Planning Nondestructive Culvert Network Condition Inspections: A Dual-Risk Assessment Approach

Mearg Ngusse Sahle[1*] and Zachery Taylor Howell[2]

[1*]Department of Global Architecture, Osaka University, Suita, Osaka, Japan.

[2] Independent Researcher, Tampa, Florida, USA.

*Corresponding author(s). E-mail(s): u645342h@ecs.osaka-u.ac.jp;

Abstract

Culvert inspection programs often rely on uniform, calendar-based visual inspections informed by culvert condition or age. However, visual methods are subjective, incapable of detecting internal defects, and inefficient as they fail to account for variations in hazard exposure and deterioration rates. Condition-based prioritization typically supports reactive maintenance by identifying damage only after deterioration has occurred, whereas hazard-exposure-based monitoring enables proactive or prescriptive maintenance by addressing failure mechanisms in advance. Nondestructive testing (NDT) offers a more diagnostic alternative but is constrained by cost and logistical complexity. Full-network deployment of NDT is neither feasible nor necessary; targeted planning is essential for efficiency. This study presents a dual-risk reinforced concrete culvert NDT deployment framework that integrates independently estimated hydraulic inadequacy and structural weakness serviceability hazard rates into a joint-risk metric. Hazard rates for 2,190 culverts in Ethiopia were derived from $\mathbf{\Gamma}$-frailty Cox proportional hazards models. Kendall's $\boldsymbol{\tau \approx 0.043 (p = 0.0102)}$ indicated negligible dependence between hazards, supporting aggregation via a weighted geometric mean and retaining a copula-based fallback for stronger dependence. Culverts were classified into four joint-risk quadrants using 66th percentile cutoffs, and composite scores were mapped to five NDT intervals: annual, biennial, triennial, quinquennial, and decennial. Optimized through brute-force grid search and benchmarked against a uniform three-year cycle over 30 years, the strategy reduced inspection labor hours by 35% while capturing 80% of cumulative joint risk within the top 53% of culverts. This framework offers transportation agencies a scalable, data-driven approach to prioritize inspections, enhance safety, and reduce costs.

Keywords: Reinforced concrete culvert, non-destructive testing, inspection planning, serviceability hazard, joint risk metric.

1 Introduction

Failure, defined as a structure's inability to perform its intended function [1], stems from two primary serviceability hazards in culverts: hydraulic inadequacy (scour, sedimentation, flooding) and structural weakness (mechanical or chemical degradation). Both hazards are mitigated through periodic inspections. During such inspections, the hydraulic and structural adequacy of culverts is assessed using either visual inspections or Non-Destructive Tests (NDTs).

Visual inspections are the most commonly employed method, with culvert condition qualitatively rated based on the engineering judgment of experienced inspectors. These assessments are often supported by optical tools such as field glasses, drones, cameras, and digital binoculars. Visual inspections allow for a rapid, relative evaluation of deterioration by identifying visible defects. However, despite their acceptable level of sensitivity, visual inspections are subject to possible misrepresentation of culvert conditions due to human bias, structural complexity, or limited accessibility. For instance, a 2-year study by Moore [2], involving 49 bridge inspectors from 25 U.S. state agencies, found that 68% of the condition ratings varied within one rating point of the average, and 95% varied within two points, highlighting the variability inherent in visual assessments. Additionally, condition-rating ambiguities may arise from the subjective aggregation of sub-component ratings.

Conversely, NDTs enable the quantitative characterization of surface and subsurface concrete defects, corrosion activity, and sedimentation or soil cavitation in reinforced concrete culverts. Rebound hammer and pendulum tests are commonly used to assess the relative strength and surface hardness of concrete, which can indicate the likelihood of delamination. Radiographic testing, magnetic particle inspection, and pulse-echo methods, such as radar, impact-echo, and ultrasonic impulse-echo, are effective for detecting cracks and internal cavities in reinforced concrete structures. To evaluate rebar corrosion activity, techniques such as Ground Penetrating Radar (GPR), magnetic flux leakage, half-cell potential measurements, and chloride ion penetration tests are frequently employed. Additional analyses of compressive strength, density, and petrographic characteristics can be performed in laboratory settings using cored samples. Other techniques, including X-ray imaging, infrared thermography, and low-frequency ultrasonic testing, are useful for identifying sedimentation or cavitation within fill materials.

Unlike visual inspection, nondestructive condition assessments offer significantly greater accuracy, providing reliable data to support maintenance decisions that enable timely retrofitting or preventive maintenance of defects, potentially avoiding critical failures and extensive future repairs. However, network-level culvert NDT inspection is an inherently expensive and labor-intensive process, requiring substantial skilled man-hours for fieldwork, testing, data management, and decision-making. Field NDT inspections typically demand specialized equipment such as bucket cranes, scaffolding

platforms, ladders, and mobility vehicles to ensure safe and effective access to culvert sites. Additional resources, including safety gear, testing instruments, and digital or analog data collection devices, are also necessary. Consequently, supplementary data from visual inspections or environmental monitoring should be strategically used to guide the timing and frequency of NDT deployment.

In terms of operationalizing NDT strategies within culvert inspection programs, manuals and field studies have highlighted both their diagnostic value and practical deployment limitations. The National Cooperative Highway Research Program Culvert and Storm Drain System Inspection Manual recommends a tiered approach, beginning with visual assessments and escalating to NDT methods such as acoustic sounding, GPR, and half-cell potential testing when specific risk triggers are identified [3]. Similarly, the Minnesota Department of Transportation (DOT) Enhanced Culvert Inspection Guidebook encourages the use of robotic CCTV, pulse-echo methods, and sonar technologies, but notes the challenges of equipment costs, access limitations, and technician availability [4]. Purdue University's large-culvert guidelines emphasize risk-based prioritization over fixed schedules, although they do not prescribe specific NDT modalities [5]. Other inspection standards, such as the Government of the Northwest Territories and Ontario structure inspection manuals, advocate adjusting inspection frequencies based on condition ratings or deterioration likelihood [6, 7]. These findings are echoed in the broader infrastructure literature, which consistently underscores that NDT methods are most effective when applied selectively, typically during special event, post-event assessments or for high-risk assets, rather than as part of routine inspection schedules.

Several studies have proposed formalized logic to guide NDT scheduling. Yang and Allouche [8] mapped NDT modalities, such as GPR, ultrasonic testing, and acoustic emission, to defect types across common culvert materials, providing a reference for modality selection. Kalhor et al [9] applied infrared thermography to detect early-stage cavity formation near buried culverts. Schroeder et al [10] developed a non-contact ultrasonic system capable of detecting wall thinning in corroded steel culverts. These studies focus primarily on tool performance and diagnostic application, but do not provide scheduling strategies.

More recent research has explored predictive frameworks for inspection scheduling. Mohammadi et al [11] introduced a hybrid system combining Extreme Gradient Boosting (XGBoost) with Grey Wolf Optimizer (GWO) to prioritize culvert inspections based on predicted deterioration. Applied to a Utah DOT dataset of 272 culverts, the model reduced inspection costs by over 15% and flagged less than 1.5% of culverts for immediate inspection. Similarly, decision-tree models developed using 12,400 culvert records from the Ohio DOT achieved over 75% classification accuracy on the test set and enabled a 44% reduction in inspection workload by predicting which culverts were likely to be in poor condition [12]. A South Carolina DOT study used the Analytic Hierarchy Process (AHP) to develop priority indices based on hydraulic vulnerability and structural condition for a network of 5,200 culverts [13].

Although these models significantly enhance visual inspection planning, they rarely include logic for when and how to deploy NDT methods. For example, while risk-based inspection is widely used in industrial applications to align NDT deployment

with failure probability and consequence [14], its integration into culvert management frameworks remains limited. Studies on bridges and pipelines using fuzzy logic, probabilistic modeling, or partially observable Markov decision processes further show that formal scheduling of NDT remains an emerging research area [15]. This gap underscores the value of the current study, which seeks to explicitly connect deterioration risk with optimal, targeted NDT deployment across a culvert network.

1.1 Problem Statement

Despite the development of advanced deterioration modeling techniques, inspection practices in most culvert management systems remain disconnected from hazard forecasting. Culverts are typically inspected at fixed intervals using visual methods, with frequency adjustments often based on observed condition, but generally irrespective of their hazard exposure, deterioration rate, or vulnerability classification. This uniform approach is often inefficient, as it may fail to detect early-stage deterioration in high-hazard locations while allocating resources to culverts exposed to lower levels of hazard. Besides, condition-based inspection prioritization often supports reactive maintenance strategies by identifying damage after deterioration has occurred, whereas hazard-exposure-based monitoring enables proactive or even prescriptive maintenance by forecasting and addressing failure mechanisms in advance.

Moreover, conventional life-expectancy-based models often fall short in accurately prioritizing culvert inspections and maintenance. An aging culvert is not necessarily a deteriorated one, as these structures are typically engineered to withstand specific environmental conditions throughout their intended service life. Relying solely on chronological age as a proxy for risk may misrepresent actual vulnerability, particularly when a culvert has operated under relatively benign conditions. Conversely, newer culverts exposed to high hazard conditions, such as elevated hydraulic loads, corrosive soils, or heavy vehicle traffic, can deteriorate much more rapidly than anticipated. This discrepancy highlights the limitations of age-based approaches and reinforces the need for hazard-exposure-informed inspection frameworks that better capture real-world deterioration dynamics.

In Ethiopia, the context of this study, and in other similar settings, culvert maintenance strategies continue to rely on periodic visual inspections. While these methods are simple and low-cost, they are inherently subjective and often ineffective at detecting internal or subsurface deterioration. This diagnostic limitation reduces the likelihood of identifying culverts exposed to elevated hazard in time to prevent serviceability failure.

NDT offers a more objective and comprehensive alternative for culvert diagnostics. However, its widespread use is constrained by cost, feasibility, and logistical complexity. Full-network deployment of NDT is neither feasible nor necessary. With proper planning, NDT inspections can be selectively scheduled based on hazard indicators such as elevated failure probability or signs of accelerated deterioration. When targeted effectively, NDT improves diagnostic accuracy and may reduce overall inspection costs compared to cyclic visual inspections. This highlights the need for inspection programs that prioritize structures exposed to high hazard, guided by data-informed strategies rather than rigid calendar schedules.

Part I of this research introduced a dual-hazard survival-analysis framework to estimate two relative serviceability hazard rates, namely hydraulic inadequacy and structural weakness, for a dataset of 2,190 reinforced concrete culverts in Ethiopia. The marginal hazard rates were modeled using Γ-frailty Cox proportional hazards models, with environmental variables as covariates. Specifically, hydraulic hazard rates $h_i^{(1)}$ were estimated as functions of discharge depth, barrel slope, and barrel length. Structural hazard rates $h_i^{(2)}$ were analogously modeled using average daily truck traffic and soil pH, with soil electrical conductivity excluded following diagnostic evaluation.

Part I demonstrated that culvert serviceability hazard can be effectively decomposed into two independent failure mechanisms, thus providing a detailed, fine-grained, network-wide view of vulnerability. However, while these marginal serviceability hazard estimates, which reflect the rate of deterioration under each distinct failure mechanism, effectively distinguish structures experiencing faster deterioration due to higher hazard rates from those with slower deterioration associated with lower hazard rates, they do not directly inform when and where to deploy costly NDT resources. The use of marginal serviceability hazard estimates, rather than an aggregated overall deterioration rate, enables a more precise characterization of failure mechanisms that compromise a culvert's ability to perform its intended drainage or structural function. This targeted insight supports the early identification of serviceability deterioration rates that precede functional failure. Nonetheless, translating these hazard metrics into actionable inspection strategies remains essential for optimizing resource allocation.

1.2 Objectives

This study develops a flexible, data-adaptive dual-risk inspection planning framework that replaces rigid calendar-based inspection cycles with schedules informed by quantified serviceability risk. By integrating marginal hazard estimates for hydraulic inadequacy and structural weakness with their joint risk, the proposed methodology offers a comprehensive risk metric for each culvert, enabling the targeted deployment of resource-intensive NDTs. To translate these hazard estimates into actionable inspection plans, the following objectives will be pursued. First, each culvert is classified into one of four binary joint-risk quadrants by applying tertile thresholds to estimated hydraulic and structural hazard rates. Second, the framework quantifies the statistical dependence between the two hazard types using Kendall's τ and selects an appropriate joint-risk model: a copula-based density score if dependence is moderate to strong, or a weighted geometric mean if dependence is weak or negligible. Third, the resulting joint-risk scores are translated into actionable inspection schedules by ranking culverts and assigning inspection intervals that balance risk coverage with resource constraints. Together, these objectives define a transparent and reproducible methodology for prioritizing condition assessments across culvert networks. The framework ensures efficient allocation of NDT resources by directing timely inspections toward structures with high hazard exposure, while minimizing unnecessary evaluations of those with low hazard exposure.

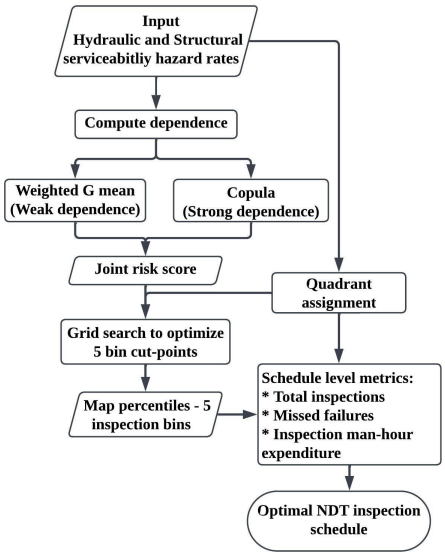

Fig. 1 Dual-risk NDT inspection scheduling framework

2 Methodology

Building on the previously estimated relative hazard rates for hydraulic inadequacy and structural weakness, this section presents a methodology for translating these marginal hazard metrics into a prioritized schedule of NDT inspections. In this framework, risk is directly equated to the deterioration rate, as the consequence of a missed inspection is assumed to be equal for all culverts. Two joint-risk metrics are proposed to accommodate different dependence structures between hydraulic and structural hazards: (i) a copula-based density score, suitable for capturing nonlinear and tail-dependent co-behavior when the hazards exhibit significant concordance; and (ii) a transparent, distribution-free geometric mean index, which is robust when the hazards are weakly correlated or independent.

The resulting joint-risk scores are used to generate actionable inspection schedules through a three-stage workflow. First, each culvert is classified into one of four binary joint-risk quadrants, $[0,0]$, $[0,1]$, $[1,0]$ or $[1,1]$, based on threshold marginal hazard rates. Second, within each quadrant, culverts are ranked according to their joint-risk scores. Third, the percentile-based rankings are mapped to discrete inspection intervals that balance risk coverage with available inspection resources. Figure 1 summarizes the full framework, from quadrant classification through joint-risk modeling to the final generation of risk-informed inspection schedule.

2.1 Dual-Risk Classification

The two series of unitless, frailty-adjusted relative hydraulic and structural serviceability hazard rates $\left\{(h_i^{(1)},\ h_i^{(2)})\right\}_{i=1}^{N}$ were previously estimated for a network of N

culverts. Each culvert i ($i = 1, \cdots, N$) is first classified into one of four binary 'low (0) /high (1)' joint-risk quadrants, $[0,0]$, $[0,1]$, $[1,0]$, and $[1,1]$, by applying tertile thresholds to the relative hydraulic $h_i^{(1)}$ and structural $h_i^{(2)}$ serviceability hazard rates. This classification facilitates the identification of assets with combined vulnerability and thereby prioritizes inspections for culverts at elevated risk under one or both deterioration modes. The use of tertile cut-offs aligns with the Ethiopian Roads Administration's three-tier condition grading (poor, fair, good), ensuring consistency with national inspection guidelines [17]. Specifically, the hazard rates were divided at the 33rd and 66th percentiles: culverts with values below the 66th percentile are coded as 0 (lower risk), and those at or above the 66th percentile are flagged as 1 (higher risk, corresponding to poor condition) Eq. (1). Thus, each culvert i is assigned a two-bit code $(b_i^{(1)},\ b_i^{(2)}) \in \{0,1\}^2$.

$$b_i^{(1)} = \begin{cases} 0, & h_i^{(1)} < P_{66}(h^{(1)}) \\ 1, & h_i^{(1)} \geq P_{66}(h^{(1)}) \end{cases} \qquad b_i^{(2)} = \begin{cases} 0, & h_i^{(2)} < P_{66}(h^{(2)}) \\ 1, & h_i^{(2)} \geq P_{66}(h^{(2)}) \end{cases} \tag{1}$$

2.2 Assessing Dependence Between Serviceability Hazards

To decide whether an explicit dependence model is warranted, the empirical concordance between $h_i^{(1)}$ and $h_i^{(2)}$ is first quantified using Kendall's rank correlation coefficient (τ). Each marginal series is transformed to the unit interval via its empirical cumulative distribution function (ECDF) so that pseudo-observations $(u_i, v_i) \in (0,1)^2$ preserve only the ordinal ranks of each hazard, isolating the dependence structure from scale differences Eq. (2). Kendall's τ is then computed empirically (via observed ranks) and by evaluating each unordered pair of distinct culverts i, j (with $i < j$) and checking whether their hazard-rank differences in both modes move in the same direction (concordant) or in opposite directions (discordant). A pair is concordant if $(u_i - u_j)(v_i - v_j) > 0$ and discordant if $(u_i - u_j)(v_i - v_j) < 0$. Summing these counts across the $\frac{1}{2}N(N-1)$ comparisons yields the numbers of concordant pairs C and discordant pairs D from which an approximate two-sided p-value for the null hypothesis $\tau = 0$ is obtained via the large sample normal approximation in Equation (4), where Φ denotes the standard normal CDF. τ varies from -1 (perfect negative association) through 0 (independence) to $+1$ (perfect positive association) Eq. (3).

$$u_i = \frac{\text{rank}(h_i^{(1)})}{N+1}, \quad v_i = \frac{\text{rank}(h_i^{(2)})}{N+1} \tag{2}$$

$$\tau = \frac{C - D}{\frac{1}{2}N(N-1)} \tag{3}$$

$$z = \frac{3\tau\sqrt{N(N-1)}}{\sqrt{2(2N+5)}}, \quad p \approx 2\left\{1 - \Phi(|z|)\right\} \tag{4}$$

Once τ is established, the framework branches according to the strength of dependence. If $|\tau| \gtrsim 0.15$, indicating moderate to strong concordance, a fully parametric

copula model is fitted to capture nonlinear and tail-specific co-behavior. If $|\tau| \lesssim 0.15$, reflecting weak or negligible association, the hazards are aggregated into a single, geometrically weighted mean index. This index preserves marginal ranking while avoiding the over-fitting risk of an unnecessary copula. Conventional effect-size benchmarks [16] classify τ values near 0.10 as small and those near 0.30 as medium, so a cutoff around 0.15 ensures that copula complexity is invoked only when it delivers nontrivial gains in both fit and risk-prioritization performance, avoiding overfitting under weak dependence.

2.3 Joint Risk Metrics

2.3.1 Copula for Moderate-to-Strong Concordance

When the empirical Kendall's τ exceeds an effect size threshold ($|\tau| \gtrsim 0.15$), a full bivariate copula model is fitted to capture any non-linear or tail-specific dependence between the hydraulic and structural hazard rates. Four copula families were evaluated for their ability to capture different patterns of dependence between the hydraulic and structural serviceability hazard rates: the Clayton, Frank, and Gumbel Archimedean copulas, as well as the Gaussian elliptical copula Table 1. These copulas were selected based on their suitability for modeling dependence structures relevant to culvert performance, environmental factors, and failure risks. Each copula's dependence parameter θ (or correlation parameter ρ, for the Gaussian copula) was estimated by maximizing the copula log-likelihood function over the transformed pseudo-observations (u_i, v_i).

Table 1 Candidate copula families for joint modeling

Family	**Type**	**Functional form, Parameter range**	**Dependence**
Clayton	Archimedean	$C(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0$	Asymmetric (lower tail)
Frank	Archimedean	$C(u,v) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right), \quad \theta \in \mathbb{R}\setminus\{0\}$	Symmetric
Gumbel	Archimedean	$C(u,v) = \exp\left\{-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{1/\theta}\right\}, \quad \theta \geq 1$	Asymmetric (upper tail)
Gaussian	Elliptical	$C(u,v) = \Phi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v)), \quad \rho \in (-1,1)$	Symmetric

Model fit was assessed using three criteria. First, Akaike's Information Criteria was computed as $AIC = -2\ell_{\max} + 2\kappa$, where $\ell_{\max}$ is the maximized log-likelihood and κ is the number of parameters. Lower AIC values indicate a better balance of goodness of fit and model complexity. Second, the CvM statistic was used to quantify the squared-integral distance, W^2, between the empirical copula $C_n(u,v)$ and the fitted copula $C_\theta(u,v)$ Eq. (5). Smaller values of W^2 indicate that the fitted copula more closely reproduces the empirical joint rank structure, providing a distribution-free goodness of fit metric that is particularly sensitive to deviations throughout the entire dependence surface, not just in the extremes. Third, Kendall's τ was used to assess the overall

concordance between the two hazard dimensions Eq. (3). For each bivariate copula family, Kendall's τ has a known analytical form in terms of the copula parameter θ. For example, $\tau = \frac{\theta}{\theta+2}$ for the Clayton copula, $\tau = 1 - \frac{1}{\theta}$ for the Gumbel copula, $\tau = \frac{2}{\pi}\arcsin(\rho)$ for the Gaussian copula (where ρ is the correlation parameter). For the Frank copula, τ is computed numerically via the Debye function of order 1, $D_1(\theta)$, where t is the dummy integration variable Eq. (6). Estimating τ analytically (via θ) allowed quantification of the overall monotonic dependence, with $\tau \in [-1, 1]$ reflecting perfect negative to perfect positive association. Together, these measures, AIC, CvM distance, and Kendall's τ, offered a robust basis for selecting the best-fitting copula model to characterize joint deterioration risk across the culvert network.

$$W^2 = \int_0^1 \int_0^1 \left[C_n(u, v) - C_\theta(u, v)\right]^2 \, du \, dv \tag{5}$$

$$D_1(\theta) = \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} \, dt \tag{6}$$

Marginal densities $p_1(x_1)$ and $p_2(x_2)$ of the raw hazard-rate series were estimated by Gaussian Kernel Density Estimation (KDE) Eq. (7), where h is the bandwidth selected by Silverman's rule of thumb Eq. (8), with σ the sample standard deviation and IQR the interquartile range. IQR is a robust measure of statistical dispersion defined as the difference between the 75th percentile (the third quartile, Q_3) and the 25th percentile (the first quartile, Q_1) of a distribution: $IQR = Q_3 - Q_1$. Because it focuses solely on the central 50% of the data, the IQR is much less sensitive to extreme values (outliers) than the total range. In Silverman's rule of thumb for bandwidth selection, the factor $\frac{IQR}{1.34}$ provides an estimate of the standard deviation under the assumption of approximately Gaussian data but remains reliable if the tails of the distribution are heavy or contaminated by outliers. This rule balances bias and variance under the assumption of approximately Gaussian margins.

$$p(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right), \quad K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \tag{7}$$

$$h = 0.9 \min\left(\sigma, \frac{IQR}{1.34}\right) N^{-\frac{1}{5}} \tag{8}$$

The copula model exhibiting the lowest AIC and smallest CvM distance, while also yielding a statistically significant Kendall's τ, would be selected to characterize the joint hazard distribution. To visualize regions of compounded hydraulic and structural serviceability failure risk, contour plots and three-dimensional surfaces of the fitted copula density $C(u, v)$ can be generated, along with plots of the joint hazard density in the physical space. The joint density in the physical hazard space can be computed using Kernel Density Estimates (KDEs) of the marginal hazard distributions in conjunction with the fitted copula. This formulation, shown in Eq. (9), provides a comprehensive representation of dual serviceability failure risk. In this expression, $f(h_i^{(1)}, h_i^{(2)})$ denotes the joint probability density function of the hydraulic and structural serviceability hazard rates, $p_1(h_i^{(1)})$ and $p_2(h_i^{(2)})$ are the respective marginal density estimates, u_i and v_i are their corresponding cumulative distribution

functions, and $c(\cdot)$ is the copula density function that captures their dependence. A higher joint density indicates greater probability that hydraulic and structural serviceability failures co-occur, warranting higher inspection priority. These outputs enable a risk-driven inspection prioritization strategy by explicitly accounting for the co-deterioration patterns, thereby facilitating more effective and efficient NDT scheduling across the culvert network.

$$f\big(h_i^{(1)}, h_i^{(2)}\big) = c(u_i, v_i)\, p_1\big(h_i^{(1)}\big)\, p_2\big(h_i^{(2)}\big) \tag{9}$$

2.3.2 Weighted Geometric Mean for Weak or Negligible Concordance

If ($|\tau| \lesssim 0.15$), the two hazards may be regarded as effectively independent and combined into a single index. The weighted geometric mean fuses two positive risk measures, $h^{(1)}$ and $h^{(2)}$ in this case, into a single composite index by taking each hazard to a power that reflects its relative importance. Formally, the index is defined as $\mathscr{G}\big(h_i^{(1)}, h_i^{(2)}\big) = h^{(1)\alpha} h^{(2)(1-\alpha)}$ where the tunable weight $\alpha \in [0, 1]$ governs the emphasis placed on the hydraulic risk versus the structural risk. Because $\ln\big(\mathscr{G}(h_i^{(1)}, h_i^{(2)})\big) = \alpha \ln h^{(1)} + (1-\alpha) \ln h^{(2)}$ the geometric mean behaves additively on the logarithmic scale, smoothly interpolating between the two hazards. When $\alpha = 1$, the index reduces to $h^{(1)}$, and when $\alpha = 0$, it reduces to $h^{(2)}$.

A key advantage of this approach is that it preserves the rank ordering of culverts. If one culvert has strictly higher hydraulic or structural hazard than another, its composite score $\mathscr{G}$ will also be higher. This monotonicity ensures consistency when mapping $\mathscr{G}$ into percentile-based inspection schedules. Moreover, unlike an arithmetic mean, which can be dominated by an extreme value in one dimension, the geometric mean tempers extremes. A very low hazard in one mode will pull the composite score downward. This feature reflects the practical notion that a culvert with negligible risk in one mode requires relatively less urgent inspection even if the other hazard is elevated.

The weighted geometric-mean index is particularly well suited to situations of weak statistical dependence. In such regimes, fitting a full bivariate copula model introduces additional parameters that the data cannot reliably support, creating risk of overfitting. In contrast, the geometric-mean index depends only on the marginal ranks and a single parameter α, making it a robust, distribution-free alternative that still captures the essential relative severity of the two hazards (i.e., it avoids overfitting when the two hazards are essentially independent).

Finally, the parameter α offers a transparent mechanism for accommodating stakeholder priorities. An agency most concerned about catastrophic hydraulic failures might choose $\alpha > 0.5$, thereby elevating the composite score whenever hydraulic risk is high. This ensures that culverts with elevated hydraulic hazard appear earlier in the inspection ranking. In this way, the weighted geometric mean provides both interpretability and flexibility for generating a unified risk score under conditions of weak or negligible hazard dependence.

2.4 NDT Scheduling Algorithm

The continuous joint-risk scores produced by either the copula density $f(h_i^{(1)}, h_i^{(2)})$ or weighted geometric mean $\mathscr{G}(h_i^{(1)}, h_i^{(2)})$ are converted into discrete inspection intervals in order to balance safety with resource constraints. To accomplish this, each culvert's joint-risk score is first converted into a percentile rank $r_i \in [0, 1]$, which is then mapped to an inspection interval Δt_i in years using a set of flexible percentile cut-points ($p_1 > p_2 > p_3 > p_4$) Eq. (10). A brute-force grid search identifies the optimal combination of percentile thresholds that best aligns inspection frequency with available resources.

Each threshold quartet defines a complete five-level inspection regime, where culverts in the topmost percentile band ($\geq p_1$) receive annual inspections; and those in the bands (p_2, p_1), (p_3, p_2), (p_4, p_3), and $< p_4$ are scheduled biennially, triennially, quinquennially, and decennially, respectively. This ranking-based assignment guarantees that the highest-risk culverts receive the most frequent inspections.

$$\Delta t_i = \begin{cases} 1 & r_i \geq p_1 \\ 2 & p_2 \leq r_i < p_1 \\ 3 & p_3 \leq r_i < p_2 \\ 5 & p_4 \leq r_i < p_3 \\ 10 & r_i < p_4 \end{cases} \tag{10}$$

Because the four percentile thresholds live on a continuous scale, there are infinitely many possible (p_1, p_2, p_3, p_4) combinations, making it infeasible to evaluate every real-valued quartet. Instead, the search was discretized into a lattice of 25 finite lattice of candidate percentiles ranging from 0.30 to 1.0, yielding $\binom{25}{4} = 12,650$ strictly descending quadruplets. Selecting percentile cut-points between 0.30 and 1.0 focuses the grid search on the portion of the network where inspection frequency actually changes, the elevated risk region, while keeping the total number of combinations manageable. This approach provides fine discrimination among the most critical culverts without wasting resources on low-risk cases. Within this chosen resolution, the brute-force grid search exhaustively evaluates every possibility, balancing computational tractability with sufficient granularity to capture a near-optimal solution. If greater precision is required, the grid can be refined at the cost of evaluating additional combinations.

The grid-search is not an unconstrained brute-force procedure; it applies a two-step performance screen to benchmark each candidate schedule against a uniform Δ_{uni} inspection plan (with c_i labor hours per inspection over a Ψ-year evaluation horizon), ensuring that each regime satisfies both resource and reliability requirements.

The first screen evaluates labor-hour demand. For each culvert i, the number of inspections over the evaluation horizon Ψ is computed as $n_i = \left\lfloor \frac{\Psi}{\Delta t_i} \right\rfloor$, and multiplied by the per-inspection effort c_i. Summing across all N-culverts yields the total network-wide labor demand Eq. (11). This quantity is compared to a uniform inspection plan, under which each culvert undergoes $\left\lfloor \frac{\Psi}{\Delta_{\text{uni}}} \right\rfloor$ inspections and the network-wide labor demand is $H_{\text{uni}} = (N n_{\text{uni}} \cdot c_i)$. Any candidate schedule whose H exceeds H_{uni} is discarded. Schedules that pass the labor-hour screen proceed to a reliability evaluation. The probability that culvert i experiences an undetected failure between inspections

is computed as $P_{i,\text{miss}}$ (Eq. (12)), where h_i is the instantaneous hazard rate for culvert i, and Δt_i is the inspection interval that is, the length of time culvert i goes uninspected. The expected missed failures across the network are then calculated as M_{risk} , Eq. (13). By exhaustively evaluating every strictly descending set of four percentile thresholds and retaining the schedule that minimizes M_{risk} subject to the labor-hour constraint, the algorithm guarantees a globally optimal solution at the chosen resolution. The optimal threshold set is the one that minimizes expected missed failures without exceeding the labor-hour budget of a uniform baseline plan.

$$H = \sum_{i=1}^{N} n_i \cdot c_i. \tag{11}$$

$$P_{i,\text{miss}}(T \leq \Delta t_i) = 1 - \exp(-h_i \cdot \Delta t_i) \tag{12}$$

$$M_{\text{risk}} = \sum_{i=1}^{N} n_i \cdot P_{i,\text{miss}} \tag{13}$$

These metrics collectively enable an effective comparison of inspection efficiency (effort) and effectiveness (failure coverage) between the conventional and risk-based strategies. By comparing N, M, H for the uniform and risk-based approaches, the analysis reveals trade-offs between labor expenditure and failure detection. This benchmarking framework provides transportation agencies with quantitative evidence to support data-informed decisions for allocating limited NDT resources toward the most impactful risk mitigation across the culvert network.

3 Case Study Application

Relative hazard rates for two failure modes, hydraulic inadequacy and structural weakness, were obtained by fitting separate Γ-frailty adjusted Cox proportional hazards models and then computing the model's partial-hazard predictions for each culvert. In effect, these 'relative hazards' quantify the multiplicative risk of serviceability failure for culvert i relative to the network average, after adjusting for the observed covariates. Figure 2 displays histograms of the estimated partial-hazard values $h_i^{(1)}$ (hydraulic) and $h_i^{(2)}$ (structural). The hydraulic inadequacy serviceability hazards lie predominantly between 0.5 and 1.5, with a roughly symmetric distribution about unity, indicating that most culverts experience near-average hydraulic risk, with fewer extreme outliers. In contrast the structural weakness serviceability hazards $h_i^{(2)}$ are more right-skewed, ranging from about 0.6 to 2.5, reflecting a longer tail of high-risk culverts subjected to elevated truck loads or aggressive soil pH. These marginal hazards estimates capture the site-specific vulnerability of each culvert to its dominant deterioration process, providing the foundation for subsequent dependence modeling and risk-driven scheduling.

To translate the joint-hazard insights into actionable inspection priorities, each culvert's relative hazards $(h_i^{(1)}, h_i^{(2)})$ were dichotomized at their 66th percentiles. This classification yielded four joint-risk quadrants: low-low $[0, 0]$, low-high $[0, 1]$, high-low

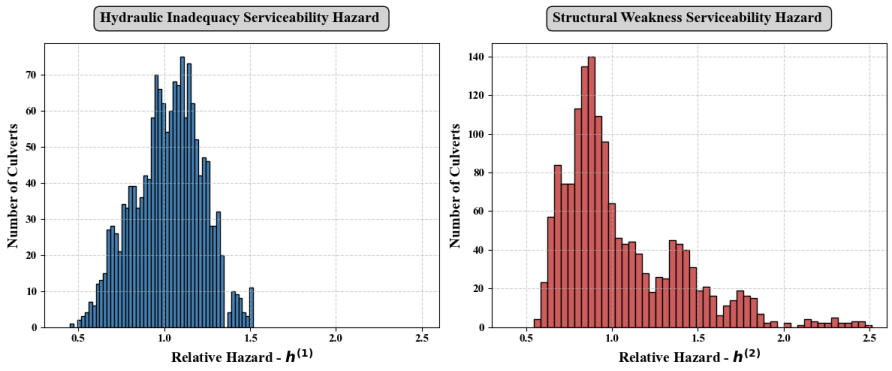

Fig. 2 Histograms of Estimated Relative Hazards

$[1, 0]$, and high-high $[1, 1]$, as illustrated in Figure 3. The resulting distribution shows that nearly half the network (46%) falls into the low-low quadrant, indicating minimal immediate concern. In contrast, 13% of culverts occupy the dual high-risk [1,1] quadrant, identifying a critical subset that may warrant more frequent inspections or enhanced monitoring under a risk-informed inspection framework. The remaining culverts are split almost evenly between the structural-only [0,1] (21%) and hydraulic-only [1,0] (21%) categories, suggesting targeted, risk-specific inspection priorities.

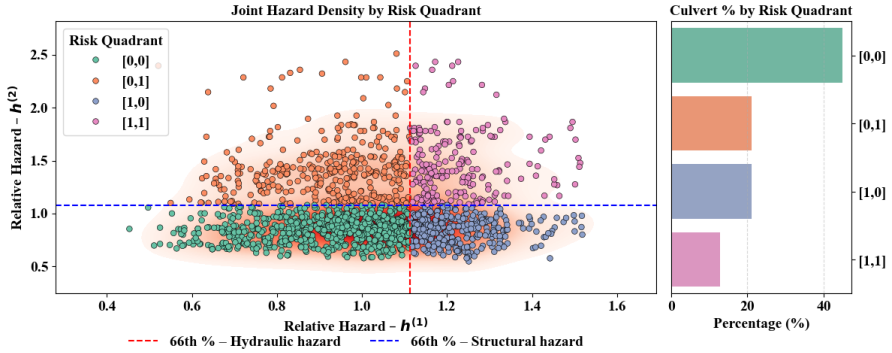

Fig. 3 Joint hazard scatter with 66th-percentile quadrants and risk-group percentages

Figure 4 plots each culvert's pair of pseudo-observations (u_i, v_i), obtained by mapping the hydraulic and structural hazard rates through their ECDFs, onto the unit square $(0, 1)^2$. The diagonal line $u = v$ marks perfect equality of ranks in two dimensions. If the two hazards were strongly dependent, the points would cluster tightly around that diagonal. Instead, the cloud of points fills the unit square almost uniformly. The visual impression is confirmed quantitatively by the nonparametric Kendall's rank correlation $\tau \approx 0.043$ (two-sided $p \approx 0.0102$), indicating a statistically

significant but practically negligible positive association. This weak dependence justified the use of weighted geometric-mean aggregation as the primary joint-risk metric. A detailed dependence threshold analysis for this dataset is provided in Appendix A.

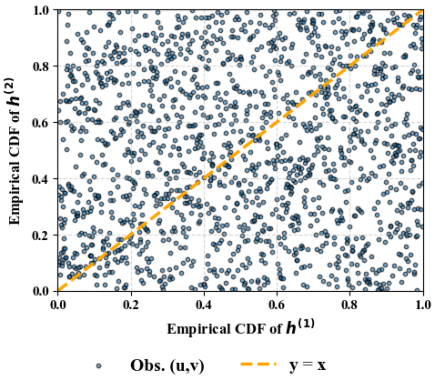

Fig. 4 Empirical CDF transforms of $(h^{(1)}, h^{(2)})$ to (u, v)

The weighted geometric mean surface computed at equal weighting ($\alpha = 0.5$) increases monotonically along both axes, indicating that the composite index grows as either hazard $h^{(1)}$ or $h^{(2)}$ increases, with the steepest gradient observed when both hazards are simultaneously high Fig. 5. Moderate hazard pairs occupy mid-range composite scores (around $1.0 - 1.2$), while only culverts with simultaneously elevated hydraulic and structural serviceability hazard rates exceed (1.5). This behavior ensures that dual-high risk culverts are distinguished clearly in the joint-risk ranking, yet isolated extremes in one mode do not unduly inflate the composite.

A grid search benchmarking every candidate schedule against the Ethiopian Roads Administration's uniform three-year plan, assuming two labor hours per inspection over a thirty-year horizon, identified optimal joint-risk percentile thresholds of $(p_1, p_2, p_3, p_4) = (0.995, 0.94, 0.68, 0.38)$. However, thresholds derived from the copula model proved highly sensitive to sampling noise under the observed weak concordance, limiting their stability and practical utility for this dataset (Appendix B).

Applying these cut-points coverts continuous joint-risk scores into annual, biennial, triennial, quinquennial and decennial inspection intervals, respectively. Under this scheme, the very highest-risk culverts fall into the annual tier, while the lowest-risk assets are deferred to quinquennial or decennial cycles. The optimized five-level schedule allocates inspection interval according to risk quadrant, Fig. 6. Culverts with both hazards below the 66th percentile fall almost exclusively into five (12.5%) and ten-year (32%) intervals. In the quadrants with one hazard above the threshold, most structures receive three to five-year inspections. Culverts in the highest-risk quadrant though are confined to annual (0.5%), biennial (4.3%), and triennial (8%) intervals, ensuring no asset with elevated hazard levels is scheduled beyond three years.

The Pareto cumulative coverage curve (Fig. 7) illustrates how network-wide risk decreases as inspection effort increases. Under the optimized schedule inspecting

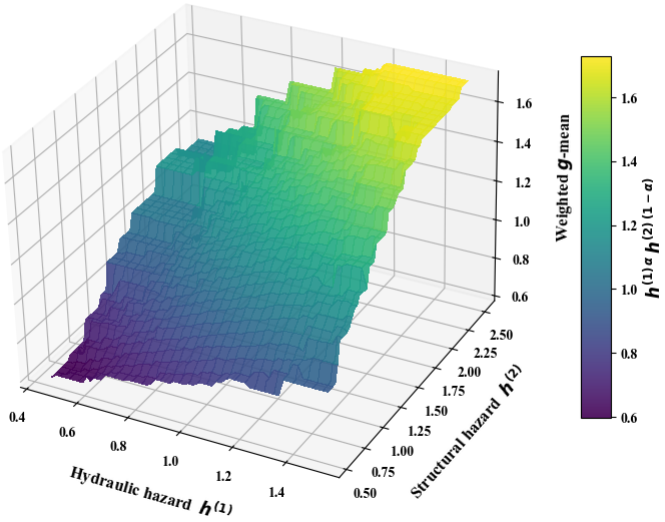

Fig. 5 Weighted geometric mean risk score ($\alpha = 0.5$)

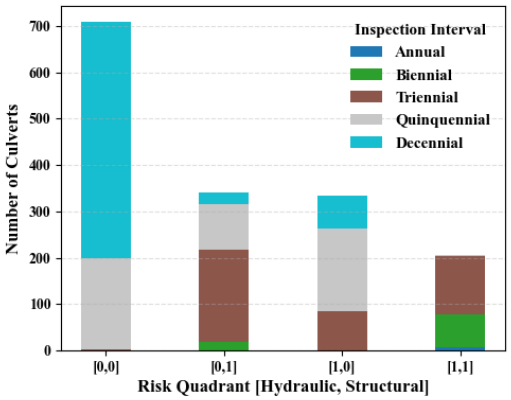

Fig. 6 Inspection frequency by risk quadrant

roughly half the network ($\approx$ 53%) already captures 80% of the cumulative joint risk, outperforming the uniform schedule at comparable effort levels. The steep initial rise shows that the highest-risk 20 percent account for roughly half of the network's risk, while beyond the 53 percent mark the curve flattens, indicating sharply diminishing returns on inspection effort. The risk-prioritized inspection regime can achieve substantial risk reduction with far fewer inspections than a uniform interval strategy.

Finally, resource and reliability benchmarking (Fig. 8) confirms that, for essentially the same labor-hour budget, the risk-based five-bin strategy substantially reduces the expected number of missed failures relative to the one size fits all three-year plan. The

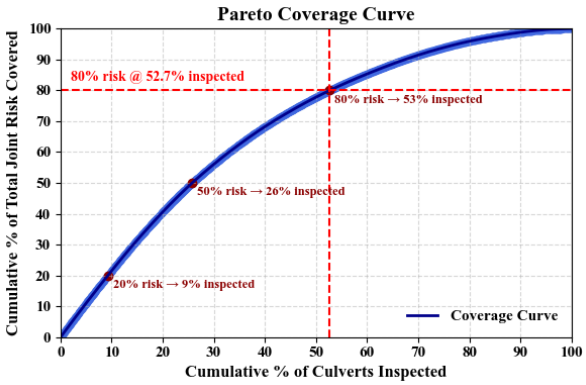

Fig. 7 Network-wide risk coverage as a function of inspection effort

optimized plan matches the labor constraint while further reducing missed failures by an additional 35% relative to the initial mapped-risk schedule.

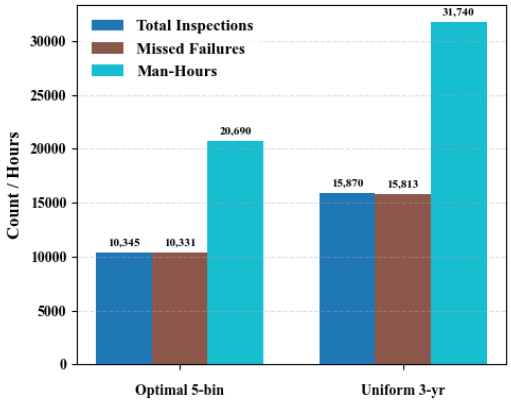

Fig. 8 Inspection resource comparison among uniform 3 year and optimal 5 bin plans

4 Conclusions

This study presents a dual risk inspection planning methodology that integrates marginal serviceability hazard rates for hydraulic inadequacy and structural weakness in reinforced concrete culverts into a unified joint risk metric that adapts to the observed dependence structure. Empirical concordance assessed through Kendall's τ of approximately 0.043 ($p = 0.0102$) indicated negligible correlation between the two hazard modes. This finding supports the use of transparent, distribution free weighted geometric mean for risk aggregation while retaining the copula-based density score as a flexible alternative when stronger dependence is observed.

Each culvert was classified into one of four risk quadrants: low-low, low-high, high-low, and high-high, based on its position relative to the 66th percentiles of the two hazard distributions. This quadrant-based classification ensures that assets with mixed or dual vulnerabilities receive inspection intervals that are tailored to their combined risk profile, rather than relying on a uniform inspection cycle. A constrained brute-force grid search over a lattice of 25 candidate percentile values ranging from 0.30 to 1.0 yielded 12,650 strictly descending threshold quadruplets. Benchmarking each candidate schedule against a uniform three-year inspection cycle, with the assumption of two labor hours per culvert inspection over a thirty-year horizon, identified optimal joint-risk cutoffs of $p_1 = 0.995$, $p_2 = 0.94$, $p_3 = 0.68$, $p_4 = 0.38$. These thresholds correspond to annual, biennial, triennial, quinquennial, and decennial inspection intervals.

Pareto analysis of the optimized schedule showed that inspecting 53% of the culverts captures 80% of the total joint risk. This confirmed the presence of diminishing returns when inspections are extended to lower-risk portions of the network. By ranking culverts in descending order of risk, the scheduling strategy prioritized inspections where they provide the greatest benefit. Resource and reliability benchmarking revealed that the risk-based schedule requires 10,345 inspections and 20,690 man-hours over thirty years. This represents an approximate 35% reduction compared to the uniform three-year cycle, while simultaneously decreasing the expected number of missed failures from 15,813 to 10,331. These results highlight significant gains in both efficiency and safety coverage under the same labor-hour budget.

The proposed methodology offers transportation agencies a practical and transparent tool for allocating NDT resources toward structures that most critically influence network-wide reliability and cost-effectiveness. By replacing rigid calendar-based cycles with adaptive, data-informed intervals, agencies can achieve significantly greater risk mitigation per unit of inspection effort.

5 Research Outlook

To broaden the applicability and robustness of the proposed dual-risk inspection framework, several avenues merit exploration. First, incorporating results from each NDT campaign back into the hazard models would allow real-time recalibration of both marginal hazards and the joint dependence structure. As fresh condition assessments arrive, Bayesian updating or sequential Cox frailty refitting can refine parameter estimates, continuously improving inspection prioritization accuracy. Second, culverts located in proximity often share environmental stressors such as watershed characteristics and soil type that induce correlated deterioration beyond what a purely copula-based dependence captures. Embedding a spatial random effect or a geographically weighted copula layer would enable clustering of inspection targets by locale, reducing travel costs and uncovering region-specific failure patterns. Finally, while this study focused on reinforced concrete culverts, the methodology readily generalizes to other infrastructure asset networks with multiple failure modes, for example, bridges with combined fatigue and scour risk or buried pipelines subject to corrosion and pressure loads. Adapting the hazard covariates and copula families to these contexts

can extend risk-based inspection planning across transportation, water, and energy infrastructure portfolios.

6 Recommendations

To ensure that limited NDT man-hour budget B are allocated where they yield the greatest network-wide risk reduction, inspection planning can be framed as a $0-1$ knapsack problem, $\max_{x_i \in \{0,1\}} \sum_{i=1}^{N} (w_i x_i)$ subject to $\sum_{i=1}^{N} c_i x_i \leq B$, where x_i indicates whether culvert i is selected for inspection, w_i is its joint-risk score, c_i is the per-inspection cost in hours. Some solution methods are recommended to guarantee that every available hour contributes optimally to risk mitigation. For moderate budgets, a dynamic programming algorithm constructs an $N \times B$ table in $O(NB)$ time, examining each of the N culverts at B budget levels (i.e., computation time grows roughly linearly with both N and B). When B is large or additional constraints are introduced, the problem can be formulated as an integer linear program and solved via branch and bound (e.g., CPLEX, Gurobi, or open-source CBC) which will deliver the global optimum. Employing these techniques ensures that NDT deployments concentrate precisely on the culverts whose inspection yields the greatest risk reduction per unit of resource.

Acknowledgments

The authors would like to thank Dr. Asnake Adraro Angelo and Dr. Kotaro Sasai for their valuable guidance in identifying suitable journals for this research. The authors also thank the Ethiopian Roads Authority for providing the reinforced concrete culvert dataset, and acknowledge the availability of public data sources that supported this work. Mearg N. Sahle gratefully acknowledges Mr. Zachery T. Howell for providing essential computing resources, and extends heartfelt gratitude to Kidane Mihret for unwavering support and encouragement throughout this research.

Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

This study builds upon hazard estimates and deterioration modeling outputs developed in a prior study. The original culvert condition dataset, including replacement records and operational variables, was obtained from the Ethiopian Roads Authority under restricted access and is not publicly distributable. Derived outputs, such as hazard scores and serviceability risk classifications, were generated using the methods described in the preceding study. Researchers interested in accessing either the underlying dataset or the derived outputs for validation purposes may contact the Ethiopian Roads Authority and the corresponding author, respectively, subject to approval.

Funding Sources

This research was supported in part by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan through the Japanese Government (MEXT) Scholarship Program, awarded to Mearg Ngusse Sahle. The authors also acknowledge institutional support provided by Osaka University. The funding source had no involvement in the study design, data collection, analysis, interpretation of data, writing of the manuscript, or the decision to submit the article for publication.

Appendix A Dependence Threshold Validation

Empirical model-selection diagnostics on the Ethiopian culvert dataset indicate that a copula framework yields appreciable gains only when the two hazard dimensions exhibit at least moderate concordance. This was assessed by computing Kendall's rank correlation coefficient τ on the full set of frailty-adjusted hydraulic and structural hazard rates, followed by recomputation across 1,000 bootstrap resamples of the pseudo-observations (i.e., the ECDF–mapped hazard ranks). For each resample, a one-parameter Frank copula was fitted and its drop in AIC and reduction in CvM were measured relative to the assumption of independence (i.e., geometric-mean baseline).

Plotting ΔAIC and ΔCvM against the absolute value of τ for each iteration produces an elbow cloud (Fig. 9). The copula fitting yielded almost no AIC improvement, or even a small penalty, when $\tau \lesssim 0.02$, and ΔCvM remains essentially flat throughout. Only as $|\tau|$ grows beyond roughly 0.08 does ΔAIC climb steeply, marking a clear inflection point where the extra flexibility of the copula structure is paid back by substantially better fit despite its added complexity. Since the observed Kendall's τ falls well below this threshold, fitting a copula here provides only minimal AIC benefit and risks over-fitting sampling noise rather than capturing genuine dependence. Accordingly, $\tau \approx 0.15$, borrowed from Cohen's benchmarks for rank correlations [**?**], serves as an effect-size demarcation below which the geometric mean index suffices, and above which a full copula-based joint-risk model is justified.

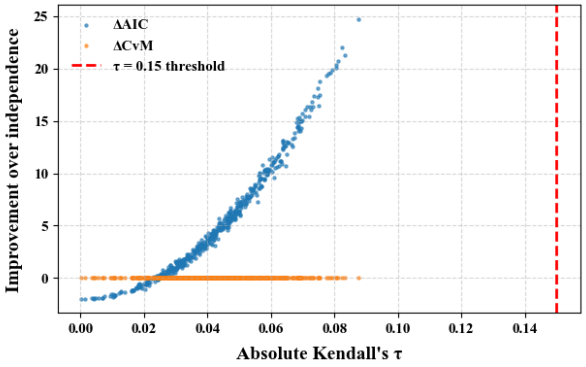

Fig. 9 Elbow Plot: Copula vs Independent Model

Appendix B Copula Density Score Modeling

This appendix presents the copula-based joint-density framework explored to keep every tested dependence structure fully transparent and reproducible. Diagnostics indicate that hydraulic and structural hazards in the current dataset are nearly independent (empirical Kendall's $\tau \approx 0.043$), so the copula offers no predictive benefit. However, the method is retained here because it becomes indispensable whenever the two hazards exhibit stronger concordance in future datasets.

A copula cleanly decouples the marginal behavior of each hazard from their joint dependence structure, allowing highly flexible modelling of non-linear or tail-specific associations without imposing parametric restrictions on either margin. To characterize the joint dependence between hydraulic and structural serviceability hazards, four candidate copula models were fitted to the transformed pseudo-observations and evaluated using both likelihood and goodness-of-fit criteria. Table 2 summarizes the fitted dependence/correlation parameter (θ/ρ), the maximized log-likelihood, and the corresponding AIC. Among the four families, the Frank copula emerged as the most suitable for modeling the joint dependence structure. It achieved the highest maximized log-likelihood value (3.262) among the four candidate families, indicating that the observed pseudo-observations (u_i, v_i) are most probable under the Frank copula density. After penalizing for model complexity using the AIC, the Frank copula remains the best-performing model, with an AIC of (-4.52), sustainably lower than that of Clayton (0.20), Gumbel (-1.91), and especially Gaussian (9033.23). This notably negative AIC suggests an excellent trade-off between goodness-of-fit and model parsimony, particularly given the Frank copula's single-parameter formulation.

In terms of goodness-of-fit, as measured by the CvM distance, the Frank copula again performs competitively, with $W^2 = 2.01 \times 10^{-5}$, closely matching Gumbel's marginally smaller value of 1.95×10^{-5}. These results indicate that both copulas approximate the empirical copula distribution with high fidelity. Furthermore, all three Archimedean copulas imply a Kendall's τ of approximately 0.043, which is in perfect agreement with the empirical rank correlation observed in the data. In contrast, although the Gaussian copula implies a similar level of positive dependence ($\tau \approx 0.041$), it performs poorly in both log-likelihood and AIC, indicating that it misrepresents the joint behavior, especially in the tails and central mass of the distribution. Taken together, the superior log-likelihood, lowest AIC, and competitive CvM distance of the Frank copula, along with its ability to reproduce the observed Kendall's τ, justify its adoption as the preferred model for joint-risk analysis of hydraulic inadequacy and structural weakness hazards. Figure 10 illustrates the fitted contour plots for each copula model, showing how well the Frank copula capture the empirical dependence pattern.

Kernel density estimation (KDE) was used to visualize the marginal distributions of hydraulic inadequacy and structural weakness serviceability hazards, offering smooth, non-parametric representations of their underlying probability densities. The KDE overlays reveal a common mode near unity but notably different dispersions (Fig. 11). The hydraulic hazard distribution is sharply peaked and largely constrained below 1.5, while the structural hazard distribution displays a heavier right tail extending beyond 2.0. The shaded overlap between approximately 0.5 and 1.3 indicates where both

Table 2 Copula fits on (u, v) pseudo-observations

Copula	Parameter	Log-likelihood	AIC	CvM	Kendall's τ
Frank	0.3892	3.262	-4.52	2.007×10^{-5}	0.043
Clayton	0.0902	0.896	0.20	2.043×10^{-5}	0.043
Gumbel	1.0451	1.957	-1.91	1.953×10^{-5}	0.043
Gaussian	0.0641	-4515.616	9033.23	1.898×10^{-5}	0.041

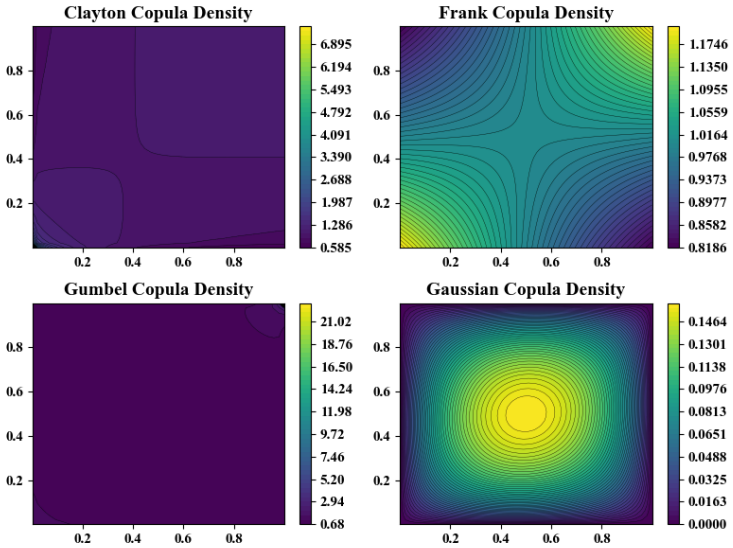

Fig. 10 Fitted copula density contours for hydraulic and structural hazards

risks co-occur for most culverts, whereas the extended green tail highlights a smaller subset with elevated structural vulnerability. Because the marginal distributions share a common mode but exhibit markedly different behavior in the tails, applying the same simple parametric form such as a normal distribution to both would be inappropriate. A nonparametric approach like KDE allowed each margin's unique shape, including its extreme values, to be learned directly from the data. Once the individual marginal distributions are flexibly captured, a copula can be employed to join them and model their dependence structure without imposing unrealistic assumptions on either margin.

The Frank-copula density surface peaks at the corners while maintaining appreciable interior mass (Fig. 12), indicating a weak but positive joint-hazard dependence. Culverts with high hydraulic-risk percentiles also tend to exhibit moderately elevated structural-risk percentiles, without pronounced tail clustering. When transformed back into physical hazard space, the joint density $f\left(h^{(1)}, h^{(2)}\right)$ reveals a clear ridge of heightened joint-failure risk around $h^{(1)} \approx 1$ and $h^{(2)} \approx 0.8$, tapering rapidly toward the extremes, indicating that most culverts occupy moderate hazard levels, with only a small subset simultaneously at high risk.

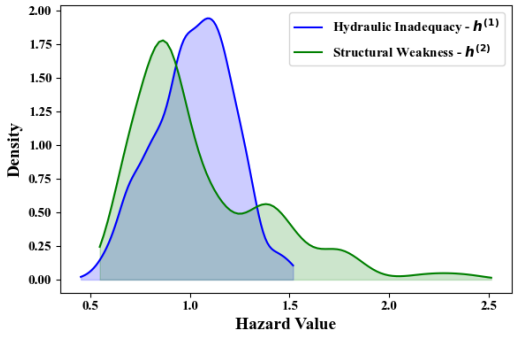

Fig. 11 Overlaid kernel density estimates of relative serviceability hazards

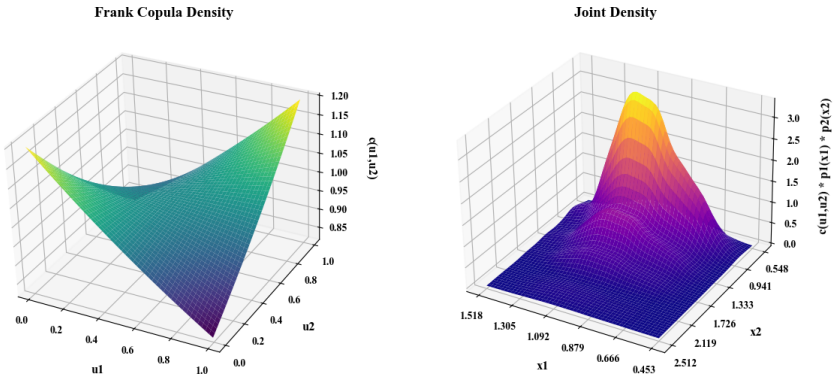

Fig. 12 Frank copula and joint hazard density surfaces

The inspection schedule was optimized by benchmarking it with a uniform inspection policy implemented by the Ethiopian Roads Authority, which mandates detailed inspections of every culvert once every three years ($\Delta_{\text{uni}} = 3$) over a 30-year horizon Ψ, and assuming a man-hour budget c_i of 2 hours per NDT inspection per culvert. Rather than relying on ad-hoc thresholds, a four-dimensional grid-search over descending joint density percentile cut-points $\{p_1 > p_2 > p_3 > p_4\}$ was performed to minimize the total missed failures. The resulting optimal breakpoints $(p_1, p_2, p_3, p_4) = (0.99, 0.9696, 0.9492, 0.9288)$ map each culvert's empirical joint-density percentile into inspection intervals where annual ($\geq p_1$), biennial ($p_2 - p_1$), triennial ($p_3 - p_2$), quinquennial ($p_4 - p_3$), and decennial ($< p_4$). These cutoffs classified only the top 1% of culverts in the annual category, followed by 2.04% each in biennial, triennial, and quinquennial tiers, and the remaining 92.88% assigned to decennial intervals. Because the joint-density distribution exhibited weak dependence (Kendall's $\tau \approx 0.043$), the optimized percentile cutoffs only captured a narrow band of culverts in the highest-frequency tiers, resulting in the misclassification of many high-risk $[1, 1]$ culverts into infrequent schedules such as quinquennial or decennial intervals.

References

[1] Watkins RK (1960) Failure conditions of flexible culverts embedded in soil. *Proc* 39:361–371

[2] Moore M (2000) Reliability of visual inspection for highway bridges. Wiss, Janney, Elstner Associates

[3] Beaver JL, Richie MC (2016) Culvert and storm drain system inspection manual. *NCHRP Project 14-26 Final Report*. Transportation Research Board, National Academies. Available at: http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP14-26_FR.pdf

[4] Youngblood D (2017) Enhanced culvert inspections – best practices guidebook. *Final Report MN/RC 2017-16*. Minnesota Department of Transportation, St. Paul, MN. Available at: http://www.dot.state.mn.us/research/reports/2017/201716.pdf

[5] Grier SM, Williams CS (2022) Large culvert inspection procedures: guidelines for INDOT. *FHWA/IN/JTRP-2022/27*. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, IN. Available at: https://docs.lib.purdue.edu/jtrp/2053/

[6] Government of the Northwest Territories, Department of Transportation (2013) Structures inspection standard operating procedures. *Technical Report*. Available at: https://www.inf.gov.nt.ca/sites/inf/files/resources/gnwt_transportation_structures_inspection_sop_v2.6_final.pdf

[7] Ministry of Transportation, Ontario (2018) Ontario structure inspection manual (OSIM). Available at: https://www.northdundas.com/sites/1/files/2021-09/OSIM%20Bridge%20and%20Culvert%20Report%2C%202020.pdf

[8] Yang C, Allouche E (2009) Evaluation of non-destructive methods for condition assessment of culverts and their embedment. In: *ICPTT 2009: Advances and experiences with pipelines and trenchless technology for water, sewer, gas, and oil applications*, pp 28–38

[9] Kalhor D, Ebrahimi S, Tokime RB, Mamoudan FA, Bélanger Y, Mercier A, Maldague X (2020) Infrared thermography for culvert inspection. In: *Proc Structural Health Monitoring–NDT 2020 Conference*. NDT.net. Paper No. 14. Available at: https://www.ndt.net/article/shmndt2020/papers/SHM-NDT_2020_paper_14.pdf

[10] Schroeder SC, Frankel J, Abbate A (1996) On-site ultrasonic inspection of highway culverts. In: *Review of Progress in Quantitative Nondestructive Evaluation: Volume 15A*, Springer, pp 1875–1881

[11] Mohammadi P, Asgari S, Rashidi A, Alder R (2025) Culvert inspection framework using hybrid XGBoost and risk-based prioritization: Utah case study. *J Constr Eng Manag* 151(6):04025052

[12] Gao C, Elzarka H (2021) The use of decision tree based predictive models for improving the culvert inspection process. *Adv Eng Inform* 47:101203. https://doi.org/10.1016/j.aei.2020.101203

[13] Piratla KR, Jin H, Yazdekhasti S (2019) A failure risk-based culvert renewal prioritization framework. *Infrastructures* 4(3):43. https://doi:10.3390/infrastructures4030043

[14] Patel RJ (2005) Risk based inspection-Qatargas Operating Company Limited. In: *Middle East Nondestructive Testing Conference & Exhibition*

[15] Morato PG, Papakonstantinou KG, Andriotis CP, Nielsen JS, Rigo P (2022) Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes. *Struct Saf* 94. https://doi.org/10.48550/arXiv.2009.04547

[16] Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Routledge. https://doi.org/10.4324/9780203771587

[17] Ethiopian Roads Authority (2001) Bridge inspection Amharic manual. Ethiopian Roads Authority, Addis Ababa, Ethiopia.