups.
2. Evaluate the generalizability of the road-class-specific deterioration structure across networks with different pavement compositions, traffic profiles, and climatic environments.
3. Evaluate whether prioritization based on PC1 improves maintenance outcomes relative to single-indicator and minimum-based measures, using future condition, treatment effectiveness, and cost as outcomes.

Overall, this study provides a transparent and transferable framework for network-level multi-indicator condition assessment and a basis for further work on multi-indicator monitoring, maintenance prioritization, and AI-assisted infrastructure management.

Author Contributions: Conceptualization, B.F.and Z.B.; methodology, B.F., M.S. and Z.B.; validation, Z.B., B.F. and M.S.; data curation, B.F., D.O.O. and Z.B.; writing—original draft preparation, B.F.; writing—review and editing, B.F., D.O.O. and Z.B.; supervision, M.S. and Z.B.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the U.S. Department of Transportation's University Transportation Centers Program through the Center for Integrated Asset Management for Multimodal Transportation Infrastructure Systems (CIAMTIS; Grant No. 69A3551847103).

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors express their sincere gratitude to their supervisor and colleagues for their continuous guidance and support throughout this study. The authors also thank the U.S. Department of Transportation's University Transportation Centers Program for funding this research through the Center for Integrated Asset Management for Multimodal Transportation Infrastructure Systems.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IRI	International Roughness Index
PMS	Pavement Management Systems
PCA	Principal Component Analysis
PC1	First Principal Component
RUT	Rutting Depth
SC_DENSITY	Structural Cracking Density
FC_DENSITY	Functional Cracking Density
CCF	Cross-Correlation Function
MDOT	Maryland Department of Transportation
FHWA	Federal Highway Administration
LTPP	Long-Term Pavement Performance
HDM-4	Highway Development and Management Model

AASHTO	American Association of State Highway and Transportation Officials
ESAL	Equivalent Single Axle Load
VAR	Vector Autoregressive
RSL	Remaining Service Life
LSTM	Long Short-Term Memory
FC	Functional Class
IQR	Interquartile Range

References

1. Sarkar, H.; Sinha, S. State-of-the-art review of pavement deterioration models for enhancing transportation infrastructure management. *Int. J. Pavement Eng.* **2026**, *27*, 2633306. [CrossRef]
2. Kulkarni, R.B.; Miller, R.W. Pavement management systems: Past, present, and future. *Transp. Res. Rec.* **2003**, *1853*, 65–71. [CrossRef]
3. Adlinge, S.S.; Gupta, A. Pavement deterioration and its causes. *Int. J. Innov. Res. Dev.* **2013**, *2*, 437–450.
4. Famewo, B.G.; Shokouhian, M. A Review of Pavement Performance Deterioration Modeling: Influencing Factors and Techniques. *Symmetry* **2025**, *17*, 1992. [CrossRef]
5. Alsheyab, M.A.; Khasawneh, M.A.; Abualia, A.; Sawalha, A. A critical review of fatigue cracking in asphalt concrete pavement: A challenge to pavement durability. *Innov. Infrastruct. Solut.* **2024**, *9*, 386. [CrossRef]
6. Meegoda, J.N.; Gao, S. Roughness progression model for asphalt pavements using long-term pavement performance data. *J. Transp. Eng.* **2014**, *140*, 04014037. [CrossRef]
7. Piryonesi, S.M.; El-Diraby, T.E. Examining the relationship between two road performance indicators: Pavement condition index and international roughness index. *Transp. Geotech.* **2021**, *26*, 100441. [CrossRef]
8. Ali, A.; Dhasmana, H.; Hossain, K.; Hussein, A. Modeling pavement performance indices in harsh climate regions. *J. Transp. Eng. Part B Pavements* **2021**, *147*, 04021049. [CrossRef]
9. Zhang, W.; Durango-Cohen, P.L. Explaining heterogeneity in pavement deterioration: Clusterwise linear regression model. *J. Infrastruct. Syst.* **2014**, *20*, 04014005. [CrossRef]
10. Onayev, A.; Swei, O. IRI deterioration model for asphalt concrete pavements: Capturing performance improvements over time. *Constr. Build. Mater.* **2021**, *271*, 121768.
11. Sun, Y.; Gao, Q.; Li, F.; Du, Y. Distress-Level Prediction of Pavement Deterioration with Causal Analysis and Uncertainty Quantification. *Appl. Sci.* **2025**, *15*, 11250. [CrossRef]
12. Nguyen, H.L.; Tran, V.Q. Data-driven approach for investigating and predicting rutting depth of asphalt concrete containing reclaimed asphalt pavement. *Constr. Build. Mater.* **2023**, *377*, 131116. Correction in *Constr. Build. Mater.* **2025**, *493*, 143265. [CrossRef]
13. Shah, Y.U.; Jain, S.; Tiwari, D.; Jain, M. Development of overall pavement condition index for urban road network. *Procedia-Soc. Behav. Sci.* **2013**, *104*, 332–341.
14. Shahnazari, H.; Tutunchian, M.A.; Mashayekhi, M.; Amini, A.A. Application of soft computing for prediction of pavement condition index. *J. Transp. Eng.* **2012**, *138*, 1495–1506. [CrossRef]
15. Qiao, Y.; Chen, S.; Alinizzi, M.; Alamaniotis, M.; Labi, S. Estimating IRI based on pavement distress type, density, and severity: Insights from machine learning techniques. *arXiv* **2021**, arXiv:2110.05413.
16. Gharaibeh, N.G.; Zou, Y.; Saliminejad, S. Assessing the agreement among pavement condition indexes. *J. Transp. Eng.* **2010**, *136*, 765–772. [CrossRef]
17. Múčka, P. Current approaches to quantify the longitudinal road roughness. *Int. J. Pavement Eng.* **2016**, *17*, 659–679. [CrossRef]
18. Albayati, A.H. A review of rutting in asphalt concrete pavement. *Open Eng.* **2023**, *13*, 20220463. [CrossRef]
19. Li, Q.J.; Zhan, Y.; Yang, G.; Wang, K.C. Pavement skid resistance as a function of pavement surface and aggregate texture properties. *Int. J. Pavement Eng.* **2020**, *21*, 1159–1169. [CrossRef]
20. Corley-Lay, J.; Mastin, J.N. Evaluation of long-term pavement performance profile data for flexible pavements. *Transp. Res. Rec.* **2009**, *2093*, 25–30. [CrossRef]
21. Dong, Q.; Huang, B. Evaluation of effectiveness and cost-effectiveness of asphalt pavement rehabilitations utilizing LTPP data. *J. Transp. Eng.* **2012**, *138*, 681–689. [CrossRef]
22. Plati, C.; Gkyrtis, K.; Loizos, A. A practice-based approach to diagnose pavement roughness problems. *Int. J. Civ. Eng.* **2024**, *22*, 453–465. [CrossRef]
23. Basnet, K.S.; Shrestha, J.K.; Shrestha, R.N. Pavement performance model for road maintenance and repair planning: A review of predictive techniques. *Digit. Transp. Saf.* **2023**, *2*, 253–267. [CrossRef]
24. Bandara, N.; Gunaratne, M. Current and future pavement maintenance prioritization based on rapid visual condition evaluation. *J. Transp. Eng.* **2001**, *127*, 116–123. [CrossRef]

25. Peshkin, D.G.; Hoerner, T.E.; Zimmerman, K.A. *Optimal Timing of Pavement Preventive Maintenance Treatment Applications*; NCHRP Report 523; Transportation Research Board: Washington, DC, USA, 2004.
26. Ferreira, A.; Picado-Santos, L.d.; Wu, Z.; Flintsch, G. Selection of pavement performance models for use in the Portuguese PMS. *Int. J. Pavement Eng.* **2011**, *12*, 87–97. [CrossRef]
27. Odoki, J.B.; Kerali, H.G.R. Analytical framework and model descriptions. In *The Highway Development and Management Series*; World Road Association (PIARC): Paris, France, 2000; Volume 4.
28. ARA, Inc., ERES Consultants Division. *Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures*; Final Report, NCHRP Project 1-37A; National Cooperative Highway Research Program, Transportation Research Board, National Research Council: Washington, DC, USA, 2004.
29. Prozzi, J.; Madanat, S. Development of pavement performance models by combining experimental and field data. *J. Infrastruct. Syst.* **2004**, *10*, 9–22. [CrossRef]
30. Madanat, S.; Ben-Akiva, M. Optimal inspection and repair policies for infrastructure facilities. *Transp. Sci.* **1994**, *28*, 55–62. [CrossRef]
31. Frangopol, D.M.; Soliman, M. Life-cycle of structural systems: Recent achievements and future directions. In *Structures and Infrastructure Systems*; Routledge: London, UK, 2019; pp. 46–65.
32. Lytton, R.L.; Uzan, J.; Fernando, E.G.; Roque, R.; Hiltunen, D.; Stoffels, S.M. *Development and Validation of Performance Prediction Models and Specifications for Asphalt Binders and Paving Mixes*; Report SHRP-A-357; Strategic Highway Research Program, National Research Council: Washington, DC, USA, 1993.
33. Huang, Y.H. *Pavement Analysis and Design*, 2nd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2004.
34. Madanat, S.; Mishalani, R.; Ibrahim, W.H.W. Estimation of infrastructure transition probabilities from condition rating data. *J. Infrastruct. Syst.* **1995**, *1*, 120–125. [CrossRef]
35. Madeh Piryonesi, S.; El-Diraby, T.E. Using machine learning to examine impact of type of performance indicator on flexible pavement deterioration modeling. *J. Infrastruct. Syst.* **2021**, *27*, 04021005. [CrossRef]
36. Sollazzo, G.; Fwa, T.; Bosurgi, G. An ANN model to correlate roughness and structural performance in asphalt pavements. *Constr. Build. Mater.* **2017**, *134*, 684–693. [CrossRef]
37. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015.
38. Odolinski, K.; Wheat, P. Dynamics in rail infrastructure provision: Maintenance and renewal costs in Sweden. *Econ. Transp.* **2018**, *14*, 21–30. [CrossRef]
39. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef] [PubMed]
40. Gewers, F.L.; Ferreira, G.R.; de Arruda, H.F.; Silva, F.N.; Comin, C.H.; Amancio, D.R.; Costa, L.d.F. Principal component analysis: A natural approach to data exploration. *ACM Comput. Surv.* **2021**, *54*, 70. [CrossRef]
41. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459.
42. Pandey, B.; Agrawal, M.; Singh, S. Assessment of air pollution around coal mining area: Emphasizing on spatial distributions, seasonal variations and heavy metals, using cluster and principal component analysis. *Atmos. Pollut. Res.* **2014**, *5*, 79–86. [CrossRef]
43. Mavungu, M. Computation of financial risk using principal component analysis. *Algorithmic Financ.* **2023**, *10*, 1–20. [CrossRef]
44. Tibaduiza, D.; Mujica, L.; Rodellar, J. Structural Health Monitoring based on principal component analysis: Damage detection, localization and classification. In *Advances in Dynamics, Control, Monitoring and Applications*; Universitat Politècnica de Catalunya, Departament deMatemàtica Aplicada: Barcelona, Spain, 2011; Volume 3, pp. 8–17.
45. Cha, Y.J.; Ali, R.; Lewis, J.; Büyüköztürk, O. Deep learning-based structural health monitoring. *Autom. Constr.* **2024**, *161*, 105328. [CrossRef]
46. Mondal, T.G.; Chen, G. Artificial intelligence in civil infrastructure health monitoring—Historical perspectives, current trends, and future visions. *Front. Built Environ.* **2022**, *8*, 1007886. [CrossRef]
47. Xu, W.; Lin, W. Selecting the public transit projects with PCA-DP technique: The example of Xiamen City. *Transp. Policy* **2016**, *46*, 56–71. [CrossRef]
48. Maryland Department of Transportation, State Highway Administration. *Pavement & Geotechnical Design Guide*; Technical report; Updated 27 June 2022; Maryland Department of Transportation, State Highway Administration, Office of Materials Technology, Pavement and Geotechnical Division: Hanover, MD, USA, 2022.
49. Henderson, B.; Dickes, J.; Cimini, G.; Olmedo, C. *FHWA LTPP Guidelines for Measuring Bridge Approach Transitions Using Inertial Profilers*; Federal Highway Administration: Washington, DC, USA, 2016.
50. Maryland Department of Transportation State Highway Administration, Office of Planning and Preliminary Engineering. MDOT SHA Roadway Functional Classification. ArcGIS Online for Maryland. 2024. Available online: https://services.arcgis.com/njFNhDsUCentVYJW/arcgis/rest/services/MDOT_SHA_Roadway_Functional_Classification/FeatureServer/0 (accessed on 15 October 2026).

51. Federal Highway Administration. *Highway Functional Classification Concepts, Criteria and Procedures*; Technical Report FHWA-PL-13-026; U.S. Department of Transportation, Federal Highway Administration: Washington, DC, USA, 2013.
52. Timmerman, M.E. Multilevel component analysis. *Br. J. Math. Stat. Psychol.* **2006**, *59*, 301–320. [CrossRef] [PubMed]
53. Labi, S.; Sinha, K.C. *The Effectiveness of Maintenance and Its Impact on Capital Expenditures*; Publication FHWA/IN/JTRP-2002/27; Joint Transportation Research Program, Indiana Department of Transportation and Purdue University: West Lafayette, IN, USA, 2003. [CrossRef]
54. Sharaf, E.A.; Shahin, M.Y.; Sinha, K.C. Analysis of the effect of deferring pavement maintenance. *Transp. Res. Rec.* **1988**, *1205*, 29–35.
55. Bhandari, S.; Luo, X.; Wang, F. Understanding the effects of structural factors and traffic loading on flexible pavement performance. *Int. J. Transp. Sci. Technol.* **2023**, *12*, 258–272. [CrossRef]
56. Adeli, S.; Najafi moghaddam Gilani, V.; Kashani Novin, M.; Motesharei, E.; Salehfard, R. Development of a relationship between pavement condition index and international roughness index in rural road network. *Adv. Civ. Eng.* **2021**, *2021*, 6635820. [CrossRef]
57. Jiang, J.; Xu, K.; Song, Y.; Gao, L.; Zhang, J. Investigation of cracking behavior in asphalt pavement using digital image processing technology. *Front. Built Environ.* **2025**, *11*, 1580379. [CrossRef]
58. Pérez-Acebo, H.; Isasa, M.; Gurrutxaga, I.; García, H.; Insausti, A. International Roughness Index (IRI) prediction models for freeways. *Transp. Res. Procedia* **2023**, *71*, 292–299. [CrossRef]
59. Chen, D.; Hildreth, J.; Mastin, N. Determination of IRI limits and thresholds for flexible pavements. *J. Transp. Eng. Part B Pavements* **2019**, *145*, 04019013. [CrossRef]
60. Lemonakis, P.; Kopelias, P.; Karlaftis, P. Introduction of a new index to assess pavement functional condition. *Environ. Dev. Sustain.* **2025**, *27*, 3265–3282. [CrossRef]
61. American Association of State Highway and Transportation Officials. *Mechanistic-Empirical Pavement Design Guide: A Manual of Practice*, interim ed.; American Association of State Highway and Transportation Officials: Washington, DC, USA, 2008.
62. Islam, K.M.; Gassman, S.L. Predicting flexible pavement distress and IRI considering subgrade resilient modulus of fine-grained soils using MEPDG. *Materials* **2023**, *16*, 1126. [CrossRef] [PubMed]
63. Al-Omari, B.; Darter, M.I. Effect of pavement deterioration types on IRI and rehabilitation. *Transp. Res. Rec.* **1995**, *1505*, 57–65.
64. Saba, C.S.; Ngepah, N.; Odhiambo, N.M. Analysis of convergence in transport infrastructure: A global evidence. *Eur. J. Transp. Infrastruct. Res.* **2021**, *21*, 137–160. [CrossRef]
65. Horak, E.; De Beer, M.; van Zyl, G.; Maina, J.; Fakra, D.A.H. Research and Measuring Technology Needs to Better Model and Measure Fatigue Crack Development of Thinly Surfaced Asphalt Road Pavements. *CivilEng* **2022**, *3*, 456–467. [CrossRef]
66. Henning, T.F.; Alabaster, D.; Arnold, G.; Liu, W. Relationship between traffic loading and environmental factors and low-volume road deterioration. *Transp. Res. Rec.* **2014**, *2433*, 100–107. [CrossRef]
67. Shtayat, A.; Moridpour, S.; Best, B.; Rumi, S. An overview of pavement degradation prediction models. *J. Adv. Transp.* **2022**, *2022*, 7783588. [CrossRef]
68. Yan, Y.; Ni, L.; Sun, L.; Wang, Y.; Zhou, J. Digital twin enabling technologies for advancing road engineering and lifecycle applications. *Engineering* **2025**, *44*, 184–206. [CrossRef]
69. Consilvio, A.; Hernández, J.S.; Chen, W.; Brilakis, I.; Bartoccini, L.; Gennaro, F.D.; van Welie, M. Towards a digital twin-based intelligent decision support for road maintenance. *Transp. Res. Procedia* **2023**, *69*, 791–798. [CrossRef]
70. Tong, S.; Wu, D.; Liu, X.; Zheng, L.; Du, Y.; Zou, D. STGAN: Spatial-Temporal Graph Autoregression Network for Pavement Distress Deterioration Prediction. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 10764–10779. [CrossRef]
71. Pan, N.; Du, Y.; Parlikad, A.K. A spatiotemporal prediction method for the evolution of pavement distress in road networks. *Comput.-Aided Civ. Infrastruct. Eng.* **2025**, *40*, 6513–6532. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.