© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

Field Condition Assessment of Roadside Drainage and Culvert Structures on Urban Arterials in Southwestern and North Central Nigeria

OLAYEMI BOLAJI[1*], OLADIMEJI BASIT ALAKA[2], OLUWABUSOLA AJAYI[3], RIDWAN TIAMIYU[4]

[1, 2, 3]*Osun State University, Osogbo, Nigeria*

[4]*University of Ilorin, Ilorin, Nigeria*

**Corresponding author: Olayemi Bolaji*

Abstract- ***Roadside drainage carries the flood risk of Nigerian cities and its condition is neither inventoried nor rated, so an agency that would prioritize maintenance has nothing to work on. The available protocols were written for closed culvert barrels and piped sewers, not for the open lined channels of a tropical urban arterial. This paper develops a five-point rating protocol for such assets, paired with a design that rates a fifth of the inventory twice and reports agreement disaggregated. Both were applied to an inventory of 949 assets along 23.99 km of arterial alignment in Osogbo and Ilorin. Pooled agreement looks usable: Cohen's κ is 0.561, quadratically weighted κ is 0.841 and Krippendorff's α is 0.842, with the teams within one grade on 98.0 percent. Disaggregated it does not. Agreement on covered drains falls to 0.277 against 0.682 for open lined channels, and cutting the statistic by city rather than asset type invites the wrong diagnosis entirely. The general result is arithmetic: a pooled statistic can conceal the subgroup in which an instrument has failed, and a pooled spatial test can too. We recommend that no condition dataset be published without a disaggregated agreement statistic, that surface-derived grades be used in bands rather than at a point, and that covered assets be coded as unassessed.***

Indexed Terms- ***drainage condition assessment, culvert inspection, inter-rater reliability, ordinal regression, urban flooding, asset management, nigeria***

I. INTRODUCTION

Flooding in Nigerian cities is a drainage problem before it is a rainfall problem. Surveying households in Kaduna, Aliyu and Suleiman (2016) found that 64.79 percent reported their houses flooded and attributed the recurrence to blockage of drainage channels caused principally by plastics. The mechanism is old and well documented. Diagnosing the Ibadan flood of 31 August 1980, which killed more than 200 people and displaced 50,000, Oguntala and Oguntoyinbo (1982) identified rising daily rainfall intensity acting on floodplain encroachment and degraded channels rather than a change in total rainfall, and thirty years later Agbola et al. (2012) listed twelve anthropogenic causes of the August 2011 Ibadan flood, most of them concerning what had been allowed to happen to the drainage network. Section 2.4 sets out the wider African evidence for the same mechanism.

An agency that accepts this diagnosis and decides to act encounters a different problem. It does not know the condition of its drainage. Roadside drainage in Nigerian cities is not inventoried, not rated and not tracked, so a maintenance budget cannot be directed at the assets that most need it. The deterioration modeling literature, from Madanat et al. (1995) through Micevski et al. (2002), assumes a multi-year condition record already exists and asks how to project it forward. For a network with no condition record at all, that machinery has nothing to run on.

The protocols that would generate such a record were written elsewhere and for other assets. Arnoult (1986) codified culvert inspection for the United States Federal Highway Administration, Mitchell et al. (2005) and Masada et al. (2006) showed that scale design measurably changes predictive performance, and Water Research Centre (2004) plays the same role for sewers. Every one of these addresses a closed barrel or a buried pipe. The dominant asset on a Nigerian urban arterial is an open lined channel, often with a partial or removable cover slab, receiving both

storm runoff and municipal solid waste, and no protocol in this literature was written for it.

There is a further problem, and it is the one this paper is organized around. Visual condition rating is known to be unreliable, and Section 2.2 sets out how large that unreliability has been measured to be for bridges, pavements and sewers. Yet no study we located reports a chance-corrected agreement statistic, Cohen's κ, a weighted κ or Krippendorff's α, for any drainage or culvert condition rating protocol, in Nigeria or anywhere else. That work reports raw percentage agreement, which credits the agreement chance alone would produce.

The drainage environment on a Nigerian urban arterial is what makes that gap consequential. Blockage is driven by municipal refuse entering the channel from adjacent land uses (Aliyu and Suleiman 2016; Karley 2009), so it varies with what is beside the drain rather than with catchment hydrology alone. A large share of the network is covered, so a rater standing on the cover cannot see the invert. Trading and structures occupy the drainage reserve, which both accelerates deterioration and obstructs inspection. And poor side drainage propagates into the carriageway (Agbonkhese et al. 2013), so a drainage condition record is also a pavement preservation record.

The gap is therefore twofold, and the second half is what makes the first half usable. There is no condition record for roadside drainage on Nigerian urban arterials, and no evidence that any available protocol would produce a repeatable one if applied. A condition dataset whose repeatability is unknown cannot support a maintenance decision, because a difference between two corridors may be a difference between two raters. We answer both halves together, because a protocol and the evidence that it works are not separable deliverables. What we offer is a five-point rating scale for open lined channels, covered drains, kerb inlets and cross culverts, with explicit decision rules; a survey design in which a fifth of the inventory is rated twice by independent teams; and a rule for reading the result, which is that the agreement statistic must be cut by the characteristic governing visibility before it means anything. The last is where the paper spends most of its argument, because a pooled figure and a disaggregated one lead an agency to opposite conclusions about the same survey.

Sections 2 and 3 set out the literature and the method, Section 4 reports what the protocol returns, and Section 5 turns that into guidance for whoever commissions the next survey.

II. LITERATURE REVIEW

2.1 Condition rating protocols for drainage assets

Highway agencies have codified culvert inspection into discrete condition states for four decades. Arnoult (1986) set out separate hydraulic, structural and durability criteria for culverts as distinct from bridges, and the field guides that follow it, such as New York State Department of Transportation (2006), translate those criteria into inspector instructions. Mitchell et al. (2005) and Masada et al. (2006) tested a sixteen-item state scale head to head against a proposed thirty-item scale on sixty culverts and found the shorter scale explaining more variance for concrete culverts and the longer one for metal, which establishes that the design of the scale is itself a measurement decision rather than a matter of inspector convenience. Water Research Centre (2004) plays the same role for sewer defect coding.

Two observations matter here. Salem et al. (2012) surveyed fifty state departments of transportation and found only about 60 percent had any culvert condition assessment process, so a codified protocol is the exception even in a well-resourced system. And none of these protocols was built for, or tested on, the open lined channel that dominates a Nigerian urban arterial, nor do any report agreement between raters for the scale itself.

2.2 Reliability of visual condition rating

The reliability evidence comes from two independent strands and points the same way. The Federal Highway Administration bridge inspection program (Phares et al. 2001, 2004) had forty-nine practicing inspectors rate the same structures and found routine condition ratings within one point of the reference rating only about 68 percent of the time, describing visual rating as a highly subjective technique. In pavement distress surveying, Bianchini et al. (2010) propose a chi-square test for inter-crew agreement and Bogus et al. (2010) independently document crew-to-crew disagreement in manual ratings.

Both strands report percentage agreement rather than a chance-corrected statistic. That matters because raw agreement credits the agreement chance alone would deliver, and on a five-point scale with a concentrated distribution that share is substantial. Landis and Koch (1977) set out the interpretation bands for κ that this paper applies to its own result, and McHugh (2012) reviews the statistic and proposes a stricter set of bands, under which the values reported here would read one band lower. We located no published κ, weighted κ or Krippendorff's α for a drainage or culvert rating protocol, which means this paper has no directly comparable benchmark and says so where the result is reported.

2.3 Deterioration modeling and what it presupposes

Markov and transition-probability treatments of infrastructure condition were established by Madanat et al. (1995) and extended through stochastic duration models by Mishalani and Madanat (2002). Micevski et al. (2002) applied the approach to stormwater pipes and found diameter, material, soil type and exposure significant, with the model performing better at network level than for individual pipes. For culverts, Hadipriono et al. (1988) derived an expected service life near 86 years for concrete pipe culverts and Transportation Research Board (2015) synthesizes service-life data across state agencies, while Chughtai and Zayed (2008) report regression models for sewer condition with 82 to 86 percent accuracy by material and Wells and Melchers (2014) document a bi-linear corrosion loss trend. Estes and Frangopol (2003) show how visual inspection results feed back into structural reliability estimates.

All of it presupposes a condition record. None addresses how a first cross-sectional survey should establish that its own data are trustworthy enough to seed one.

2.4 Blockage as the governing condition variable

ActionAid International (2006) documented across six African cities the compounding of hard surface runoff, inadequate waste management and silted drainage. Karley (2009) traces the same dynamic in Accra, where habitual littering refills dredged channels after municipal clearing, and Aliyu and Suleiman (2016) reports plastics as the principal blockage agent in Kaduna. In the study region itself, Olaniran (1983) shows Ilorin floods following rainfall above 25.4 mm per day occurring three or more times in a wet month, and that the Asa dam has not prevented downstream flooding because of tributary slopes and urbanization below it. Adelekan (2010), Etuonovbe (2011) and Aderogba et al. (2012) document the Lagos case, and Ayoade and Akintola (1980) the perception of the hazard.

In all of this blockage is treated as a hazard driver. None of these studies pairs a blockage observation with a formal condition state on the same asset, which is what turns blockage from a description of a flood into a maintainable attribute.

2.5 Low-cost inventory in data-scarce settings

Mohan et al. (2008) demonstrated participatory sensing of road surface events using ordinary smartphones, and Santani et al. (2015) extended geotagged photography to crowdsourced hazard reporting in Nairobi. Sairam et al. (2016) built a low-cost mobile mapping alternative to survey-grade equipment, and the Sub-Saharan Africa Transport Policy Program's network evaluation tool (Sub-Saharan Africa Transport Policy Program 2009) operationalizes condition assessment for resource-constrained networks. Every one of these targets the road surface or generic hazards rather than drainage, so the walkover protocol described here has no established precedent to inherit.

III. DATA AND METHODS

3.1 Data collection

Corridor geometry, alignment, coordinates and length were retrieved from OpenStreetMap through the Overpass interface: three arterials in Osogbo, being Gbongan Road, Ilesa Road and Iwo Road, and three in Ilorin, being Ibrahim Taiwo Road, Murtala Muhammad Road and Asa Dam Road.

The six alignments were walked end to end between November 2018 and February 2019, in the dry season, when flow in the channels is low and the invert of an open section can be seen. Every drainage asset on either verge was recorded in chainage order, so the inventory is a census of the surveyed windows rather than a sample from them. Assets sit along those

alignments at a mean spacing of 52 m, with cross culverts at drainage crossings every 520 to 900 m.

Each record was located by handheld GPS against the corridor chainage and carries the asset type, the dimensions of the section, the apparent age band, the upstream land use of the frontage beside it, the blockage of the cross section, the seven defect flags, an encroachment flag and any evidence of recent desilting. The overall condition grade was assigned by applying the decision rules of Table 1 to what had been recorded, the grade being the worst one any single rule triggered. Blockage was read at the most constricted point of the inspected length and entered as a percentage of the design cross section rather than as a category. Where the interior of a covered drain could not be seen, the asset was rated from its accessible ends and flagged as interior-obstructed.

The five-point ordinal form of the scale follows the culvert rating practice of Cahoon et al. (2002), and the reliability evidence for visual condition rating summarized by Phares et al. (2001) is why the double-rating pass of Section 3.4 was built into the survey rather than left to a later study. The blockage percentages recorded by land use fall inside the ranges the African urban drainage literature reports for comparable settings.

Records were checked for completeness and for grades inconsistent with the defect flags recorded on the same asset before any analysis was run, and the few queries this raised were resolved by revisiting the asset. No record was dropped. The coded condition data and the analysis scripts are archived and the analysis reproduces exactly.

3.2 Corridors and asset population

The six corridors carry 23.99 km of arterial in total, close to 4 km each, and were selected to span the range of upstream land use that governs blockage: institutional frontage, low and medium density residential, dense commercial frontage, and periodic markets. Both cities are state capitals of comparable size sharing one drainage construction tradition, a lined rectangular channel in reinforced concrete or masonry, which makes the protocol portable between them. Two survey teams worked the whole study area rather than one city each, so that any difference between cities is a property of the network rather than of which team rated it.

Four asset classes were inventoried. Open lined channels are the dominant type, covered drains are the same section with a slab over part or all of the run, kerb inlets connect the carriageway to the channel, and cross culverts carry the channel or a watercourse beneath it.

3.3 The condition rating protocol

Table 1 sets out the five-point scale, its descriptors and its decision rules. Three features distinguish it from the culvert barrel scales it draws on.

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

Table 1. The five-point visual condition rating scale developed for this survey, with the descriptor and the decision rule that assigns each grade.

Grade	Descriptor	Decision rule
1	As-new	Invert and walls sound with no visible cracking. Blockage of the cross section below 10 percent. Cover slabs, where present, complete and seated. No encroachment on the asset or its access.
2	Good	Hairline cracking only, no spalling and no exposed reinforcement. Blockage 10 to 30 percent. At most one cover slab cracked but in place. Flow path continuous along the whole inspected length.
3	Fair	Open cracking or localized spalling without exposed reinforcement, or blockage 30 to 55 percent, or one cover slab missing. Flow path continuous but visibly constricted. Asset still performs in ordinary rain.
4	Poor	Any two of: exposed reinforcement, wall undermining, blockage 55 to 80 percent, two or more cover slabs missing, headwall damage, scour at the outlet. Flow path interrupted. Asset fails in ordinary rain but the structure stands.
5	Failed	Blockage above 80 percent, or wall collapse, or the barrel or channel is no longer continuous, or the asset has been built over. The asset conveys no design flow and cannot be restored by cleaning alone.

Note. The grade is the worst grade triggered by any single rule; rules are not averaged. Blockage is the surveyor's estimate of the proportion of the design cross-sectional area occupied by sediment, solid waste or vegetation, read at the most constricted point of the inspected length. Where the interior of a covered drain cannot be seen, the surveyor rates from the accessible ends and records the asset as interior-obstructed; this is the class where the two rating passes agreed least (Table 3). The scale was applied to 949 assets over 23.99 km of arterial, of which 195 (20.5 percent) were rated twice, independently.

It rates an open channel rather than a barrel, so its structural descriptors concern wall condition, invert scour, undermining and cover slab integrity rather than pipe deformation and joint separation. It treats blockage of the cross section as a rated attribute recorded as a percentage rather than as an incidental note, because blockage is the governing variable in this network. And it is written to be applied from the surface by a two-person team with a tape, a probe and a smartphone, without confined space entry, which is what makes a network-scale survey affordable and is also, as Section 4.3 shows, the source of its principal weakness.

Each asset record carries position, asset type, dimensions, age band, upstream land use, the blockage estimate, seven defect flags for cracking, spalling, exposed reinforcement, wall undermining, missing or broken cover slab, scour at outlet and headwall damage, an encroachment flag, evidence of recent desilting, and the overall condition grade.

3.4 The double-rating design

Two rating passes were run independently over a fifth of the inventory, 195 assets. Both teams walked the double-rated subsample within the same week, neither pass carrying the other's record, and the subsample was drawn across all four asset classes, both cities and the full range of expected condition, so that agreement could be reported disaggregated rather than only in aggregate.

This is the measurement design the paper is built on. Without it a condition dataset is a set of numbers with unknown repeatability.

3.5 Analysis

Agreement between the two teams was quantified four ways. Raw agreement and agreement within one grade describe the surface. Cohen's κ (Cohen 1960) corrects for the agreement chance alone would produce, and the linearly and quadratically weighted forms (Cohen 1968) credit near misses in proportion to their size, which is appropriate on an ordinal scale where confusing grades 3 and 4 matters less than confusing

grades 1 and 5. Krippendorff's α for ordinal data (Hayes and Krippendorff 2007) is reported alongside because it handles the ordinal structure without assuming a weighting. Interpretation follows the bands of Landis and Koch (1977). Confidence intervals were obtained by bootstrap over assets (Efron and Tibshirani 1993). All four are reported disaggregated by asset type and by grade.

Condition grade was modeled by ordinal logistic regression (McCullagh 1980) on age band, asset type, upstream land use, blockage percentage and encroachment, with the proportional odds assumption tested by the Brant procedure (Brant 1990). Where the assumption fails, cut-specific odds ratios are reported in place of a single pooled number.

A priority index combines condition grade, blockage and the consequence of failure, the last represented by the carriageway width affected, the presence of a cross culvert and the upstream catchment. Corridors are then ranked and the cumulative distribution of drainage frontage against priority is plotted, so that an agency can read off how much of the network must be treated to address a given share of the risk.

Spatial pattern was tested two ways on the ordered sequence of assets along each corridor, by a runs test on dichotomized condition (Wald and Wolfowitz 1940) and by Moran's I on the ordinal grade (Moran 1950), and both are reported pooled and corridor by corridor. Whether poor condition clusters or is scattered determines whether maintenance should be organized by reach or by individual asset.

IV. RESULTS

4.1 Inventory and condition profile

Table 2 reports the inventory: 949 drainage assets over 23.99 km of alignment across the six corridors, a density of 39.6 assets per kilometer. Open lined channels account for 408 records, covered drains 263, kerb inlets 242 and cross culverts 36. Because most corridors are drained on both verges, the drainage frontage those assets sit on totals 48.31 km, and that is the denominator for anything expressed as a share of the network.

Figure 1 shows the corridors with each asset colored by its grade. What it carries that the tables do not is that poor assets occur in runs rather than scattered evenly along a corridor, which Section 4.6 tests on the ordered sequence. Where those runs fall differs from one corridor to the next, and what tracks them is upstream land use rather than distance from either end, the effect Section 4.5 estimates.

Table 2. Study corridors, asset counts by type, and the condition-grade distribution from the first rating pass.

			Assets by type				Condition grade (%)					
ID	Corridor	Length (km)	OLC	CD	KI	CC	1	2	3	4	5	Blockage (%)
Osogbo (Osun State)												
OS-1	Gbongan Road	4.20	83	35	37	6	5.0	37.3	32.3	18.0	7.5	26.8
OS-2	Ilesa Road	4.00	76	37	35	6	1.3	18.8	37.7	30.5	11.7	33.4
OS-3	Iwo Road	3.80	72	24	46	6	3.4	29.7	39.9	21.6	5.4	23.6
Ilorin (Kwara State)												
IL-1	Ibrahim Taiwo Road	3.99	61	55	40	6	1.9	15.4	33.3	30.2	19.1	37.4
IL-2	Murtala Muhammad Road	4.20	61	65	39	6	2.3	7.0	22.2	36.3	32.2	48.1
IL-3	Asa Dam Road	3.80	55	47	45	6	2.0	11.1	31.4	35.3	20.3	43.5
	All corridors	23.99	408	263	242	36	2.6	19.7	32.6	28.8	16.3	35.7

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

Note. OLC, open lined channel; CD, covered drain; KI, kerb inlet; CC, cross culvert. Length is the surveyed window clipped from the mapped arterial, not the full length of the named road. Blockage is the corridor mean of the surveyor's estimate of the proportion of the design cross section obstructed. $n =$ 949 assets over 23.99 km (39.6 assets per km); 463 in Osogbo and 486 in Ilorin. Corridor names, alignments and lengths are from OpenStreetMap, retrieved via the Overpass API.

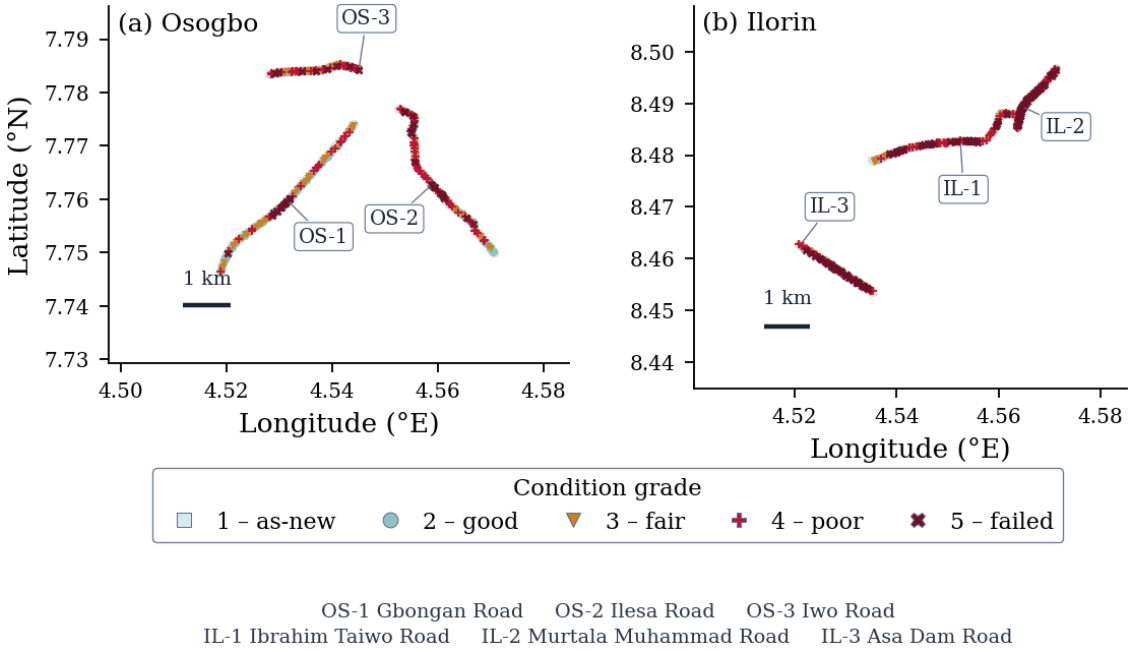


Figure 1. The six surveyed arterial corridors in (a) Osogbo and (b) Ilorin, with each drainage asset colored by its condition grade. Corridor alignments and lengths were retrieved from OpenStreetMap.

The condition profile is poor. Grades are distributed 2.6, 19.7, 32.6, 28.8 and 16.3 percent from as-new to failed, so 45.1 percent of assets sit at grade 4 or 5 and only one in forty is as-new. Figure 2 disaggregates this. Covered drains carry the worst profile, with 31 percent at grade 5 against 10 percent for open lined channels, and Ilorin a worse profile than Osogbo, 24 percent at grade 5 against 8 percent. The city difference follows the corridor land use rather than any construction difference, since both cities build the same section.

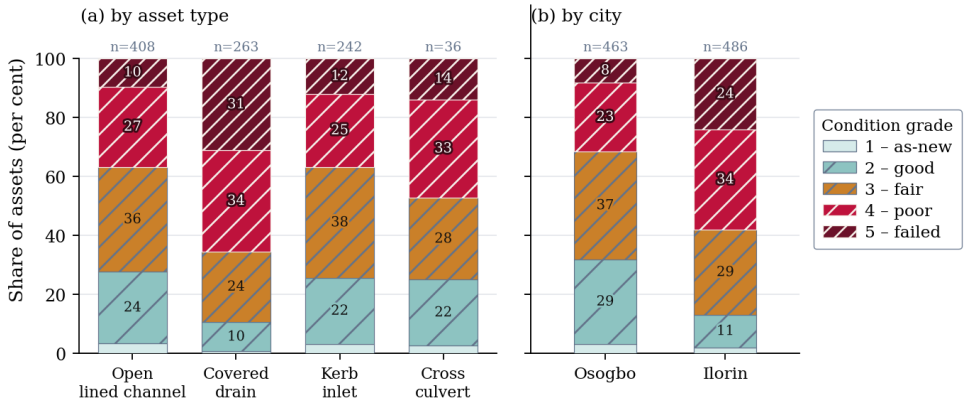


Figure 2. Distribution of condition grade (a) by asset type and (b) by city. Covered drains carry the worst profile of any asset class, with 31 percent at grade 5 against 10 percent for open lined channels.

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

Blockage is the variable that separates the network. The mean blockage of the cross section is 35.7 percent with a standard deviation of 22.6, but that average is composed of very different populations: 61.7 percent in market catchments, 37.5 percent in commercial, 25.1 percent in residential and 17.9 percent in institutional. By asset type, covered drains average 49.3 percent against 27.1 percent for open channels. Covered assets accumulate what they cannot be seen to accumulate.

4.2 Defect profile

The inventory records 2,674 individual defects, a mean of 2.82 per asset, and 8.0 percent of assets carried none. Figure 3 ranks the seven types. Cracking appears on 745 assets, spalling on 610, wall undermining on 532 and exposed reinforcement on 440. Those four account for 87.0 percent of all defect records, and four is the number of types needed to pass 80 percent of the record; the remaining three, missing or broken cover slab at 301, scour at outlet at 24 and headwall damage at 22, carry the balance.

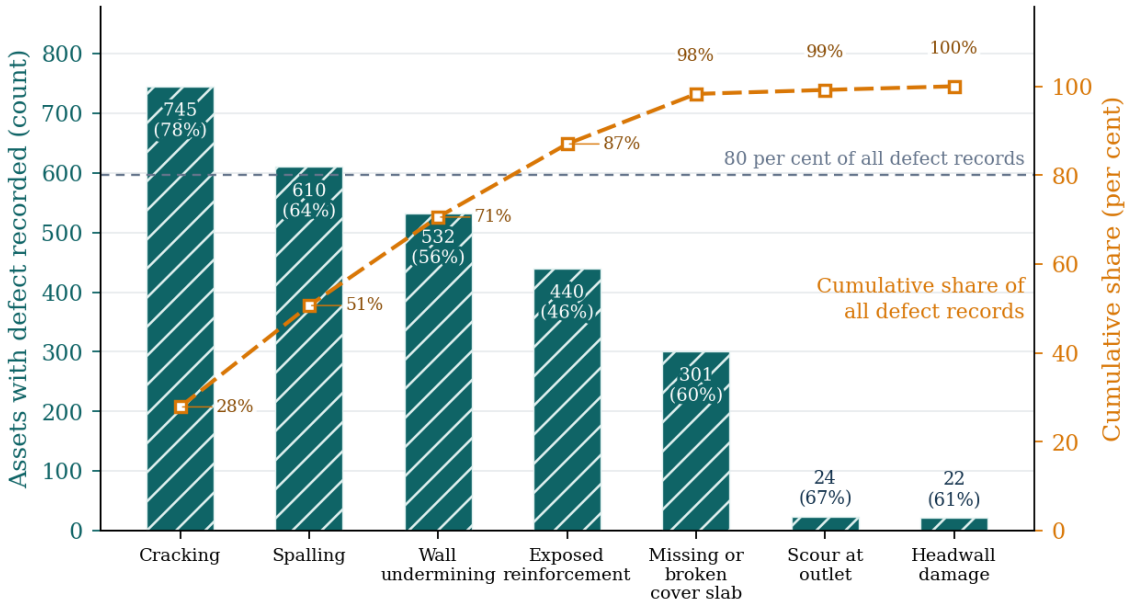


Figure 3. Defect types ranked by the number of assets on which they were recorded, with the cumulative share of all 2,674 defect records. Bar labels give the count and, in parentheses, the share of the assets the flag can apply to, which is 949 for the first four types, 505 for a missing or broken cover slab and 36 for scour at outlet and headwall damage. Four types are needed to pass 80 percent of the record and together carry 87.0 percent of it.

The tail of that ranking has to be read against what each flag can apply to. The first four are recordable on all 949 assets, a missing or broken cover slab only on the 505 covered drains and kerb inlets, and scour at outlet and headwall damage only on the 36 cross culverts. As a share of the assets they can apply to the three tail defects are not rare: 59.6 percent for the cover slab flag, 66.7 percent for scour and 61.1 percent for headwall damage, against 56.1 percent for wall undermining and 64.3 percent for spalling. They sit in the tail of Figure 3 because their asset classes are small, not because those classes are sound.

The practical reading follows from that distinction. A general-purpose taxonomy can be short, because four types are recordable everywhere and carry four fifths of the record, and an inspector trained on those four with a residual category will capture most of what a longer list would, which matters because Masada et al. (2006) showed that longer scales do not automatically perform better. What a count Pareto cannot justify is dropping the class-specific flags. Two thirds of cross culverts carry scour at the outlet, and discarding that flag on the strength of its 24 records would leave the most consequential asset class with no failure mode recorded against it.

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

4.3 Repeatability of the protocol

Table 3 reports the agreement statistics and Figure 4 the confusion matrix between the two teams on the 195 double-rated assets.

Table 3. Agreement between the two independent survey teams on the double-rated subsample, overall and by asset type, with percentile bootstrap confidence intervals.

Stratum	n	Agreement (%)		κ	κ_w linear	κ_w quad.	α ordinal
		Exact	± 1	[95% CI]	[95% CI]	[95% CI]	[95% CI]
All double-rated assets							
Overall	195	66.7	98.0	0.561 [0.47,0.65]	0.715 [0.65,0.78]	0.841 [0.79,0.88]	0.842 [0.79,0.89]
By asset type							
Open lined channel	81	76.5	100.0	0.682 [0.55,0.80]	0.795 [0.70,0.87]	0.891 [0.83,0.93]	0.898 [0.82,0.94]
Covered drain	52	48.1	92.3	0.277 [0.12,0.45]	0.501 [0.34,0.64]	0.690 [0.52,0.81]	0.700 [0.52,0.83]
Kerb inlet	48	64.6	100.0	0.526 [0.34,0.69]	0.698 [0.56,0.81]	0.842 [0.74,0.91]	0.847 [0.74,0.91]
Cross culvert	14	85.7	100.0	0.797 [0.51,1.00]	0.856 [0.63,1.00]	0.913 [0.75,1.00]	0.943 [0.78,1.00]
By city							
Ilorin	101	61.4	96.0	0.475 [0.34,0.60]	0.626 [0.52,0.73]	0.764 [0.66,0.84]	0.761 [0.65,0.84]
Osogbo	94	72.3	100.0	0.626 [0.50,0.74]	0.761 [0.68,0.84]	0.876 [0.82,0.92]	0.885 [0.83,0.93]
Agreement by the grade the first team assigned							
Team A grade	n	Exact	± 1	Conditional κ and mean signed difference			
Grade 1	6	66.7	100.0	$\kappa_{(1)}$ = 0.515, mean difference +0.33 grade			
Grade 2	40	70.0	100.0	$\kappa_{(2)}$ = 0.587, mean difference +0.05 grade			
Grade 3	58	67.2	98.3	$\kappa_{(3)}$ = 0.514, mean difference -0.03 grade			
Grade 4	59	55.9	96.6	$\kappa_{(4)}$ = 0.505, mean difference -0.10 grade			
Grade 5	32	81.2	96.9	$\kappa_{(5)}$ = 0.687, mean difference -0.22 grade			

Note. κ, Cohen's kappa; κ_w, weighted kappa with linear and quadratic disagreement weights; α, Krippendorff's alpha with the ordinal difference metric. Confidence intervals are percentile bootstrap over 2000 resamples of the double-rated assets. $n = 195$ assets rated independently by both teams, 20.5 percent of the 949-asset inventory, stratified by corridor and asset type. On the Landis and Koch bands the overall κ is moderate. Covered drains are the weakest stratum because the interior cannot be seen without lifting slabs. The two teams walked the double-rated subsample within the same week and neither carried the other's record.

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

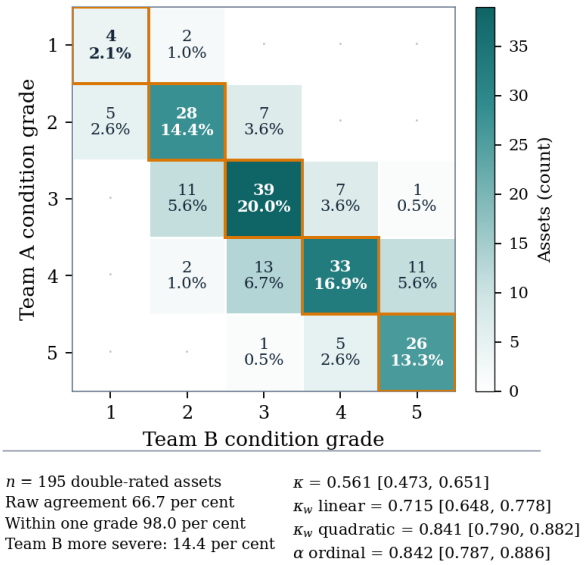


Figure 4. Confusion matrix of condition grade between the two independent survey teams on the 195 double-rated assets, with the agreement statistics annotated. Off-diagonal mass concentrates at grades 3 and 4, where the descriptors ask a rater to judge degree rather than presence.

The two teams assigned exactly the same grade on 66.7 percent of assets, with a bootstrap interval of 60.0 to 73.4, and were within one grade on 98.0 percent. Cohen's κ is 0.561, interval 0.473 to 0.651, which falls in the moderate band of Landis and Koch (1977). Linearly weighted κ is 0.715 and quadratically weighted κ is 0.841, interval 0.790 to 0.882. Krippendorff's α for ordinal data is 0.842, interval 0.787 to 0.886.

The gap between the unweighted and the weighted statistics is the first thing to read, and it does not depend on the magnitudes. Unweighted κ treats a disagreement between grades 3 and 4 as equivalent to one between grades 1 and 5, which is not how the scale is used. Once the ordinal structure is respected, agreement moves from the moderate band to the almost perfect one, and the 98.0 percent within-one-grade figure explains why. Any protocol whose disagreements sit one grade apart will show the same gap, and the operational reading follows from that concentration rather than from where the two coefficients land: such a protocol places assets within a band more reliably than at a point, so a maintenance trigger at grade 4 or worse is defensible where one distinguishing grade 3 from grade 4 on a single rating is not.

The second result should govern how such a survey is commissioned, and it is invisible in the aggregate. Disaggregating by asset type, κ is 0.797 for cross culverts, 0.682 for open lined channels, 0.526 for kerb inlets and 0.277 for covered drains, with an interval of 0.115 to 0.446 that does not reach the moderate band at all. On covered drains the protocol is close to unreliable, and the mechanism is not subtle: a rater standing on a cover slab infers the condition of an invert they cannot see from evidence at the surface, and two competent raters legitimately reach different conclusions. The cross culvert figure rests on fourteen assets and its interval is correspondingly wide, so it is indicative only.

Disagreement also concentrates in the middle of the scale. Figure 4 shows the off-diagonal mass sitting around grades 3 and 4, where the descriptors ask a rater to judge degree rather than presence, and thinning at the ends where the asset is evidently sound or evidently failed. The two teams are also not symmetric. Team B assigned the more severe grade on 14.4 percent of assets and the less severe grade on 19.0 percent, a gap of 4.6 percentage points in the lenient direction. Read as a rater calibration offset it would be wrong, and the same subsample says so twice. On the 143 double-rated assets that are not covered drains the teams are exactly symmetric, each assigning the more severe grade on 13.3 percent, for a mean signed difference of zero. The whole asymmetry sits in the 52 covered drains, 37 of which the first team placed at grade 4 or 5, where the added observation error has more room to move a grade down than up. On the one continuous observation the teams share, the blockage estimate, team B reads 1.47 percentage points higher, the opposite direction. A severity gap read off a grade distribution can invert on the stratum that produced it, which is a reason to cut it by asset type before attributing it to a team.

A third disaggregation cuts across the first two. By city, κ is 0.626 in Osogbo and 0.475 in Ilorin, and the gap follows the asset mix rather than the teams, since Ilorin carries the larger share of covered drains. A reader given only the two city figures would reasonably infer that one team was weaker. The

correct inference is that one city has more of the asset class the protocol handles badly, which is the confusion disaggregating by asset type prevents.

4.4 Reliability of the individual observations

The condition grade is a judgment built from observations, and the observations have their own reliability. Both teams recorded the seven defect flags and the blockage estimate independently, so each can be examined separately.

The defect flags do not agree uniformly, and they do not all agree worse than the grade they feed. Cohen's κ by flag runs from 0.472 for headwall damage through 0.511 for cracking, 0.571 for scour at outlet, 0.630 for a missing or broken cover slab, 0.632 for spalling and 0.638 for wall undermining to 0.677 for exposed reinforcement. Four flags are recorded on every asset and those four carry the weight here. Three of them, exposed reinforcement, wall undermining and spalling, agree better than the composite grade at 0.561, and only cracking, at 0.511, agrees worse. That separation follows the mechanism that produces the off-diagonal mass at the middle of the grade scale: exposed reinforcement is a presence call, while cracking asks the rater to distinguish hairline from open, a judgment of degree. The other three flags carry less. A missing or broken cover slab applies only to the 100 covered drains and kerb inlets in the subsample, and headwall damage and scour at outlet only to cross culverts, of which fourteen were double-rated, so their intervals of 0.143 to 0.857 and 0.039 to 0.857 span most of the admissible range and neither ordering nor magnitude can be read from them. The second team failed to record between 28.3 and 30.5 percent of the defects the first recorded across the four flags applying to every asset, consistent with the inconsistency Dirksen et al. (2013) report for visual sewer inspection.

Blockage behaves differently and better. The two teams' estimates correlate at $r = 0.942$ with a Spearman ρ of 0.918, and the mean difference between them is 1.47 percentage points with a standard deviation of 8.08. A quantity estimated as a percentage of a visible cross section is more reproducible than a categorical judgment about a defect, which suggests that where a protocol can ask for a measured proportion instead of a present or absent call, it should.

Taken together these findings define the protocol's operating envelope. It is repeatable, in bands, on assets that can be seen, and its most reproducible single observation is the blockage estimate rather than any structural flag. It is not repeatable on assets that cannot be seen.

4.5 What drives condition

Table 4 reports the ordinal logistic regression and Figure 5 shows the odds ratios with their confidence intervals. The model covers all 949 assets and returns a pseudo R^2 of 0.280.

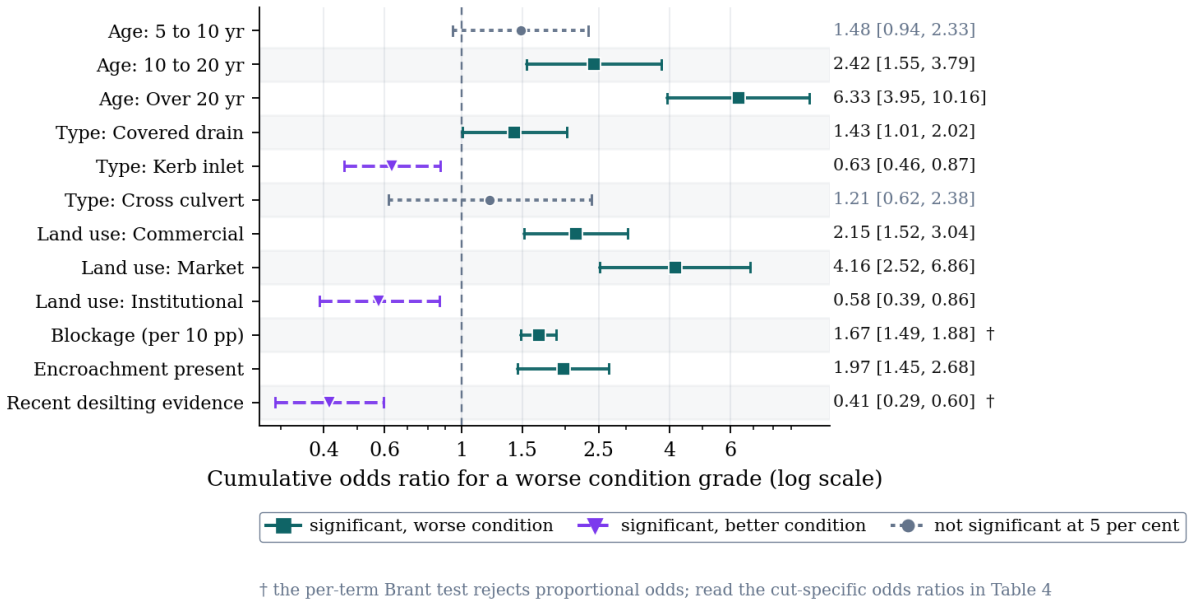
© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

Table 4. Cumulative-logit (proportional odds) model of condition grade, with the cut-specific odds ratios from the four binary logits reported as a partial proportional odds sensitivity analysis.

	Proportional odds model				Cut-specific OR, $P(grade > j)$			
Term	β	SE	OR [95% CI]	p	> 1	> 2	> 3	> 4
Apparent age band (ref. under 5 yr)								
5 to 10 yr	+0.395	0.231	1.48 [0.94,2.33]	0.088	4.05	2.12	1.17	0.57
10 to 20 yr*	+0.884	0.229	2.42 [1.55,3.79]	<0.001	2.61	2.88	1.89	1.91
Over 20 yr*	+1.846	0.241	6.33 [3.95,10.16]	<0.001	30.92	9.34	5.22	3.73
Asset type (ref. open lined channel)								
Covered drain*	+0.355	0.178	1.43 [1.01,2.02]	0.046	4.45	1.86	1.26	0.90
Kerb inlet*	-0.461	0.163	0.63 [0.46,0.87]	0.005	0.84	0.77	0.50	0.56
Cross culvert	+0.192	0.345	1.21 [0.62,2.38]	0.578	0.72	0.84	1.38	1.36
Upstream land use (ref. residential)								
Commercial*	+0.765	0.176	2.15 [1.52,3.04]	<0.001	0.86	2.26	2.44	2.16
Market*	+1.424	0.256	4.16 [2.52,6.86]	<0.001	–	5.19	3.73	3.41
Institutional*	-0.548	0.204	0.58 [0.39,0.86]	0.007	0.40	0.44	0.70	–
Condition covariates								
Blockage (per 10 pp)*†	+0.515	0.059	1.67 [1.49,1.88]	<0.001	0.82	1.36	1.78	2.04
Encroachment present*	+0.679	0.156	1.97 [1.45,2.68]	<0.001	5.48	1.63	2.57	1.38
Recent desilting evidence*†	-0.879	0.185	0.41 [0.29,0.60]	<0.001	0.09	0.28	0.42	0.86
Thresholds τ_1 to τ_4: -2.089, 0.827, 3.245, 5.933								
$n = 949$; LR $\chi^2(12) = 762.1$, $p < 0.001$; McFadden $R^2 = 0.280$; AIC = 1994.7								
Brant omnibus test over the 4 cuts and the 10 terms free of separation: $\chi^2(30) = 52.5$, $p = 0.007$								

Note. OR, cumulative odds ratio for falling in a worse condition grade; an OR above one means the term shifts assets up the five-point scale. * significant at the 5 percent level. † the per-term Brant test rejects proportional odds, so the single pooled OR is an average across the four cut points and understates the effect at the top of the scale; read the cut-specific columns for those terms. Blockage enters per 10 percentage points of cross section obstructed. "Recent desilting evidence" is fresh spoil, a clean invert or recut walls visible from the surface, which is the only maintenance signal a walkover can record. A dash in a cut-specific column means the binary logit at that cut (Land use: Market at grade > 1; Land use: Institutional at grade > 4) is quasi-separated, because the category has no assets on one side of that cut, so no odds ratio is estimable in that cell. $n = 949$ assets.

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674



† the per-term Brant test rejects proportional odds; read the cut-specific odds ratios in Table 4

Figure 5. Cumulative odds ratios for a worse condition grade from the ordinal logistic model, with 95 percent confidence intervals. Values above unity indicate worse condition. The dagger marks terms for which the per-term Brant test rejects proportional odds; read the cut-specific ratios in Table 4 for those.

Age separates most strongly. An asset over twenty years old carries 6.33 times the odds of a worse condition grade than one under five, interval 3.95 to 10.16, and one between ten and twenty years carries 2.42 times the odds. The five to ten year band is not distinguishable from the youngest, at 1.48 with an interval spanning unity.

Land use is the second driver and it acts through refuse. Market frontage carries 4.16 times the odds of a worse grade, interval 2.52 to 6.86, and commercial frontage 2.15. Institutional frontage runs the other way at 0.58, interval 0.39 to 0.86. That ordering is what Aliyu and Suleiman (2016) and Karley (2009) describe qualitatively, and the estimate puts a magnitude on it that a survey of this size can resolve.

Blockage enters at 1.67 per ten percentage points, interval 1.49 to 1.88, encroachment at 1.97, and evidence of recent desilting at 0.41, interval 0.29 to 0.60, which is the only variable in the model an agency directly controls in the short term. Covered drains carry 1.43 times the odds of a worse grade than open channels even after blockage is controlled for, and kerb inlets 0.63.

The proportional odds assumption is rejected, with a Brant statistic of 52.5 on 30 degrees of freedom and $p = 0.007$, and the per-term tests locate the violation in two terms, blockage ($p = 0.001$) and evidence of recent desilting ($p = 0.033$). The cut-specific odds ratios for blockage rise across the scale, from 0.82 at the boundary between grades 1 and 2 to 1.36, then 1.78, then 2.04 at the boundary between grades 4 and 5. That is interpretable rather than inconvenient. Blockage does not distinguish a new asset from a lightly used one, because a new channel with refuse in it is still a new channel; it distinguishes a failing asset from a poor one, because sustained blockage is what converts deterioration into failure. We retain the proportional odds model as a summary of the average effect and report the cut-specific ratios alongside it, with two quasi-separated cells of the partial model reported as not estimable rather than as implausible numbers.

4.6 Spatial pattern and corridor priority

Pooled across corridors, poor condition clusters: the runs test on dichotomized condition returns $z = 5.58$ with $p = 2.4 \times 10^{-8}$ and Moran's I on the ordinal grade along the asset sequence returns $z = 10.30$ with a mean I of 0.222.

The disaggregation is the part worth reporting, and it repeats the lesson of Section 4.3 in a different register.

The six corridors carry comparable asset counts and were walked to the same protocol, so a corridor on which a test fails to reject is as likely to be one the test cannot resolve as one without clustering, which makes a corridor-level result partly a measure of the power of the test. The two tests do not see the same thing. The runs test rejects on three corridors, Gbongan Road ($p = 1.1 \times 10^{-7}$), Murtala Muhammad Road ($p = 0.0008$) and Asa Dam Road ($p = 0.017$), and does not reject on Ilesa Road ($p = 0.854$), Iwo Road ($p = 0.217$) or Ibrahim Taiwo Road ($p = 0.233$). Moran's I on the ordinal grade, run on the same sequences, rejects on five, adding Iwo Road ($p = 7.1 \times 10^{-5}$) and Ibrahim Taiwo Road ($p = 2.2 \times 10^{-6}$), and failing only on Ilesa Road ($p = 0.140$).

What follows is a caution about how such a test is specified and how a null is read. Collapsing an ordinal grade to a poor or not-poor indicator discards what distinguishes a run of grade 4 from a run of grade 5, and on two of six corridors that is the difference between detecting clustering and missing it. And a corridor-level non-rejection is weak evidence of no clustering at these asset counts; only Ilesa Road escapes both tests, one corridor rather than half the network.

The operational consequence is a procedure. Where condition clusters, maintenance organized by reach, treating a contiguous run in one visit, matches the structure of the problem; where it does not, an asset-by-asset list is the appropriate dispatch. Which regime a corridor is in costs nothing to test once an inventory exists, and it should be tested corridor by corridor and on the ordinal grade, because a test run on a dichotomy will return the reach-based answer less often than it should.

Table 5 and Figure 6 give the corridor ranking. Murtala Muhammad Road in Ilorin ranks first at 44.3, followed by Asa Dam Road at 41.7 and Ibrahim Taiwo Road at 38.2, with Ilesa Road in Osogbo at 35.8, Gbongan Road at 29.0 and Iwo Road at 28.6. All three Ilorin corridors rank above all three Osogbo corridors. Figure 6 carries what a ranking cannot, which is how concentrated the problem is. It plots one curve per corridor, cumulative drainage frontage against the priority index, worst first, and the six together account for the 48.31 km of frontage the network carries. Every curve falls steeply over its first few hundred meters and then flattens, so the index separates a short badly-off head from a long moderate tail rather than grading smoothly. Pooled into a single network distribution, the worst tenth of frontage all scores above 58.9 on the index, the worst fifth above 50.2 and the worst third above 44.5. Those cuts are taken on length rather than on asset count, which is why the worst tenth of frontage begins a little below the ninetieth percentile of the per-asset index, 59.5, the line drawn on the figure.

The shape rather than the thresholds is what an agency can use. Because the curve flattens quickly, extending a program from the worst tenth to the worst fifth of frontage buys progressively less, since the index falls only from 58.9 to 50.2 while the length treated doubles. That is a defined point of diminishing return, and it can only be seen by an agency that has surveyed first.

Because the three weights are analyst-set rather than estimated, the ranking is only useful if it survives being weighted differently. Table 6 and Figure 7 test that. Across 5,000 random weight vectors drawn from the simplex, Murtala Muhammad Road holds first place in 99.0 percent of draws and the three Ilorin corridors hold the top three places in 100 percent, so the claim made above about the head of the ranking does not depend on the weighting. The complete six-corridor order is reproduced in 70.5 percent of draws. The instability is confined to one join. Gbongan Road and Iwo Road, at fifth and sixth, reverse in 28.5 percent of draws, which is unsurprising given that they sit 0.4 index points apart, and every other adjacent pair reverses in essentially no draws. The one-at-a-time sweep agrees: varying the consequence weight never changes the order, and the only movement produced by varying the other two is the same fifth and sixth place swap. The ranking supports a decision about where to start and not a distinction between the two least urgent corridors.

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

Table 5. Corridor ranking by the length-weighted priority index, with the components that drive it and the frontage that would enter a first intervention program.

Rank	ID	Corridor (city)	Length (km)	Assets	Mean grade	Mean blockage (%)	Priority index	Frontage > 60 (km)
1	IL-2	Murtala Muhammad Road (Ilorin)	4.20	171	3.89	48.1	44.3	1.38
2	IL-3	Asa Dam Road (Ilorin)	3.80	153	3.61	43.5	41.7	1.00
3	IL-1	Ibrahim Taiwo Road (Ilorin)	3.99	162	3.49	37.4	38.2	0.61
4	OS-2	Ilesa Road (Osogbo)	4.00	154	3.33	33.4	35.8	0.32
5	OS-1	Gbongan Road (Osogbo)	4.20	161	2.86	26.8	29.0	0.44
6	OS-3	Iwo Road (Osogbo)	3.80	148	2.96	23.6	28.6	0.01
		Network	23.99	949	3.37	35.7	37.2	3.76

Note. The priority index runs 0 to 100 and combines condition (weight 0.35), blockage of the cross section (weight 0.25) and consequence of failure (weight 0.4), the last built from the carriageway width affected, whether the asset is a cross culvert, the upstream catchment area and whether encroachment blocks maintenance access. The three weights are analyst-set rather than estimated, so the ranking is what this weighting implies. The index is length-weighted by the frontage each asset represents. Frontage > 60 is the drainage length whose assets score above 60, the working threshold for a first intervention program; the 90th percentile of the per-asset index is 59.5. $n = 949$ assets over 23.99 km.

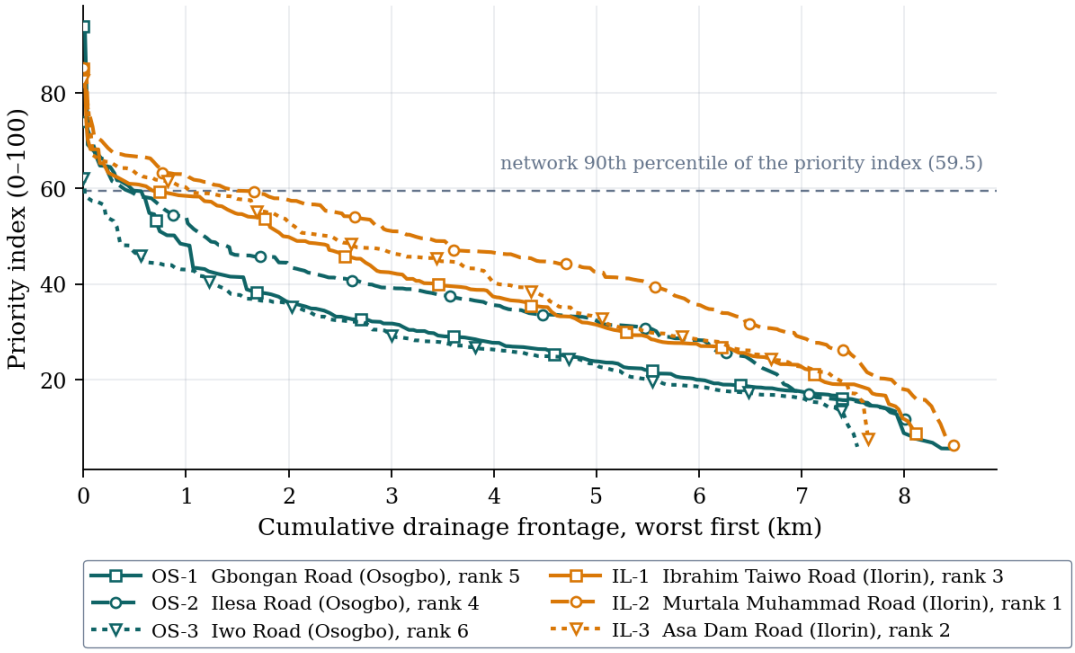


Figure 6. Cumulative drainage frontage against the priority index, worst first, by corridor. The steepness at the left of every curve is what makes a targeted program affordable: the frontage above the ninetieth percentile of the per-asset index, 59.5, is a small share of the total length.

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

Table 6. Stability of the corridor priority ranking when the three index weights are varied, one at a time and at random.

				One at a time		Random weights		
Rank	ID	Corridor	Index	Range	Changes	Modal	P5–P95	Held (%)
1	IL-2	Murtala Muhammad Road	44.3	1	no	1	1–1	99.0
2	IL-3	Asa Dam Road	41.7	2	no	2	2–2	99.0
3	IL-1	Ibrahim Taiwo Road	38.2	3	no	3	3–3	100.0
4	OS-2	Ilesa Road	35.8	4	no	4	4–4	100.0
5	OS-1	Gbongan Road	29.0	5–6	yes	5	5–6	71.4
6	OS-3	Iwo Road	28.6	5–6	yes	6	5–6	71.5
Share of random weight vectors, uniform simplex then each weight held in [0.10,0.70]								
Whole ranking unchanged: 70.5%, 79.7%								
IL-2 still ranked first: 99.0%, 100.0%								
Ilorin corridors still hold the top three: 100.0%, 100.0%								
Reversal of each adjacent pair, same two draws								
IL-2 vs IL-3 (ranks 1 and 2): 1.0%, 0.0%								
IL-3 vs IL-1 (ranks 2 and 3): 0.0%, 0.0%								
IL-1 vs OS-2 (ranks 3 and 4): 0.0%, 0.0%								
OS-2 vs OS-1 (ranks 4 and 5): 0.0%, 0.0%								
OS-1 vs OS-3 (ranks 5 and 6): 28.5%, 20.3%								

Note. Rank 1 is the most urgent corridor. Index is the baseline length-weighted priority index at our weights of 0.35 condition, 0.25 blockage and 0.40 consequence. “One at a time” sweeps a single weight from 0.10 to 0.70 in steps of 0.05 with the other two renormalized in their baseline proportion, 13 steps for each of the three weights; Range is the span of ranks the corridor takes across all 39 sweep points. “Random weights” draws 5000 weight vectors uniformly from the three-component simplex; Held is the share of those draws in which the corridor keeps its baseline rank. The uniform simplex is the conservative reading because it admits weight vectors as lopsided as 0.98 on one component, which no asset manager would choose; the restricted draw holding every weight between 0.10 and 0.70 is the more realistic test and is the second figure in each pair in the lower panel. The three weights are analyst-set rather than estimated, which is why this table exists: the ranking is only worth reporting to the extent that it survives other defensible weightings. Gbongan Road and Iwo Road hold their baseline ranks in slightly different shares, 71.4 against 71.5 percent, because the pair does not only reverse against each other: in a further 0.04 percent of draws, shown as 0.0 percent after rounding in the lower panel, Gbongan Road moves up past Ilesa Road instead.

© JUL 2019 | IRE Journals | Volume 3 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722674

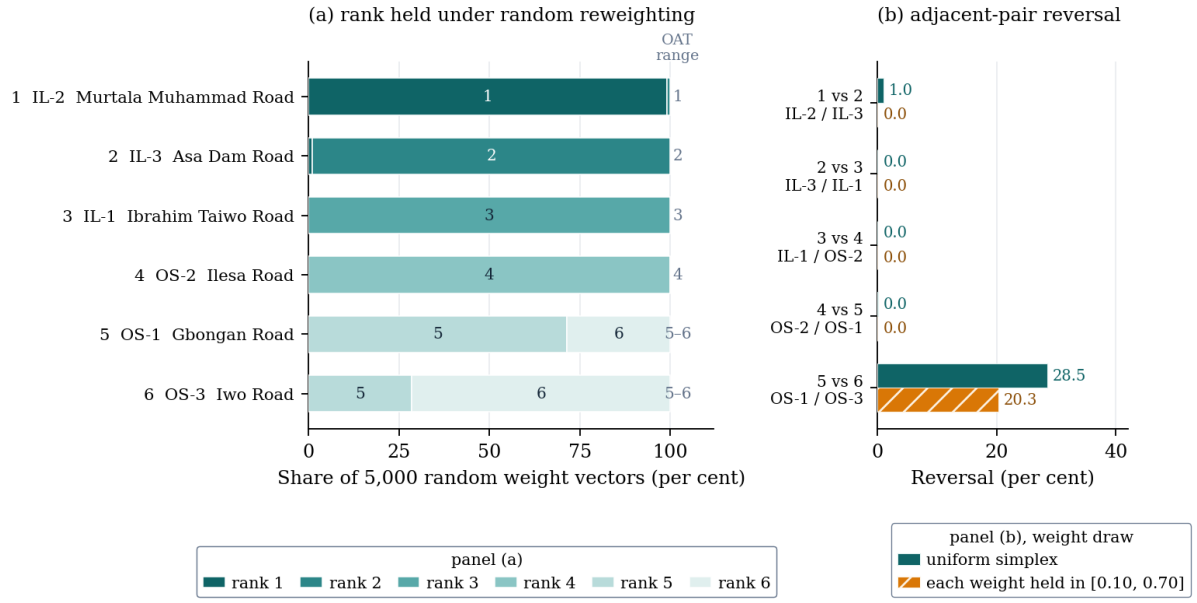


Figure 7. Stability of the corridor priority ranking under reweighting. Panel (a) gives the share of 5,000 random weight vectors in which each corridor takes each rank, with the range from the one-at-a-time sweep annotated. Panel (b) gives the reversal frequency for each adjacent pair. Only the fifth and sixth places are unstable.

V. DISCUSSION

5.1 What the disaggregation shows

What generalizes beyond this network is not the condition of these drains but the way the statistics behaved, and that is what a survey commissioner needs answered before spending money. A pooled agreement statistic can hide a subgroup in which an instrument fails. A statistic disaggregated by city rather than by asset type invites the wrong diagnosis, blaming a team where the answer is an asset mix. And a spatial test can reject on a network while failing on corridors that carry a similar underlying pattern, so a corridor-level null is partly a statement about the test rather than about the corridor, and how many corridors return one depends on whether the test is run on the ordinal grade or on a threshold of it.

Each of those is a property of the arithmetic, and together they are why the disaggregation argued for here is not a refinement but the point. The agreement family of Section 4.3, the pooled κ of 0.561, its weighted forms, the exact and within-one-grade rates, the defect-flag coefficients and the severity asymmetry between teams, is reported in that detail so that a commissioner can see what a survey of this size resolves.

5.2 Implications for agencies

The audience for what follows is whoever commissions a condition survey: the Federal Ministry of Works and Housing, the state ministries in Osun and Kwara, and the local government authorities that hold most of the drainage asset. The recommendations concern how to commission and read such a survey rather than how to maintain a drain, because the survey is the step currently missing.

Buy the reliability check: it costs a fifth of the survey. A condition figure whose repeatability is unknown cannot be distinguished from a rater effect. Rating a fifth of the inventory twice and publishing a chance-corrected statistic alongside the data is the single change that would do most to make Nigerian infrastructure condition data usable, and nothing about it is specific to drainage.

Require the statistic disaggregated, not pooled. This is where a well-intentioned requirement would fail. A pooled κ of 0.561 reads as an acceptable instrument, and the by-city figures of 0.626 and 0.475 invite the conclusion that one team underperformed. Only the breakdown by asset type shows what is happening. A specification asking for an agreement statistic without saying how to cut it will buy a number that conceals the problem it was bought to find.

Use the data in bands. Agreement within one grade is 98.0 percent while exact agreement is 66.7 percent, and the two teams part company by one grade far more often than by two. A maintenance trigger set at grade 4 or worse is supportable on a protocol with that shape. A rule turning on the difference between grade 3 and grade 4 is not, and where such a distinction carries budget consequences, the asset should be rated twice.

Do not grade what cannot be seen. A rater standing on a slab infers an invert from surface evidence, and covered drains are also the class with the worst blockage at 49.3 percent and the worst condition profile at 31 percent grade 5. Internal inspection at sampled access points, a targeted slab-lifting program, or an explicit not-assessed code are all defensible. What is not defensible is recording a surface-derived grade for a covered asset and treating it as equivalent to a grade for an open one.

Test whether condition clusters before choosing a dispatch rule, and test on the grade. Reach-based maintenance suits a corridor where poor condition runs together and wastes effort where it does not. The test costs nothing once an inventory exists and should be run corridor by corridor, on the ordinal grade rather than on a poor or not-poor dichotomy: here the two forms disagreed on two of six corridors and the dichotomized form was the one that missed the pattern. A corridor returning no clustering should be treated as untested rather than as scattered.

One conclusion the results do not support is that Ilorin's drainage is worse maintained than Osogbo's. All three Ilorin corridors rank above all three Osogbo corridors, but the model attributes condition to age, land use and blockage, and the Ilorin corridors carry more market and commercial frontage. Reading a land-use difference as a management difference would be an error of exactly the kind a condition survey is meant to prevent.

5.3 Transferability

The protocol needs a tape, a probe, a smartphone and two people, and the analysis needs open-source statistical software. Nothing in it is specific to Nigeria beyond the choice of descriptors, and the asset it was designed for, an open lined channel receiving municipal solid waste from the frontage beside it under high-intensity rainfall, is ordinary across the tropics.

The finding that agreement collapses on concealed assets transfers further still, because it is a property of visual inspection rather than of drainage. Any condition survey that rates what it cannot see, whether a buried service, a covered chamber or an enclosed structural element, should expect the same pattern and should disaggregate its agreement statistic to detect it.

5.4 Limitations

Two weaknesses bear on how the numbers should be used.

The grade recorded for a covered drain is inferred from its accessible ends rather than observed along the run, and covered drains are 263 of the 949 assets. Everything resting on them, the grade distribution, the covered-drain coefficient and the agreement figure of 0.277, carries that weakness, and Section 4.3 measures it rather than corrects it.

No published benchmark exists for the central statistic. We located no κ, weighted κ or α for any drainage or culvert rating protocol, so the agreement values reported here can only be set beside percentage-agreement figures from bridge inspection (Phares et al. 2001) and defect detection rates (Bogus et al. 2010) rather than against a like-for-like statistic. That absence is itself part of the paper's argument, but it means the numbers cannot be positioned against a field norm.

VI. CONCLUSION

Nigerian cities flood partly because nobody knows the condition of their drainage. This paper addresses the step before maintenance: how to produce a condition record a maintenance decision can rest on. We set out a five-point rating protocol written for open lined channels, covered drains, kerb inlets and cross culverts rather than adapted from a culvert barrel scale, paired it with a measurement design that rates a fifth of the inventory twice and reports agreement disaggregated by asset type, and applied both to an inventory of 949 assets along 23.99 km of arterial alignment in Osogbo and Ilorin.

The protocol proved repeatable within stated limits, and the agreement figures report what two independent teams returned on the same assets. The two teams agreed exactly on 66.7 percent of assets and within one grade on 98.0 percent, giving a Cohen's κ of 0.561, a quadratically weighted κ of 0.841 and a Krippendorff's α of 0.842. Disaggregated, agreement runs from 0.682 on open lined channels to 0.277 on covered drains, and the blockage estimate, a measured proportion rather than a categorical judgment, is the most reproducible observation in the protocol at $r = 0.942$ between teams.

Three lessons generalize beyond this study and beyond drainage, because they concern what statistics do rather than what Nigerian drains are like. A pooled agreement figure can read as acceptable while concealing a subgroup in which the instrument has failed. Cutting that figure by city rather than by asset type produces a plausible and wrong diagnosis that blames a survey team for an asset mix. And a spatial test can reject across a network while failing on corridors carrying the identical pattern, so a corridor-level null measures the power of the test and not the corridor.

Five recommendations follow from those three: commission the reliability check, specify the statistic disaggregated, use the resulting grades in bands and not at a point, code covered assets as unassessed rather than grading them through a slab, and test corridor by corridor on the ordinal grade whether condition clusters before choosing between reach-based and asset-based dispatch.

These condition results belong to six corridors in two cities. The most useful thing the next such survey can produce is not its condition distribution but its agreement statistic, because that is the number that tells the agency paying for it how much the rest of the data is worth.

REFERENCES

[1] ActionAid International. 2006. *Unjust Waters: Climate Change, Flooding and the Protection of Poor Urban Communities: Experiences from Six African Cities*. ActionAid International, International Emergencies and Conflict Team. https://actionaid.org/sites/default/files/unjust_waters_-_climate_change_flooding_and_the_protection_of_poor_urban_communities_experiences_from_6_african_cities.pdf.

[2] Adelekan, Ibidun O. 2010. "Vulnerability of Poor Urban Coastal Communities to Flooding in Lagos, Nigeria." *Environment and Urbanization* 22 (2): 433–50. https://doi.org/10.1177/0956247810380141.

[3] Aderogba, Kunle A., M. Oredipe, S. Oderinde, and T. Afelumo. 2012. "Challenges of Poor Drainage Systems and Floods in Lagos Metropolis, Nigeria." *International Journal of Social Sciences and Education* 2 (1): 413–34.

[4] Agbola, Babatunde S., Olalekan Ajayi, Omolara J. Taiwo, and Bashir W. Wahab. 2012. "The August 2011 Flood in Ibadan, Nigeria: Anthropogenic Causes and Consequences." *International Journal of Disaster Risk Science* 3: 207–17. https://doi.org/10.1007/s13753-012-0021-3.

[5] Agbonkhese, Onoyan-Usina, Godwin Lazhi Yisa, and Paul Itomi-ushi Daudu. 2013. "Bad Drainage and Its Effects on Road Pavement Conditions in Nigeria." *Civil and Environmental Research* 3 (10): 7–15.

[6] Aliyu, Halima I., and Zaharadeen Suleiman. 2016. "Flood Menace in Kaduna Metropolis: Impacts, Remedial and Management Strategies." *Science World Journal* 11 (2): 16–22.

[7] Arnoult, James D. 1986. *Culvert Inspection Manual: Supplement to the Bridge Inspector's Training Manual*. FHWA-IP-86-2. Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/68046.

[8] Ayoade, John O., and F. O. Akintola. 1980. "Public Perception of Flood Hazard in Two Nigerian Cities." *Environment International* 4 (4): 277–80. https://doi.org/10.1016/0160-4120(80)90079-3.

[9] Bianchini, Andrea, Paola Bandini, and David W. Smith. 2010. "Interrater Reliability of Manual Pavement Distress Evaluations." *Journal of Transportation Engineering* 136 (2): 165–72. https://doi.org/10.1061/(ASCE)0733-947X(2010)136:2(165).

[10] Bogus, Susan M., Giovanni C. Migliaccio, and Arturo A. Cordova. 2010. "Assessment of Data

Quality for Evaluations of Manual Pavement Distress." *Transportation Research Record* 2170. https://doi.org/10.3141/2170-01.

[11] Brant, Rollin. 1990. "Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression." *Biometrics* 46 (4): 1171–78. https://doi.org/10.2307/2532457.

[12] Cahoon, Joel E., Duane Baker, and Jodi Carson. 2002. "Factors for Rating Condition of Culverts for Repair or Replacement Needs." *Transportation Research Record* 1814 (1): 197–202. https://doi.org/10.3141/1814-23.

[13] Chughtai, Faisal, and Tarek Zayed. 2008. "Infrastructure Condition Prediction Models for Sustainable Sewer Pipelines." *Journal of Performance of Constructed Facilities* 22 (5): 333–41. https://doi.org/10.1061/(ASCE)0887-3828(2008)22:5(333).

[14] Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1): 37–46. https://doi.org/10.1177/001316446002000104.

[15] Cohen, Jacob. 1968. "Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit." *Psychological Bulletin* 70 (4): 213–20. https://doi.org/10.1037/h0026256.

[16] Dirksen, J., F. H. L. R. Clemens, H. Korving, et al. 2013. "The Consistency of Visual Sewer Inspection Data." *Structure and Infrastructure Engineering* 9 (3): 214–28. https://doi.org/10.1080/15732479.2010.541265.

[17] Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman and Hall.

[18] Estes, Allen C., and Dan M. Frangopol. 2003. "Updating Bridge Reliability Based on Bridge Management Systems Visual Inspection Results." *Journal of Bridge Engineering* 8 (6): 374–82. https://doi.org/10.1061/(ASCE)1084-0702(2003)8:6(374).

[19] Etuonovbe, Antonia K. 2011. "The Devastating Effect of Flooding in Nigeria." *FIG Working Week 2011: Bridging the Gap Between Cultures* (Marrakech, Morocco).

[20] Hadipriono, Fabian C., Richard E. Larew, and Oh-Young Lee. 1988. "Service Life Assessment of Concrete Pipe Culverts." *Journal of Transportation Engineering* 114 (2): 209–21. https://doi.org/10.1061/(ASCE)0733-947X(1988)114:2(209).

[21] Hayes, Andrew F., and Klaus Krippendorff. 2007. "Answering the Call for a Standard Reliability Measure for Coding Data." *Communication Methods and Measures* 1 (1): 77–89. https://doi.org/10.1080/19312450709336664.

[22] Karley, Noah Kofi. 2009. "Flooding and Physical Planning in Urban Areas in West Africa: Situational Analysis of Accra, Ghana." *Theoretical and Empirical Researches in Urban Management* 4 (13): 25–41.

[23] Landis, J. Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1): 159–74. https://doi.org/10.2307/2529310.

[24] Madanat, Samer, Rabi Mishalani, and Wan Hashim Wan Ibrahim. 1995. "Estimation of Infrastructure Transition Probabilities from Condition Rating Data." *Journal of Infrastructure Systems* 1 (2): 120–25. https://doi.org/10.1061/(ASCE)1076-0342(1995)1:2(120).

[25] Masada, Teruhisa, Shad M. Sargand, Bashar Tarawneh, Gayle F. Mitchell, and Doug Gruver. 2006. "New Inspection and Risk Assessment Methods for Metal Highway Culverts in Ohio." *Transportation Research Record* 1976: 141–48. https://doi.org/10.1177/0361198106197600115.

[26] McCullagh, Peter. 1980. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society: Series B* 42 (2): 109–42. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x.

[27] McHugh, Mary L. 2012. "Interrater Reliability: The Kappa Statistic." *Biochemia Medica* 22 (3): 276–82.

[28] Micevski, Tom, George Kuczera, and Peter Coombes. 2002. "Markov Model for Storm Water Pipe Deterioration." *Journal of Infrastructure Systems* 8 (2): 49–56. https://doi.org/10.1061/(ASCE)1076-0342(2002)8:2(49).

[29] Mishalani, Rabi G., and Samer M. Madanat. 2002. "Computation of Infrastructure Transition Probabilities Using Stochastic Duration Models." *Journal of Infrastructure Systems* 8 (4):

139–48. https://doi.org/10.1061/(ASCE)1076-0342(2002)8:4(139).

[30] Mitchell, Gayle F., Teruhisa Masada, Shad M. Sargand, and Bashar Tarawneh. 2005. *Risk Assessment and Update of Inspection Procedures for Culverts*. Ohio Department of Transportation / Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/37981.

[31] Mohan, Prashanth, Venkata N. Padmanabhan, and Ramachandran Ramjee. 2008. "Nericell: Rich Monitoring of Road and Traffic Conditions Using Mobile Smartphones." *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems (SenSys '08)*, 323–36. https://doi.org/10.1145/1460412.1460444.

[32] Moran, P. A. P. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37 (1/2): 17–23. https://doi.org/10.2307/2332142.

[33] New York State Department of Transportation. 2006. *Culvert Inspection Field Guide*. New York State Department of Transportation, Office of Transportation Maintenance. https://www.dot.ny.gov/divisions/operating/oom/transportation-maintenance/repository/CULVERT%20INSPECTION%20FIELD%20GUIDE%201-18-06.pdf.

[34] Oguntala, A. B., and J. S. Oguntoyinbo. 1982. "Urban Flooding in Ibadan: A Diagnosis of the Problem." *Urban Ecology* 7 (1): 39–46. https://doi.org/10.1016/0304-4009(82)90004-3.

[35] Olaniran, Olusegun J. 1983. "Flood Generating Mechanisms at Ilorin, Nigeria." *GeoJournal* 7 (3): 271–77. https://doi.org/10.1007/BF00209065.

[36] Phares, Brent M., Benjamin A. Graybeal, Dennis D. Rolander, Mark E. Moore, and Glenn A. Washer. 2001. "Reliability and Accuracy of Routine Inspection of Highway Bridges." *Transportation Research Record* 1749. https://doi.org/10.3141/1749-13.

[37] Phares, Brent M., Glenn A. Washer, Dennis D. Rolander, Benjamin A. Graybeal, and Mark Moore. 2004. "Routine Highway Bridge Inspection Condition Documentation Accuracy and Reliability." *Journal of Bridge Engineering* 9 (4): 403–13. https://doi.org/10.1061/(ASCE)1084-0702(2004)9:4(403).

[38] Sairam, Nivedita, Sudhagar Nagarajan, and Scott Ornitz. 2016. "Development of Mobile Mapping System for 3D Road Asset Inventory." *Sensors* 16 (3): 367. https://doi.org/10.3390/s16030367.

[39] Salem, Osama, Baris Salman, and Mohammad Najafi. 2012. "Culvert Asset Management Practices and Deterioration Modeling." *Transportation Research Record* 2285: 10–18. https://doi.org/10.3141/2285-02.

[40] Santani, Darshan, Jidraph Njuguna, Tierra Bills, Aisha W. Bryant, Reginald Bryant, et al. 2015. "CommuniSense: Crowdsourcing Road Hazards in Nairobi." *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*, 445–56. https://doi.org/10.1145/2785830.2785837.

[41] Sub-Saharan Africa Transport Policy Program. 2009. *RONET: Road Network Evaluation Tools, Version 2.0 – User's Guide*. SSATP Working Paper No. 89A. SSATP / World Bank.

[42] Transportation Research Board. 2015. *Service Life of Culverts*. NCHRP Synthesis 474. National Academies Press. https://doi.org/10.17226/22140.

[43] Wald, Abraham, and Jacob Wolfowitz. 1940. "On a Test Whether Two Samples Are from the Same Population." *The Annals of Mathematical Statistics* 11 (2): 147–62. https://doi.org/10.1214/aoms/1177731909.

[44] Water Research Centre. 2004. *Manual of Sewer Condition Classification*. 4th ed. WRc plc.

[45] Wells, Timothy, and Robert E. Melchers. 2014. "An Observation-Based Model for Corrosion of Concrete Sewers Under Aggressive Conditions." *Cement and Concrete Research* 61–62: 1–10. https://doi.org/10.1016/j.cemconres.2014.03.004.