Viability of bridge inspectors determining defect ratings using photographic images

Nepomuceno, David T.; Vardanega, Paul J.; Pregnolato, Maria; Tryfonas, Theo; Bennetts, John

DOI
10.1680/jinam.23.00004

Publication date
2023

Document Version
Final published version

Published in
Infrastructure Asset Management

Citation (APA)
Nepomuceno, D. T., Vardanega, P. J., Pregnolato, M., Tryfonas, T., & Bennetts, J. (2023). Viability of bridge inspectors determining defect ratings using photographic images. *Infrastructure Asset Management*, *10*(3), 144-156. Article 2300004. https://doi.org/10.1680/jinam.23.00004

Cite this article
Nepomuceno DT, Vardanega PJ, Pregnolato M, Tryfonas T and Bennetts J
Viability of bridge inspectors determining defect ratings using photographic images.
Infrastructure Asset Management,
https://doi.org/10.1680/jinam.23.00004

Research Article
Paper 2300004
Received 23/01/2023; Accepted 13/06/2023
First published online 16/06/2023

Infrastructure Asset Management

emerald PUBLISHING ice Publishing

Viability of bridge inspectors determining defect ratings using photographic images

David T. Nepomuceno MEng, PhD (Bristol)
Formerly PhD student, Department of Civil Engineering, University of Bristol, Bristol, UK (Orcid:0000-0001-5612-1176)

Paul J. Vardanega BE, MEngSc (QldUT), PhD (Cantab), GMICE, MASCE, MIEAust, FHEA
Associate Professor in Civil Engineering, Department of Civil Engineering, University of Bristol, Bristol, UK (Orcid:0000-0001-7177-7851)
(corresponding author: p.j.vardanega@bristol.ac.uk)

Maria Pregnolato MSc, PhD (Newc), FHEA
Associate Professor, Department of Hydraulic Engineering, TU Delft, Delft, the Netherlands; Honorary Associate Professor, Department of Civil Engineering, University of Bristol, Bristol, UK (Orcid:0000-0003-0796-9618)

Theo Tryfonas PhD (Athens), FRSA
Professor of Infrastructure Systems and Urban Innovation, Department of Civil Engineering, University of Bristol, Bristol, UK (Orcid:0000-0003-4024-8003)

John Bennetts MA (Cantab), MSc (Sur), EngD (Bristol), CEng, MICE
Associate, WSP, Bristol, UK

The visual inspection of bridges is a major undertaking for asset owners and operators. In the UK, visual inspections require inspectors to visit bridges on site and often at night and under unfavourable weather conditions. Therefore, it would be beneficial to move some of the visual inspection process off site. This paper studies whether the defect classification aspects of the inspection process could be conducted remotely using photographs. This study examines the defect ratings assigned by ten survey participants who were tasked with examining photographs from visual inspections of ten UK bridges. The survey results were compared with the results from the general inspections previously carried out for the bridges in question. From this data set, the differences in the ratings given and the extent to which defects are missed were examined. The results show that a higher number of defects were identified for a given bridge by the remote inspectors. Statistical analysis shows that aggregated defects rated by off-site inspectors tend to be more severe and of a higher priority rating compared with those from the on-site inspectors. The results also indicate that there is closer agreement between on-site and off-site inspectors for defects of a higher severity rating.

Keywords: bridges/monitoring/service life

Notation

N	number of data points
p	p-value from a statistical test
μ	mean
σ	standard deviation
σ^2	variance

Introduction

Bridge condition

Bridges are crucial national infrastructure assets. As they age, appropriate routine maintenance is necessary to ensure their continued serviceability and safety. In the UK, of the approximately 8000 bridges on the English Strategic Road Network, about 60% are approaching 50 years in service (Mahut and Woodward, 2005). Maintaining a bridge stock in the most efficient and cost-effective way is a significant challenge for asset owners.

For infrastructure-asset-owning organisations, visual inspection continues to be the most prevalent method of assessing the condition of bridges (e.g. Bennetts, 2019; Bennetts *et al*., 2016, 2020; Lea and Middleton, 2002; Phares *et al*., 2004). However, studies have shown that data gathered from visual inspections are highly subjective and that quality can be largely dependent on the inspector's prior experience, as well as their knowledge of the structural behaviour of the relevant bridge (e.g. Bennetts *et al*., 2016, 2018a, 2020; Lea and Middleton, 2002; Moore *et al*., 2001; Phares *et al*., 2004). Visual inspection procedures can be time consuming and unsafe for inspectors, particularly for large structures located in difficult terrain. Bhreasail *et al*. (2019) presented a review focusing on how remote sensing can assist with geotechnical asset management for highways. Nepomuceno *et al*. (2022a) presented a recent review of future technologies that may be considered for visual inspection of bridge assets.

In the context of bridge management, McRobbie *et al*. (2007) found that it was possible to assess a bridge visually using only captured images. The findings of this image-based assessment of bridge condition were comparable with those of the on-site inspection. Subsequently, this research was expanded into a series of initiatives investigating the automation of highway infrastructure inspection (McRobbie, 2009; McRobbie *et al*., 2007, 2015). The work in this paper follows on directly from a pilot study in which a basic comparison of results derived from an on-site and off-site assessment of a small group of defects was conducted (Nepomuceno *et al*., 2021). The pilot study (Nepomuceno *et al*., 2021) also helped shape the study protocol presented in this paper. Some other parts of the collected survey data were recently published in the paper by Nepomuceno *et al*. (2022b), showing how aspects of the visual inspection process may be incorporated into the digital work environment. This paper addresses an important research need by examining the feasibility and effectiveness of conducting defect classification remotely using photographs, thereby investigating an approach that has the potential to enhance the efficiency, accuracy and prioritisation of bridge maintenance strategies.

Study aims and objectives

This paper focuses on examining how feasible it is for an off-site inspector to assess a bridge for defects using two-dimensional (2D) photographs alone. The objective of this paper is to present a proof of concept showing that some aspects of the visual inspection workflow can be conducted off site without significant loss of data accuracy. Central to this work is a survey wherein ten participants were tasked with examining photographs taken during on-site inspections and asked to identify and grade the defects observed. The results were then compared with the assigned ratings from the original on-site inspections, allowing an analysis of the differences between the two approaches. This research considers the human component of defect identification and grading, assessing the differences between results obtained from physical and remote inspections. The results offer an indication of how reliable this method of defect assessment can be and what (if any) implications more widespread implementation would have on the visual inspection process, with a view towards greater automation by, for example, introducing pattern recognition/machine learning algorithms at a later stage. Additionally, the findings would illustrate the proficiency with which human inspectors can assess defects remotely. If it transpires that a human is incapable of performing this task, the promised revolution of machines rating defects will be extremely difficult to achieve. Finally, while the study compares both traditional and remote procedures, it is beyond the scope of this research to make a judgement on which is superior.

Visual inspection of bridges

The current industry standard for visual bridge inspection in the UK involves a cycle of routine inspections, comprising general inspections (GIs) and principal inspections (PIs) (see HE, 2021). Bennetts *et al*. (2018a, 2018b, 2022) and Bennetts (2019) presented recent works on the use of visual inspection data to study trends at the stock/regional level. The work in this paper primarily focuses on implications for the GI procedure.

Limitations of visual inspections

Wallbank (1989: p. 17) noted that '[w]hen recording and comparing the visual condition of a wide variety of bridges it is difficult to be precise and consistent'. It can be argued that compared with other aspects of bridge-monitoring research, the reliability of visual inspections has received less attention. However, multiple studies (i.e. Bennetts, 2019; Lea and Middleton, 2002; Middleton, 2004; Moore *et al*., 2001; Phares *et al*., 2004) detailed progress in this research space, in part highlighting the significance of the human variables that influence the reliability of visual inspection.

Study motivation

A case can be made to enhance the routine inspection process with off-site defect assessment. Viewed from a human resource standpoint, the division of labour can be optimised. A system that allows image capture to be conducted by highly competent photographers, thus ensuring that higher-quality images are produced, is a potential alternative. These images can then be provided to qualified inspectors in an office setting to identify and rate defects. This modification to the procedure would involve the use of readily available technology such as high-resolution cameras, making this regime more readily adoptable in industry practice and more easily implemented in day-to-day operations. Demonstrating that structures can be adequately inspected remotely in this manner may also instil confidence to enable future innovation to replace or augment the image capture, defect identification or defect grading operations.

Study protocol

The study presented in this paper comprised an online trial designed to compare the defect ratings received from an on-site GI against those from a remote inspection for a given bridge. The trial was designed and led by the first author (see the thesis by Nepomuceno (2022) for further details on the study methodology). As also explained in the paper by Nepomuceno *et al*. (2022b), the third author of this paper acted as an academic representative in the survey to assist with benchmarking, whose results were aggregated with the other results in the paper. This section briefly outlines the study protocol and format of the trial.

Trial overview

The trial comprised three parts, which were to be completed in succession:

- a stakeholder questionnaire survey (questionnaire A)
- a remote inspection survey (main survey)
- an evaluation questionnaire survey (questionnaire B).

The trial was designed to be completed in approximately 1 h. This section outlines the development of each of these three trial components. Participants of the trial were key stakeholders in the bridge-management field. Out of the 19 individuals that were initially approached, full results from ten participants were received and analysed in this paper.

The main survey was specifically designed and developed to compare the results of an on-site GI and a remote inspection of a specific bridge structure. In 2018–2019, GIs were conducted on ten bridges that are part of the highway network in the south-west of England. For this paper, the inspectors who carried out the original inspections are referred to as 'group A'. The bridge ratings obtained by these inspectors are referred to as 'set A'. As is standard procedure, during these inspections, inspectors also took photographs of the bridge and subsequently stored them on their organisational network. These photographs formed the basis of this trial. The remote inspection survey presents the digital photographs taken by group A to the participants of the study, who were then tasked with observing the photographs for defects, thus 'remotely' inspecting a bridge. For brevity, the participants of the study are collectively referred to as 'group B' in this paper. The subsequent defect ratings from these remote inspections are termed 'set B'. These key terms are summarised in Table 1. An overall schematic diagram of the trial is shown in Figure 1.

Questionnaires

Questionnaire A was developed to obtain information on the profile of participants. This included key questions on their perceived stakeholder role in the bridge-management process and their level of bridge inspection experience. Questionnaire B aimed to obtain the participant's perception of the main survey and additional evaluation of the process. In addition, both questionnaires allowed the participant to give comments for any of the questions. These comments are included in parts of the discussion in the section headed 'Discussion'.

Material presented to participants

To complete the trial, each participant was presented with the following:

- photographs of the bridge (taken by inspectors in group A), each labelled with a unique identification number
- for structures of asset owner 1 (AO1) (see the section headed 'Structure profiles'), forms containing background structural information; the relevant forms for structures under asset owner 2 (AO2) were not available for this trial
- a spreadsheet to input their observations
- a step-by-step guide on how to complete the survey, as well as a reference for defect types and severity degrees
- two questionnaires given as Google Forms.

Table 1. Definition of key terms

Terms	Definition
Group A	On-site bridge inspectors conducting GIs
Group B	Participants of this study who inspected the bridge using photographs
Set A	Defect ratings obtained from on-site inspections
Set B	Defect ratings obtained from remote inspection

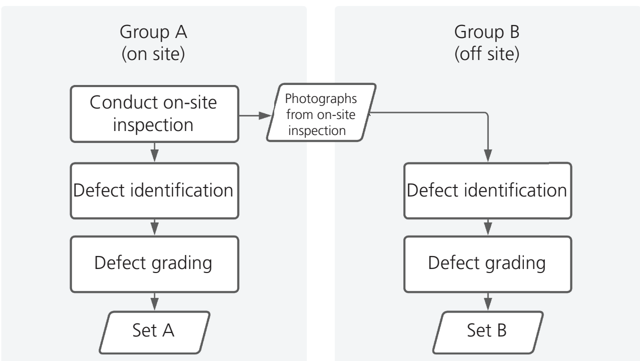

Figure 1. High-level schematic diagram of the trial procedure showing how photographs taken from the on-site inspections are used by group B to identify and grade defects. The resulting set A and set B are compared in this paper

Participants were sent a OneDrive link containing the aforementioned materials, which they could access at any point in time. Using the provided spreadsheet, participants were required to record the following defect attributes:

- photograph ID: filename of photograph
- component: the bridge element on which the defect was located
- type: defect type
- severity: the resulting amount of damage or loss of functionality to a component; this is measured on a scale of 1–5 (HA, 2007)
- extent: a measure of how widespread a defect is on a component; this is measured on a scale of A–E (HA, 2007)
- priority: estimated priority of defect repair
- in another photograph?: indicate whether the defect is seen in another photograph (yes or no)
- relevant photograph ID: if yes to above, record photograph IDs
- comments: any additional information.

Each cell was pre-populated for ease of input. If a participant had zero confidence in recording an attribute, an option was available to indicate 'insufficient data available'.

Participant profiles

All ten participants at the time of the survey worked in the UK for various organisations, including local authorities, independent consultancies and universities. Respondents identified themselves as holding various professional roles from graduate engineer to inspector to asset owner (see Table 2). When asked which bridge-management stakeholder roles they could best identify with, half identified themselves as 'inspectors' and/or 'engineers'. Three participants identified as 'asset owners', and one participant identified as a 'researcher'. Seventy per cent of participants had at least 4 years of inspection experience, with 4–6 years being the most indicated experience level (30%). Sixty per cent had undertaken an inspection in the 12 months prior to completing the trial. One participant had never undertaken an official GI before. Furthermore, only one participant had taken part in the Bridge Inspector Certification Scheme (BICS) (Lantra, 2021), achieving a 'Senior Inspector' level.

Structure profiles

For this study, the on-site defect data were taken from GIs of two groups of structures: (*a*) highway bridges in the south-west of

Table 2. Participant profiles

Participant reference	Stakeholder role(s)	Length of inspection experience: years	Time since most recent inspection
P1	Asset owner	10–14	Over 5 years ago
P2	Bridge inspector	4–6	In the past 12 months
P3	Bridge inspector, bridge engineer, asset owner	Over 20	In the past 12 months
P4	Assistant engineer	1–3	In the past 12 months
P5	Academic	Less than 1	Never
P6	Bridge engineer	4–6	In the past 3 years
P7	Bridge inspector	7–9	In the past 12 months
P8	Bridge inspector, bridge engineer	4–6	In the past 12 months
P9	Bridge inspector, bridge engineer	7–9	In the past 12 months
P10	Asset owner	Less than 1	Over 5 years ago

England and (*b*) bridges in a county in the east of England (see also the pilot study reported by Nepomuceno *et al.* (2021)). In each of these instances, two different asset owners are represented, which are referred to as AO1 and AO2, respectively. Each participant was randomly assigned a structure for the trial, with the only condition being that they had not taken part in the GI themselves. The ten structures subsequently included in the study can be found in Table 3. The majority of the structures (80%) were owned by AO1, with the remainder being owned by AO2. The structures range in length from a 10 m culvert to an 82 m highway overbridge. The structures chosen for this study were dictated by the GI information made available at the time of the authors' selection. For a future study, it was suggested that structures be selected more deliberately to represent a range of construction types. This selection might involve including a single type of structure or a wider variety of structures.

In the UK, inspectors generally take photographs of a structure and specific defects during an on-site inspection. A subset of these photographs is then used in the final GI report to the asset owner. The quantity of photographs taken from the most recent GI of each structure can also be seen in Table 3. It is worth noting that these inspection photographs had been taken prior to the inception of this study and so were not influenced by the study protocol implemented in this work. Finally, the number in the structure reference is numerically equivalent to the participant that remotely inspected it (e.g. participant P3 inspected structure S3).

Results

This section presents the data in both set A and set B, outlining the main observations found in each data set. The results were then compared to check for any differences in the ratings given.

Set A (on-site inspection data)

In total, 196 defects were recorded from the on-site inspections of the ten structures. See Table 4 for the definitions of the statistical measures used in this paper. Figure 2(a) shows the distribution of the severity and extent recorded. It should be noted that score progression from A to E indicates an increasing extent and score progression from 1 to 5 indicates an increasing severity (see e.g. the thesis by Bennetts (2019) for more detailed information on defect extent and severity classification in the UK visual inspection process). From this figure, it is observed that 2B is the most common rating, accounting for 33% (N = 65) of the defects. When considering severity ratings only, 67% of defects were rated 2.

Table 4. Descriptive statistics for severity for set A and set B

Statistic	Set A	Set B
Number of data points, N	196	199
Mean, μ	2.19	2.55
Mode	2.0	2.0
Median	2.0	2.0
Standard deviation, σ	0.72	0.74
Variance, σ^2	0.51	0.55

Table 3. Structure profiles

Structure reference	Asset owner	Structure description	Length: m	Number of spans	Number of photographs
S1	AO1	Highway overbridge (foot)	76	3	54
S2	AO1	Highway overbridge	76	4	95
S3	AO1	Highway overbridge	74	3	72
S4	AO1	Highway overbridge	82	3	87
S5	AO1	Culvert	10	2	50
S6	AO1	Highway underbridge	59	3	145
S7	AO2	Culvert	10	1	65
S8	AO2	Highway overbridge	30	1	83
S9	AO1	Highway overbridge	60	2	101
S10	AO1	Highway underbridge	28	1	76

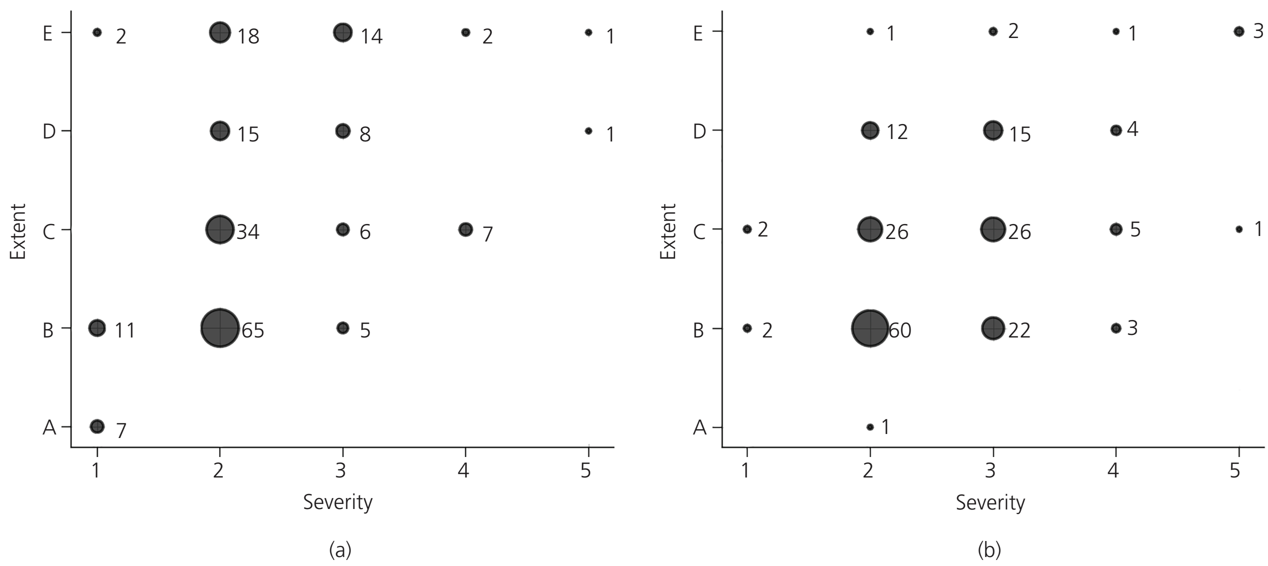

Figure 2. (a) Distribution of severity and extent ratings for set A. Note: score progression from A to E indicates increasing extent, and score progression from 1 to 5 indicates increasing severity. An increasing marker size indicates an increased number of observations corresponding to that point on the graph. (b) Distribution of severity and extent ratings for set B. Note: score progression from A to E indicates increasing extent, and score progression from 1 to 5 indicates increasing severity. An increasing marker size indicates an increased number of observations corresponding to that point on the graph

When analysing extent, an equivalent numerical rating was assumed where A = 1 and E = 5. Taking the averages of the severity and extent ratings separately, values of 2.19 and 3.02 are obtained, respectively (see Table 4). Figure 3 shows the frequency of each defect class for both set A and set B. The five most common defects in set A are (*a*) corrosion (24%; N = 40); (*b*) cracks (19%; N = 32); (*c*) concrete defects (16%; N = 27); (*d*) vegetation/maintenance (15%; N = 25); and (*e*) water (10%; N = 16).

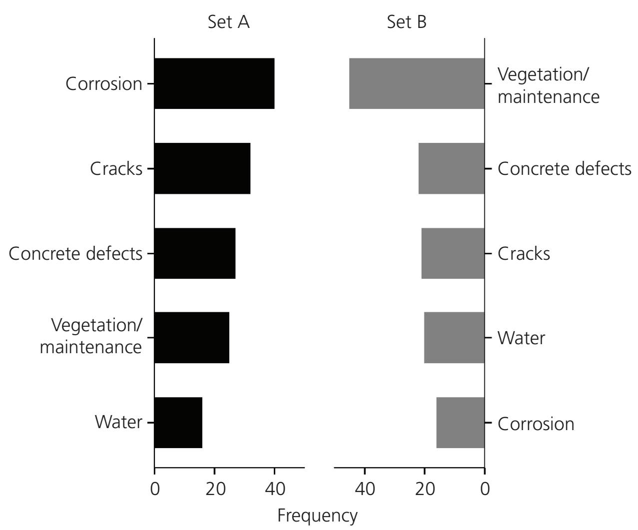

Figure 3. Five most frequent defect classes recorded for each data set

Set B (off-site inspection data)

In total, 206 defects were recorded from the remote inspections of the ten structures. Only 186 (out of 206) defects could be confidently classified, where both a severity and an extent rating were assigned by a remote inspector. The severity and extent distribution of these 186 defects is shown in Figure 2(b). Similar to set A, the most common rating recorded was 2B, making up 32% (N = 60) of the ratings. However, it is notable that there is a higher proportion of defects rated 3 and 4 in severity. Compared to set A, a higher average severity of 2.56 is observed, while a lower average extent of 2.76 is noted. This value seems to indicate that on average, set B inspectors rated defects more severely compared with set A inspectors. This result is further examined in the section headed ‘Severity’. Surveying the most frequent defect classes recorded, the same five classes from set A are also observed, albeit in a different order: (*a*) vegetation/maintenance (25%; N = 45); (*b*) concrete defects (12%; N = 22); (*c*) cracks (12%; N = 21); (*d*) water (11%; N = 20); and (*e*) corrosion (9%; N = 16).

Proportion of defects recorded by both groups

Table 5 shows the percentages of unique defects recorded by each group. This analysis was conducted by listing the unique defects recorded by both on-site and off-site groups and calculating the proportion found by each group. Table 5 shows that the on-site inspectors (group A) found 196/249 (79%) of the total collective unique defects and that the off-site inspectors (group B) found 206/249 (83%) of the total collective unique defects. It was observed that a slightly higher percentage of unique defects were recorded by the off-site inspectors. This result may be explained

Table 5. Percentages of unique defects recorded by each group

Structure	Group A	Group B	Defects in A, indicated by group B	Collective unique defects	AB agreement: %
S1	16	20	13 (81%)	23	57
S2	30	22	20 (67%)	32	63
S3	16	28	15 (94%)	29	52
S4	24	36	20 (83%)	40	50
S5	2	7	2 (100%)	7	29
S6	11	14	11 (100%)	14	79
S7	13	8	5 (38%)	16	31
S8	17	13	6 (35%)	20	30
S9	41	40	40 (98%)	41	98
S10	26	18	17 (65%)	27	63

by the off-site inspectors erring on the ‘safe’ side and pointing out a somewhat minor defect ‘just in case’ it becomes a problem. This outcome is a positive result, as it could be said that by recording a higher volume of defects, the likelihood of discovering a severe defect would increase.

Attribute comparisons

This section compares set A and set B for the defect attributes of severity, extent, defect class and priority. To clarify, the defects studied in this section relate to the collective unique defects shown in Table 5, to investigate any variation in the data. Therefore, comparisons are made without using those data from set A or set B as a benchmark for ‘true’ identifications and ratings. This may be investigated in the future.

Severity

Considering the severity ratings from both data sets only, it is observed that a rating of 2 was the most frequent for both set A and set B (see Figure 4(b)). As previously noted, higher frequencies of defects rated 3 and above are seen in set B, which suggests that group B inspectors tend to rate defects more severely. To quantify this rate, the Mann–Whitney *U*-test can be used to compare each data set statistically. This analysis is a rank-based test for comparing the values of two independent groups which does not require normality (Dodge, 2008; Mann and Whitney, 1947). It is appropriate for both continuous and ordinal data. A significant result indicates that the values of the two groups are distinct. A recent example of the test being used in a civil engineering context can be found in the paper by Huang *et al*. (2021).

A p-value can be calculated to measure this statistical difference, which can then be compared with a significance level of 0.05. The p-values in this study were computed using the SciPy package in the Python programming language (SciPy, 2021). For the severity ratings of set A and set B, this yielded a p-value of 5.07×10^{-8}. Given that the p-value is significantly less than the 0.05 level of significance, it is highly likely that the higher values acquired by off-site inspectors are statistically significant. This result supports the suggestion that off-site inspectors will often

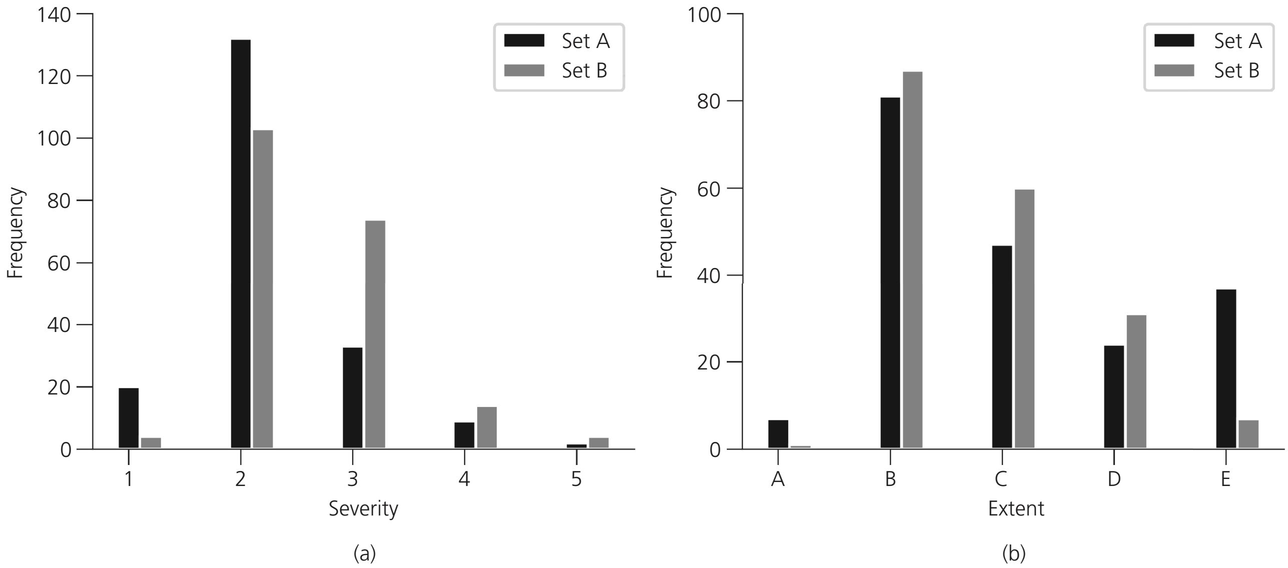

Figure 4. (a) Frequency of severity ratings in sets A and B; (b) frequency of extent ratings in sets A and B

assign a higher severity to defects compared with their on-site counterparts. This is further examined in the section headed ‘Rating difference of comparable defects’. The authors note that where the Mann–Whitney *U*-test was employed in this paper, the unequal variances *t*-test was also used.

Extent
Figure 4(b) shows that for both sets, A and B are the least and most frequent extent ratings, respectively. In set B, the count decreases with higher extent ratings. There was, however, a much higher number of E extent ratings recorded in set A compared with that in set B. This resulted in a *p*-value of 0.06. This also indicated that the extent ratings from both sets were statistically less different when comparing severity.

Defect class
A plot of the frequency distribution by defect class is shown in Figure 5. Both inspector groups had the same five most frequent defect classes. Group B inspectors recorded much more vegetation/maintenance defects compared with group A inspectors. Similar disproportion was seen in the defect classes ‘others’ and ‘paint/element surface’. A more in-depth analysis of the data showed that these were generally more superficial defects. Group A inspectors logged higher proportions for the defect classes ‘concrete defects’, ‘corrosion’ and ‘cracks’. This result suggested that group A inspectors were more confident in classifying these defects. The variability of defect classes between the two groups pointed to ambiguities in the current list of defects, which could cause inspectors to be unsure which defect to select.

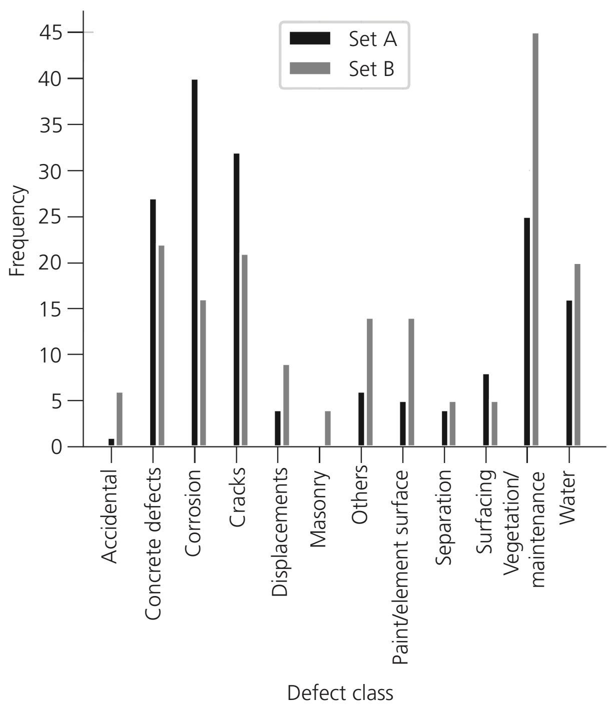

Figure 5. Frequency of defect classes recorded in set A and set B

Priority
Figure 6 indicates that when observing the priority ratings given, defects in set B tend to be of higher priority when compared with those in set A. This trend reinforces the suggestion that the remote inspectors tend to have a more pessimistic assessment for a given defect than the original on-site inspectors. This tendency may be due to wider situational awareness and a more holistic appreciation of bridge issues for the inspectors that are on the ground.

Defect indications
This section examines the agreeability between the ratings from the two data sets. Many of the defects recorded in set A consisted of one defect rating description and rating, mapped to various components in the structure. See Figure 7 and Table 6 as an example. Corrosion of a parapet mesh was recorded with a rating of 1B and was then mapped to six components. This resulted in six recorded defects, contributing to the overall total of 196 defects. For this comparative analysis, a distinction was made between these recorded defects and the actual unique ratings given (i.e. solely taking the unique rating as one defect). From the 196 defects recorded in set A, this resulted in 100 unique defect ratings. Analysis was conducted on what proportion of these 100 defects in set A were ‘indicated’ in set B. A defect was judged to be an indication if the remote inspector was clearly referring to a defect in set A. This result was determined using a combination of the comments and photograph IDs provided by the off-site inspector.

Out of the 100 unique defects in set A, 67 were indicated by group B inspectors. The severity and extent ratings of these 67 defects were examined further. Figure 8(a) shows the frequency count of set A defects by severity rating (denoted by a grey bar) and what proportion of these defects were indicated by group B inspectors (denoted by a dashed line). From the plot, it is observed that when the severity of a defect in set A is higher,

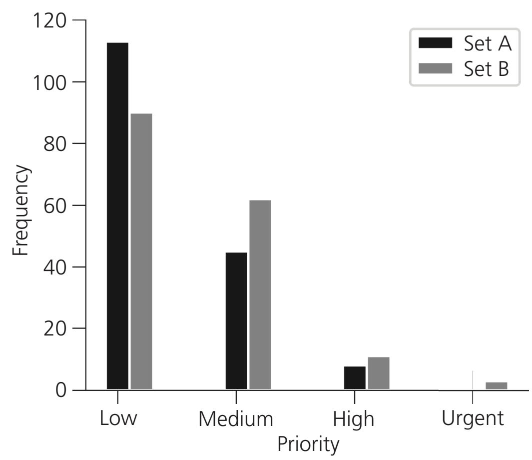

Figure 6. Frequency of priority ratings recorded in set A and set B

Defect type	Defect class	Severity	Extent	Priority	Component type	Component name
RS	Corrosion	1	B	Low	Centre deck	North parapet
RS	Corrosion	1	B	Low	Centre deck	South parapet
RS	Corrosion	1	B	Low	West deck	North parapet
RS	Corrosion	1	B	Low	West deck	South parapet
RS	Corrosion	1	B	Low	East deck	North parapet
RS	Corrosion	1	B	Low	East deck	South parapet

Figure 7. Example of the format of a defect record in Set A. One unique defect rating is assigned to six components. Photograph courtesy of WSP, used with permission (see Table 6 for recorded attributes)

Table 6. Recorded attributes for the defects shown in Figure 7

Defect attribute	Set A	Set B
Defect type	RS	RCo
Defect class	Corrosion	Corrosion
Severity	1	2
Extent	B	Insufficient data available
Priority	Low	Low
Component type	Multiple (see Figure 7)	Road vehicle restraint
Component name	Multiple (see Figure 7)	Insufficient data available
Comments	Minor corrosion and distortion to parapet mesh infill panels	Not clear which parapet mesh infill this is showing. Mesh infill has surface corrosion present

RCo, rusty/corroded; RS corroded/rusty

the number of defects indicated by a group B inspector also increases. This result suggests that although remote inspectors were found in this study to miss about one-third of the defects reported by on-site inspectors, they were very unlikely to miss high-severity defects but highly likely to miss low-severity defects. A similar trend could also be seen for the extent ratings (Figure 8(b)): the higher the extent rating, the higher the proportion of defects indicated by an off-site inspector.

Defect agreement

The percentage agreement for different attributes of the 67 independently recorded defects between group A and group B is shown in Table 7. It was observed that 63% of the defects had matched for 'class'. This was followed by 'severity', with a 49% match, and then 'extent', with 33% agreement. This considerable variation in the way defects were recorded by independent inspectors was noteworthy, highlighting the subjectivity of the visual inspection procedure.

Table 7. Percentage agreement between defect attributes

Defect attribute	Agreement: %
Class	63
Severity	49
Extent	33
Severity and extent	21
Class, severity and extent	16

These results were similar to the findings of the study by Bennetts *et al*. (2018a), which compared two independent inspectors' ratings of defects on 200 bridges on the Highways England network. Each bridge structure was given two inspectors: one from the pertinent service provider, who completed the regularly scheduled PI, and one from WSP Ltd, who detected and rated defects independently but without a thorough inspection (see the paper by Bennetts *et al*. (2018a) and the discussion of this study in the paper by Nepomuceno *et al*. (2021)). The results showed that individual inspectors documented defects in a highly variable manner (Bennetts *et al*., 2018a; Nepomuceno *et al*., 2021). Nearly 30% of the defects reported by WSP inspectors could not be matched with a comparable defect in the service provider's PI reports (Bennetts *et al*., 2018a; Nepomuceno *et al*., 2021). The present study findings revealed that remote inspectors missed 30% of defects; this value was no worse than sending a second independent on-site inspector to conduct the inspection. This result provides some evidence that remote inspection does not degrade inspection quality, although a larger data set would be required to give further confidence in this observation.

Statistical difference of comparable defects

The results of the Mann–Whitney *U*-test for these comparable defects for a significance level of 0.05 is shown in Table 8. For severity, a *p*-value of 0.001 denotes significant statistical difference between on-site and off-site inspectors. Considering extent, a larger significant value of 0.390 is observed, suggesting better agreement among ratings. This aspect is further explored in the following section.

Rating difference of comparable defects

Figure 9 shows the distribution of rating difference for these comparable defects. This difference was calculated by subtracting

Table 8. Results of the Mann–Whitney U-test for comparable defects

Attribute	Set A				Set B				Significant value
	N	μ	σ	σ^2	N	μ	σ	σ^2	Set A–Set B
Severity	67	2.34	0.72	0.52	67	3.15	1.16	1.35	0.001
Extent	67	2.75	0.85	0.73	65	2.98	1.99	1.02	0.390

the pertinent set A value from the corresponding set B value. Thus, a positive value indicates that a given defect received a more severe and extensive assessment in set B. A normal distribution can be fitted to the data, with the highest proportion of defects having zero rating difference. Again, the data indicate that more severe ratings are given in B, but a more symmetrical distribution is observed when it comes to extent. There are slightly more negative values for extent, meaning that off-site inspectors tend to rate less extensively.

Structure

This section examines the defects recorded by structure. Figure 10 shows the defect counts by structure for sets A and B. Structure S9 had the highest number of recorded defects, while S5 had the lowest. The off-site inspectors recorded more defects on structures S1, S3, S4, S5 and S6. It was also observed that S2, S7, S8 and S10 had notably higher defects in set A. Figure 11 shows the percentage of defects in set A indicated by the corresponding group B inspector. From the plot, the low percentages for S7 (38%) and S8 (35%) are notable. Both these structures were in the remit of AO2, and it was observed by the first author that the photographs for these structures were of lower quality compared with the photographs of the structures in AO1.

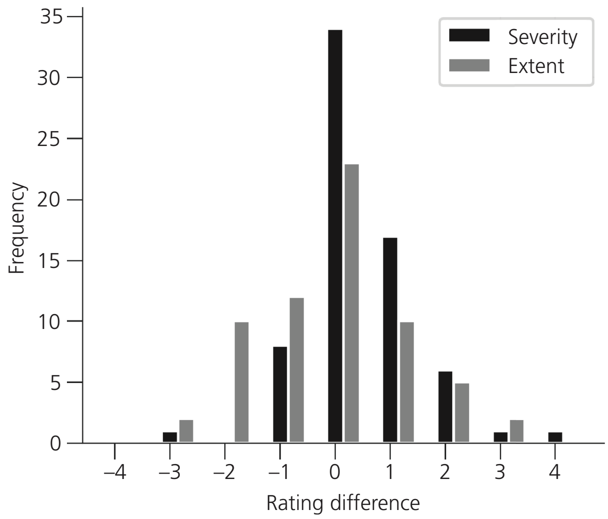

Figure 9. Plot showing the rating difference between directly comparable defects

Confidence levels

Finally, participants were asked to give an indication of their confidence level when rating the 'type', 'severity' and 'extent' of

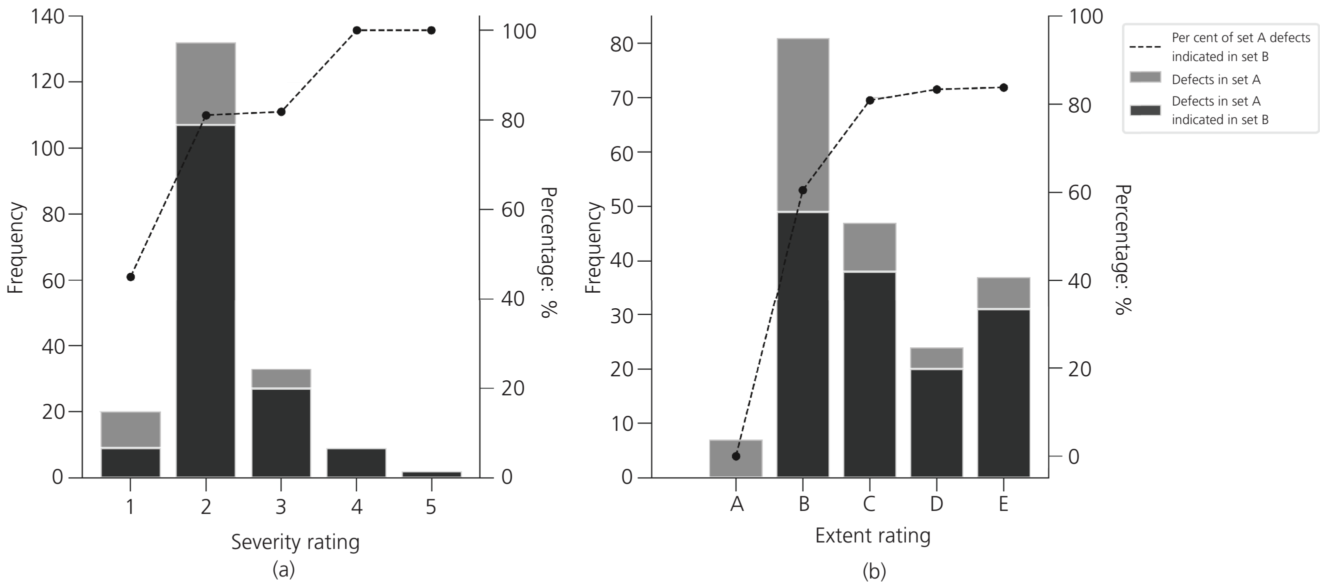

Figure 8. (a) Plot showing the number and proportion of defects (dashed line) in set A that are indicated by inspectors in group B, by severity. The dark column bars represent defects indicated in both set A and set B. (b) Plot showing the number and proportion of defects (dashed line) in set A that are indicated by inspectors in group B, by extent

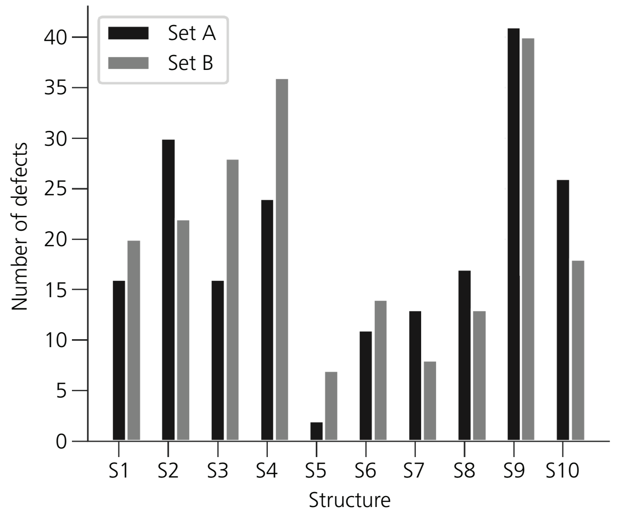

Figure 10. Number of defects recorded per structure

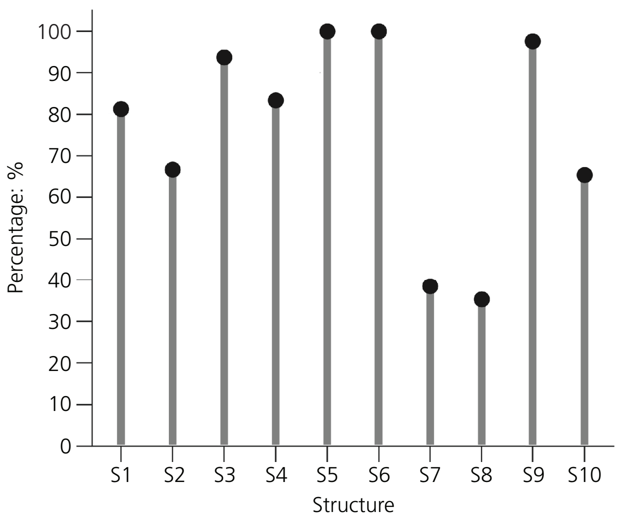

Figure 11. Percentage of set A defect ratings indicated by the off-site inspector per structure

each defect. Choices of 'low', 'medium' and 'high' were given. The results are shown in Figure 12; the results suggested that off-site inspectors were the least confident in their ability to rate extent. This sentiment was reflected in several of the comments included in questionnaires A and B:

- '[t]he extent was the hardest due to what seemed like a lack of photos showing the true extent of some of the defects' (P7)
- '… extent [was] practically impossible [to rate] without knowing which other photos related to the same component' (P8).

This outcome is perhaps unsurprising, as one of the key challenges affecting an off-site inspector is a loss of sense of scale.

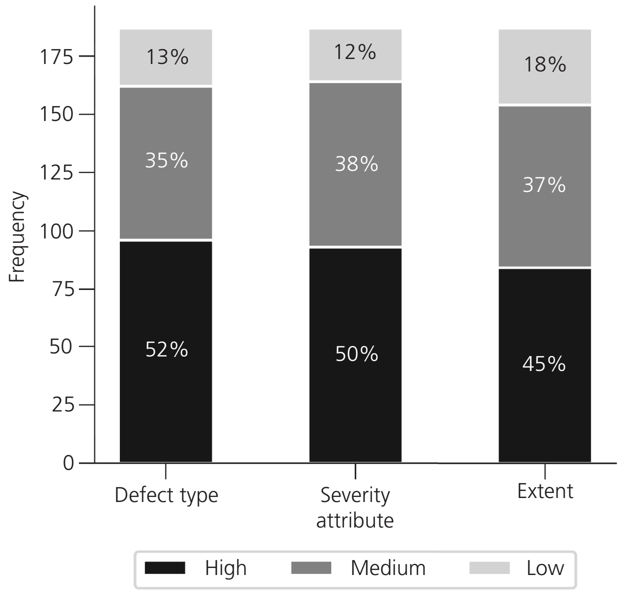

Figure 12. Confidence levels indicated by group B when rating defect type, severity and extent

Discussion

The results from this study offer useful information on the feasibility of human inspectors rating bridge defects using photographic images. Additionally, it adds to the body of data regarding visual inspection subjectivity. The following sections (i.e. the next four sections) provide some key discussion points that emerged from examination of the results of the study. The section headed 'Evaluating the remote inspection process' gives some comments on how the study participants thought the off-site inspection protocol could be improved. The section headed 'Looking ahead' gives suggestions for future research directions as a result of this study.

Off-site inspectors recorded more defects compared with their on-site counterparts

When considering the number of unique defects listed by inspectors in group A and group B, overall, a slightly higher percentage of unique defects were recorded by the off-site inspectors (see Table 5). This issue may be explained by the off-site inspectors erring on the 'safe' side and pointing out a somewhat mild defect 'just in case' it is a problem. It could be argued that increasing the number of defects recorded increases the likelihood of discovering a serious defect.

Off-site inspectors are effective at recording onerous defects

The plots in Figures 8(a) and 8(b) indicated that while off-site inspectors were effective at recording more severe defects, they were less successful at identifying less onerous defects. Out of the 73 defects in set A rated 2 or lower, 63% were indicated by set B inspections, whereas 78% of the set A defects rated 3 and above were indicated by set B. This result may be viewed as a promising outcome, as one could argue that it is the higher-severity defects that are critical to detect and the most vital to consider when implementing maintenance actions, particularly for GIs.

Off-site inspectors tend to rate individual defects more severely

Of the 67 comparable defects in both data sets, it can be shown that individual defects are typically assigned a higher severity rating by off-site inspectors. This tendency is further supported by the plot shown in Figure 6, which shows the distribution of priority ratings assigned; group B inspectors tended to rate defects of a higher priority. As mentioned, this might be due to a form of task bias within off-site inspectors. Because they had the aim of identifying defects (particularly the more onerous instances), they might naturally be inclined towards assigning more severe ratings to be on the safe side. It was acknowledged that an increased number of defects might make it more difficult for an off-site inspector to appreciate the wider picture and be too focused on individual defects. However, this process would rely on inspectors being appropriately qualified (e.g. BICS (Lantra, 2021)). A qualified inspector who had working knowledge of structural articulation would be expected to understand the criticality of individual aspects of various components with regard to the overall structural integrity of a bridge.

Loss of sense of scale and orientation is a significant barrier to off-site inspectors

When participants were surveyed about their degree of confidence, 'defect type' ratings were rated as the most confident, while extent ratings were rated as the least confident. There was also particular difficulty when assessing where a defect was on a structure, making location and orientation very difficult to determine. Several participant comments alluded to this:

- '… hard to visualise where some of [the defects] were and how they relate to other defects' (P1)
- 'not all photos showed [the] location of defects making it difficult to identify' (P2)
- '[t]he orientation of the pictures was very difficult to identify' (P4)
- '[i]t is difficult to assess extent when the majority of photos are close ups with no sense of scale' (P6)
- '… found it confusing [when trying to work out] which element/direction I was looking' (P9)
- 'extent [is] impossible [to rate] when [the] whole element not in photo' (P10).

These findings, along with the comments, point to a major weakness in the method of bridge assessment as explored in this work.

Evaluating the remote inspection process

Potential solutions

A number of participants' comments provided valuable insights into how remote inspection might be improved. Many of these emphasised the need for information relating to defect location: '[a]nything to allow the person grading the defects to easily identify where on the structure it is would [be] very helpful' (P7).

One bridge inspector noted that 'if a sketch was included with some basic comments or notes where defects [were], this would greatly increase confidence [when rating defects]' (P2). Some suggested that supplementing photographs with details such as direction and scale would be of value:

- 'including both close up and wide-angle photos for each defect to aid in locating area of structure. Including a scale in photos' (P6)
- 'so long as photo location/direction are recorded, and some scaling tool is available' (P9).

One participant stated that a 'walk-through video' (P1) would be helpful. As also noted by Nepomuceno *et al*. (2022b), another participant mentioned that 'a 5-minute chat with the person that took the photos' (P8) would facilitate knowledge transfer pertaining to the bridge under investigation. The use of 360° camera imagery as an adjunct to the traditional 2D photographs could offer benefits in providing a more comprehensive view of the bridge structure. This approach could assist inspectors in gaining a better understanding of the defect location and context within the overall structure. By including both close-up and wide-angle photographs, along with a scale for reference, inspectors could identify more easily the specific area affected and assess the extent of the defect.

Foreseen barriers to the procedure

Valuable comments were provided on the actual procedure itself. Two participants mentioned the time taken to undertake the survey, stating that it 'took quite a time to do' (P1) and that it 'took more than 1 hour' (P4). One participant gave insight into how the number of photographs provided to a remote inspector would have to be carefully considered: '[o]nce structures get above a certain size, … the number of photos to be reviewed would get very daunting, and the inspector could easily get lost amongst the data. I have experience of uploading others' inspection notes that this can easily happen' (P8) (this quote is also included in the paper by Nepomuceno *et al*. (2022b)). Finally, the quality of photographs was stated plainly by one participant, saying that 'inspectors are not professional photographers, and it shows: defects out of focus, dark and only some on screen' (P10).

Looking ahead

The results from this study show promise in utilising remote defect assessment methods to supplement the routine inspection regime, rather than completely replacing it. The findings from this work indicate that 'off-site' inspectors are unlikely to miss higher-severity defects. By conducting such a process perhaps on an annual basis, the level of granularity needed to track rapidly deteriorating defects on aged structures would increase. This approach would help take a step towards being able to quantify accurately the rate of change of deterioration, which the current assessment system does not sufficiently enable (Bennetts *et al*., 2021). In turn, this may increase the quality of data taken by visual inspections.

Such a method would also align well with the changing role of a bridge inspector. In the past, the role was typically seen as a long-term career, where one would amass decades of experience and be extremely knowledgeable of the structural behaviour of many bridges in a region of the network. There is now a higher turnover in these posts, where bridge inspectors come into the role for a few years and then move on to a different position. By having a procedure where defect photographs are professionally taken and assessed remotely, less experienced inspectors may be able to undertake defect identification and grading with more confidence, facilitating 'digital stewardship' of the asset. A recently proposed schema for the remote inspection of bridges was recently published by the authors to show how these aims can be achieved (Nepomuceno *et al*., 2022b).

Regarding the experience level of inspectors, in the future, a similar experimental trial could be devised where one bridge is virtually rated by multiple inspectors of various experience levels. By virtually rating the same bridge, the scoring ratings and comments provided by inspectors with different experience levels can be compared, providing deeper insights into the relationship between experience and the quality of defect assessments.

Summary and conclusions

To investigate the effectiveness of off-site inspections for assessing bridge defect ratings, a targeted trial was designed wherein participants rated bridge defects using photographs taken during the on-site GIs of a particular structure. The results were then compared with the ratings assigned to the same structure by the on-site inspector. In all, ten structures were inspected as part of the study. The on-site inspectors recorded 196 defects across the ten structures, while the remote inspectors (i.e. participants of the study) recorded 206 defects (of which 186 could be confidently classed).

Statistical analysis shows that aggregated defects rated by off-site inspectors tend to be perceived as more severe and of a higher priority compared with those from the on-site inspectors. The results also indicate that for defects of higher severity and extent, there is closer agreement in ratings between on-site and off-site inspectors. The results from this study suggest that there may be promise in standardising remote inspections for the identification and grading of more urgent, higher-severity defects. The authors note that any move to a remote inspection process should at first be seen as a complementary approach to traditional on-site inspections, rather than replacing them entirely. A potential future is seen where aspects of GIs are made remote, while PIs will remain the same, wherein on-site inspectors can gain practical field experience. Finally, it is noted that a larger data set would be needed to give further confidence in the trends reported in this paper.

Acknowledgements

The first author would like to acknowledge support from WSP, as well as from the Engineering and Physical Sciences Research Council through the National Productivity Investment Fund (grant number EP/R51245X/1). The authors express thanks to the anonymous individuals who gave their time to participate in this study. The statements made in this paper do not necessarily reflect the opinions of the organisations involved.

Data availability statement

To maintain anonymity of the survey participants the survey data cannot be made available in a data repository. No other experimental data were generated during this study.

REFERENCES

Bennetts J (2019) *The Management of Bridges*. EngD thesis, University of Bristol, Bristol, UK.

Bennetts J, Vardanega PJ, Taylor CA and Denton SR (2016) Bridge data – what do we collect and how do we use it? In *Transforming the Future of Infrastructure through Smarter Information: Proceedings of the International Conference on Smart Infrastructure and Construction, 27–29 June 2016* (Mair RJ, Soga K, Jin Y, Parlikad AK and Schooling JM (eds)). ICE Publishing, London, UK, pp. 531–536.

Bennetts J, Webb G, Denton S, Vardanega PJ and Loudon N (2018a) Quantifying uncertainty in visual inspection data. In *Maintenance, Safety, Risk, Management and Life-cycle Performance of Bridges: Proceedings of the Ninth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2018), 9–13 July 2018, Melbourne, Australia* (Powers N, Frangopol DM, Al-Mahaidi R and Caprani C (eds)). CRC Press/Balkema, Leiden, the Netherlands, pp. 2252–2259.

Bennetts J, Webb GT, Vardanega PJ, Denton SR and Loudon N (2018b) Using data to explore trends in bridge performance. *Proceedings of the Institution of Civil Engineers – Smart Infrastructure and Construction* **171(1)**: 14–28, https://doi.org/10.1680/jsmic.17.00022.

Bennetts J, Vardanega PJ, Taylor CA and Denton SR (2020) Survey of the use of data in UK bridge asset management. *Proceedings of the Institution of Civil Engineers – Bridge Engineering* **173(4)**: 211–222, https://doi.org/10.1680/jbren.18.00050.

Bennetts J, Denton SR, Webb GT, Nepomuceno DT and Vardanega PJ (2021) Looking to the future of bridge inspection and management in the UK. In *Bridge Maintenance, Safety, Management, Life-cycle Sustainability and Innovations: Proceedings of the Tenth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2020), June 28–July 2, 2020, Sapporo, Japan* (Yokota H and Frangopol DM (eds)). CRC Press/Balkema, Leiden, the Netherlands, pp. 3858–3866.

Bennetts J, Vardanega PJ, Webb GT, Denton SR and Loudon N (2022) Analysis of visual inspection data for a sample of highway bridges in the UK. *Proceedings of the Institution of Civil Engineers – Forensic Engineering*, https://doi.org/10.1680/jfoen.21.00005.

Bhreasail AN, Pritchard O, Carluccio S *et al.* (2019) Remote sensing for proactive geotechnical asset management of England's Strategic Roads. *Infrastructure Asset Management* **6(4)**: 222–232, https://doi.org/10.1680/jinam.17.00025.

Dodge Y (2008) Mann–Whitney test. In *The Concise Encyclopedia of Statistics*. Springer, New York, NY, USA, pp. 327–329.

HA (Highways Agency) (2007) *The Inspection Manual for Highway Structures, Volume 1: Reference Manual*. The Stationery Office, London, UK.

HE (Highways England) (2021) CS 450: Inspection of highway structures. In *Design Manual for Roads and Bridges*. HE, Birmingham, UK.

See https://www.standardsforhighways.co.uk/search/c5c2c3e5-f7f3-4c94-8254-184e41ccd1a0 (accessed 12/06/2023).

Huang Z, Ma M, Tam VWY and Lang H (2021) Critical success factors for developing construction and demolition waste management in China. *Proceedings of the Institution of Civil Engineers – Engineering Sustainability* **174(5)**: 213–223, https://doi.org/10.1680/jensu.20.00053.

Lantra (2021) *National Highway Sector Scheme 31 for the Bridge Inspector Certification Scheme: Scheme Manual,* version 2. Lantra. Coventry, UK. See https://bics.skills-plus.net/assets/uploads/SCHEME-MANUAL-2021.pdf (accessed 12/06/2023).

Lea F and Middleton C (2002) *Reliability of Visual Inspection of Highway Bridges*. University of Cambridge, Cambridge, UK, CUED/D-STRUCT/TR.201.

Mahut B and Woodward RJ (2005) Comparison of bridge management practice in England and France. In *Bridge Management 5: Inspection, Maintenance, Assessment and Repair: Proceedings of the 5th International Conference on Bridge Management, Organized by the University of Surrey, 11–13 April 2005* (Parke GAR and Disney P (eds)). Thomas Telford, London, UK, pp. 163–170.

Mann HB and Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* **18(1)**: 50–60, https://doi.org/10.1214/aoms/1177730491.

McRobbie SG (2009) *Automated Inspection of Highway Structures 2008/09*. TRL Ltd, Wokingham, UK, Report PPR412.

McRobbie SG, Lodge R and Wright A (2007) *Automated Inspection of Highway Structures – Stage 2*. TRL Ltd, Wokingham, UK, Report PPR255.

McRobbie SG, Wright MA and Chan A (2015) Can technology improve routine visual bridge inspections? *Proceedings of the Institution of Civil Engineers – Bridge Engineering* **168(3)**: 197–207, https://doi.org/10.1680/jbren.12.00022.

Middleton C (2004) Bridge management and assessment in the UK. *Austroads 2004 Bridge Conference: Bridges Another Dimension, Hobart, Australia*. See https://web.archive.org/web/20221024074022/http://www.bridgeforum.org/files/pub/2004/austroads5/Middleton%20keynote.pdf (accessed 12/06/2023).

Moore M, Phares B, Graybeal B, Rolander D and Washer G (2001) *Reliability of Visual Inspection for Highway Bridges, Volume I: Final Report*. Federal Highway Administration, US Department of Transportation, Washington, DC, USA. Report No. FHWA-RD-01-020. See https://www.fhwa.dot.gov/publications/research/nde/pdfs/01020a.pdf (accessed 12/06/2023).

Nepomuceno DDT (2022) *Technology Innovation for Improving Bridge Management*. PhD thesis, University of Bristol, Bristol, UK.

Nepomuceno DT, Vardanega PJ, Tryfonas T *et al.* (2021) Viability of off-site inspections to determine bridge defect ratings. In *Bridge Maintenance, Safety, Management, Life-cycle Sustainability and Innovations: Proceedings of the Tenth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2020), June 28–July 2, 2020, Sapporo, Japan* (Yokota H and Frangopol DM (eds)). CRC Press/Balkema, Leiden, the Netherlands, pp. 3688–3695.

Nepomuceno DT, Vardanega PJ, Tryfonas T *et al.* (2022a) A survey of emerging technologies for the future of routine visual inspection of bridge structures. In *Bridge Safety, Maintenance, Management, Life-cycle, Resilience and Sustainability: Proceedings of the Eleventh International Conference on Bridge Maintenance, Safety and Management (IABMAS 2022), Barcelona, Spain, July 11–15, 2022* (Casas JR, Frangopol DM and Turmo J (eds)). CRC Press/Balkema, London, UK, pp. 846–854.

Nepomuceno DT, Bennetts J, Pregnolato M, Tryfonas T and Vardanega PJ (2022b) Development of a schema for the remote inspection of bridges. *Proceedings of the Institution of Civil Engineers – Bridge Engineering*, https://doi.org/10.1680/jbren.22.00027.

Phares BM, Washer GA, Rolander DD, Graybeal BA and Moore M (2004) Routine highway bridge inspection condition documentation accuracy and reliability. *Journal of Bridge Engineering* **9(4)**: 403–413, https://doi.org/10.1061/(ASCE)1084-0702(2004)9:4(403).

SciPy (2021) *scipy.stats.mannwhitneyu*. NumFOCUS, Austin, TX, USA. See https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html (accessed 12/06/2023).

Wallbank EG (1989) *The Performance of Concrete in Bridges: A Survey of 200 Highway Bridges*. Her Majesty's Stationery Office, London, UK.

How can you contribute?

To discuss this paper, please submit up to 500 words to the editor at journals@ice.org.uk. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial board, it will be published as a discussion in a future issue of the journal.