Structural reliability updating through proof load testing

A Bayesian methodology applied to reinforced concrete road bridges and viaducts

Rein de Vries

Structural reliability updating through proof load testing

A Bayesian methodology applied to reinforced concrete road bridges and viaducts

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus,
Prof. dr. ir. H. Bijl,
chair of the Board for Doctorates
to be defended publicly on
Wednesday, 15 April 2026 at 17:30

by

Rein DE VRIES

This dissertation has been approved by the (co)promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. ir. M. A. N. Hendriks	Delft University of Technology, Norwegian University of Science and Technology, Norway, promotor
Prof. dr. ir. R. D. J. M. Steenbergen	Netherlands Organisation for Applied Scientific Research (TNO), Ghent University, Belgium, promotor
Prof. dr. ir. E. O. L. Lantsoght	Delft University of Technology, Universidad San Francisco de Quito, Ecuador, copromotor
Independent members:	
Prof. dr. A. V. Metrikine	Delft University of Technology
Prof. dr. J. D. Sørensen	Aalborg University, Denmark
Prof. dr. J. W. Schmidt	Aalborg University, Denmark
Em. prof. ir. A. C. W. M. Vrouwenvelder	Netherlands Organisation for Applied Scientific Research (TNO), Delft University of Technology
Em. prof. dr. ir. J. G. Rots	Delft University of Technology, reserve member

Keywords: Bayesian updating, existing structures, infrastructure, load testing, proof load, reliability, time-dependence, traffic loads

ISBN 978-94-6518-275-9
https://doi.org/10.4233/uuid:73fb7b95-5b81-4362-9c14-41cc11242569

An electronic copy of this dissertation is available at https://repository.tudelft.nl, and the corresponding data and code at https://data.4tu.nl.

"The most important questions of life are indeed, for the most part, really only problems of probability."

— Pierre-Simon Laplace (Laplace, 1814)

Contents

Summary

The continuous ageing of infrastructure raises concerns about structural safety and reliability, making accurate structural assessments increasingly important. Proof load testing is presented as an effective way to verify structural reliability by intentionally subjecting bridges and viaducts to high load levels. However, challenges and research needs associated with the probabilistic substantiation of proof load testing for this purpose still remain, in particular considering time-dependent effects, varying knowledge levels, optimal use of in-situ monitoring, and system reliability aspects. These topics are centred around the main research question: *How can proof load testing establish the structural reliability of reinforced concrete bridges and viaducts?* Through exploratory analyses based on highly informed models and conservative lower-bound estimates, the transition to a flexible Bayesian methodology is made.

Annual reliability is first examined to address varying remaining life spans and to study time-dependent effects such as traffic trends and resistance deterioration. A detailed case study of a short-span concrete slab bridge shows a temporary dip in reliability during testing, followed by a substantial increase when the target load is survived. The conditional probability calculations demonstrate how proof load testing fits into a more generic reliability-updating context, where not only the resistance distribution is updated. To address situations where limited information is available, a conservative lower-bound approach solely based on the survived load effect is explored. Although it serves as the background for current standards and guidelines on proof load testing, conceptual and implementation challenges exist. An up-to-date probabilistic description is derived, but the method still lacks the ability to incorporate relevant structural information.

As neither the highly informed analyses with detailed resistance information nor the lower-bound approach accurately reflects the actual state of knowledge in many assessments, a Bayesian methodology is adopted. From a Bayesian perspective, mechanical resistance models are treated as sources of information rather than required inputs, enabling the method to address a broad spectrum of knowledge levels. In this way, any available information can be taken into account, even if it is subject to considerable uncertainty, as beliefs can be quantified through prior specification. Although a low-informative prior distribution can be adopted for the resistance, the primary value of the proposed methodology lies in viewing proof load testing within a broader Bayesian reliability updating context.

The Bayesian methodology forms a basis for developing flexible approaches to proof load testing that can incorporate monitoring data and account for spatial correlations. During a proof load test, extensive information about the structure becomes available through in-situ monitoring. Instead of solely evaluating stop criteria, these

measurements can be used to infer structural performance. In a novel approach, in-situ monitoring and laboratory testing data are combined in a Bayesian update of the structural reliability following each successful load increment. Case studies that integrate laboratory testing data and analytical modelling are shown to allow test load reductions of up to 25% compared to lower-bound estimations. In addition, the probability of failure under high test loads can be accurately quantified, supporting risk-based decisions on whether to proceed with further loading. Another application where the Bayesian methodology proves valuable is the assessment of structures with many components, critical sections, and multiple possible failure mechanisms. In these cases, it is typically challenging to demonstrate sufficient reliability with only a limited number of load tests. To address this challenge, a new system-reliability approach has been developed utilising a hierarchical Bayesian model. The approach, which explicitly accounts for spatial correlation in resistance and load effects, is demonstrated through two case studies: one derives transfer factors quantifying the required increase in target loads, and the other identifies optimal testing strategies and vehicle configurations.

Altogether, this research demonstrates that proof load testing should not be viewed as a standalone structural reliability assessment tool, but as a key element within a Bayesian reliability updating framework. When combined with advanced probabilistic modelling, monitoring, and system reliability considerations, proof load testing provides a *uniquely accurate assessment* method for establishing the reliability of existing infrastructure. This property follows from observing actual structural behaviour under high loads and integrating these observations with mechanical models. Although Bayesian methods have previously been applied in proof load testing, this work is the first to employ in-situ monitoring for evidence-based performance prediction—effectively replacing stop criteria that lack a clear reliability basis. In parallel, the proposed hierarchical Bayesian model unifies single-component proof load and multi-component sample testing, enabling the optimisation of test locations, target loads and test vehicle layouts. While Bayesian methods provide a sound mathematical basis, they must be applied carefully. Their use is demonstrated through case studies, indicating that the gap between research and practice is smaller than it may initially appear to be. Future research should focus on refining the proposed procedures through practical implementation, while international collaboration will be essential for future standardisation.

Samenvatting

De voortdurende veroudering van infrastructuur roept vragen op over de constructieve veiligheid en betrouwbaarheid, waardoor nauwkeurige beoordelingsmethoden steeds belangrijker worden. Proefbelasten wordt voorgesteld als een effectieve manier om de constructieve betrouwbaarheid te verifiëren door bruggen en viaducten opzettelijk aan hoge belastingen te onderwerpen. Er blijven echter uitdagingen en onderzoeksbehoeften bestaan met betrekking tot de probabilistische onderbouwing van proefbelastingen voor dit doel, met name gezien de tijdsafhankelijke effecten, variërende kennisniveaus, optimaal gebruik van in-situ monitoring en aspecten van systeembetrouwbaarheid. Deze onderwerpen staan centraal in de hoofdonderzoeksvraag: *Hoe kan een proefbelasting de constructieve betrouwbaarheid van gewapend betonnen bruggen en viaducten aantonen?* Door middel van verkennende analyses op basis van goed geïnformeerde modellen en conservatieve ondergrensschattingen wordt de overgang naar een flexibele Bayesiaanse methodologie gemaakt.

De jaarlijkse betrouwbaarheid wordt eerst onderzocht om rekening te kunnen houden met variaties in de resterende levensduur en om tijdsafhankelijke effecten zoals verkeerstrends en afname van de weerstand te bestuderen. Een gedetailleerde casus betreffende een betonnen brug met korte overspanning laat een tijdelijke daling van betrouwbaarheid zien tijdens het testen, gevolgd door een aanzienlijke stijging wanneer de proefbelasting wordt doorstaan. De conditionele probabilistische berekeningen laten zien hoe proefbelastingtests passen in de meer algemene context van betrouwbaarheidsupdates, waarin dus niet alleen de weerstandsverdeling wordt geactualiseerd. Voor situaties waarin beperkte informatie beschikbaar is, wordt een conservatieve benadering met een ondergrensmethode onderzocht die uitsluitend op het overleefde belastingeffect is gebaseerd. Hoewel deze benadering als achtergrond dient voor huidige normen en richtlijnen voor proefbelastingtests, zijn er conceptuele en implementatiegerelateerde uitdagingen. Een vernieuwde probabilistische beschrijving wordt afgeleid, maar de methode mist nog steeds de mogelijkheid om relevante constructieve informatie mee te wegen.

Aangezien noch de sterk geïnformeerde analyses met gedetailleerde weerstandsinformatie, noch de ondergrensbenadering een accurate weergave zijn van het kennisniveau in veel beoordelingen, wordt gekozen voor een Bayesiaanse methodologie. Vanuit Bayesiaans perspectief worden mechanische weerstandsmodellen behandeld als informatiebronnen in plaats van als vereiste invoer, waardoor de methode geschikt is voor een breed spectrum aan kennisniveaus. Op deze manier kan alle beschikbare en relevante informatie worden meegenomen, zelfs als deze aan aanzienlijke onzekerheid onderhevig is, omdat overtuigingen a priori kunnen worden gekwantificeerd. Hoewel voor de weerstand een laag-informatieve a priori verdeling kan worden gehanteerd, ligt

de primaire waarde van de voorgestelde methodologie in het beschouwen van proefbelastingtests binnen een bredere Bayesiaanse context van betrouwbaarheidsupdates.

De Bayesiaanse methodologie vormt een basis voor het ontwikkelen van flexibele toepassingen voor proefbelastingtests die monitoringgegevens kunnen integreren en rekening kunnen houden met ruimtelijke correlaties. Tijdens een proefbelastingtest komt uitgebreide informatie over de constructie beschikbaar via in-situ monitoring. In plaats van uitsluitend stopcriteria te evalueren, kunnen deze metingen worden gebruikt om de constructieve sterkte af te leiden. Met een nieuw ontwikkelde methode worden in-situ monitoring en laboratoriumproefgegevens gecombineerd in een Bayesiaanse update van de constructieve betrouwbaarheid. Studies waarin laboratoriumgegevens en analytische modellering worden geïntegreerd, tonen aan dat de testbelasting tot wel 25% kan worden gereduceerd in vergelijking met ondergrensschattingen. Bovendien kan de kans op falen bij hoge testbelastingen nauwkeurig worden gekwantificeerd, wat risicogebaseerde beslissingen over het voortzetten van de test ondersteunt. De Bayesiaanse methodologie is ook waardevol bij de beoordeling van constructies met veel componenten, kritieke doorsneden en meerdere mogelijke faalmechanismen. In dergelijke gevallen is het uitdagend om met een beperkt aantal proeven voldoende betrouwbaarheid aan te tonen. Om dit vraagstuk aan te pakken, is een nieuwe aanpak ontwikkeld waarbij gebruik wordt gemaakt van een hiërarchisch Bayesiaans model. De aanpak, die expliciet rekening houdt met ruimtelijke correlatie in weerstand en belastingseffecten, wordt gedemonstreerd aan de hand van twee casussen: de ene leidt overdrachtsfactoren af die de vereiste toename in de proefbelasting kwantificeren, en de andere identificeert optimale teststrategieën en voertuigconfiguraties.

Al met al toont dit onderzoek aan dat proefbelasten niet als een op zichzelf staand beoordelingsinstrument moet worden beschouwd voor de constructieve betrouwbaarheid, maar als een belangrijk onderdeel van een Bayesiaanse raamwerk. In combinatie met geavanceerde probabilistische modellering, monitoring en overwegingen betreffende systeembetrouwbaarheid, bieden proefbelastingtests een *uniek nauwkeurige beoordelingsmethode* voor het afleiden van de betrouwbaarheid van bestaande constructies. Deze eigenschap volgt uit het observeren van werkelijk constructief gedrag bij hoge belastingen en het integreren van deze observaties met kennis uit mechanische modellen. Hoewel Bayesiaanse methoden eerder zijn toegepast bij proefbelastingen, is dit onderzoek het eerste waarin in-situ monitoring constructieve sterkte voorspelt—zodat stopcriteria zonder duidelijke betrouwbaarheidsbasis worden vervangen. Daarnaast verenigt een nieuw hiërarchisch Bayesiaans model het proefbelasten van één component met de steekproefmatige beproeving van meerdere componenten, waarmee optimale testlocaties, doelbelastingen en voertuigindelingen kunnen worden bepaald. Hoewel Bayesiaanse methoden een wiskundig kader bieden, moeten ze zorgvuldig worden toegepast. De praktische toepasbaarheid wordt aangetoond aan de hand van casussen, waaruit blijkt dat de kloof tussen onderzoek en praktijk kleiner is dan op het eerste gezicht lijkt. Toekomstig onderzoek zal zich moeten richten op het verfijnen van de voorgestelde procedures via praktische toepassing, terwijl internationale samenwerking essentieel zal zijn voor toekomstige standaardisering.

1

Introduction

The continuous ageing of infrastructure raises concerns about structural safety and reliability, making accurate structural assessments increasingly important. Proof load testing is presented as an effective way to verify structural reliability by intentionally subjecting bridges and viaducts to high load levels. This chapter introduces the challenges and research needs associated with the probabilistic substantiation of proof load testing for reliability-based structural assessments. By formulating research questions and defining a consistent methodology, this study aims to fundamentally advance reliability-based proof load testing and support its international standardisation. An outline of the dissertation is provided to guide the reader through the key topics, including time dependence, Bayesian methods, and spatial correlation.

1.1 Assessment of existing structures

Buildings and civil engineering works are expected to meet specific reliability requirements throughout their entire design life. Reliability assessment of an existing structure becomes relevant when its original design life has passed, or doubts arise concerning its capacity to withstand future loads. The structure may reflect outdated design and construction practices or display performance issues (JCSS, 2001). Over time, degradation may have taken place, and traffic loads often increase, underscoring the need for assessment. With up-to-date knowledge about structural materials, resistance models and loads, a design that may have been sufficient in the past might not be adequate today. In the Netherlands, about 60% of bridges and viaducts were built before 1975 and are thus at least 50 years old (Lantsoght et al., 2013). Internationally, similar situations unfold with regard to the ageing bridge stock. In Europe, efforts are underway to incorporate the assessment of existing structures in the Eurocode standards (Holický, 2019).

In the assessment of existing structures, the original design reliability, based on prior knowledge, should be updated to reflect our current knowledge and the current state of the structure (Diamantidis et al., 2025; Faber, 2000). To judge whether the calculated reliability is sufficient, it is important to consider the differences in requirements between new and existing structures, as well as the time dependence of reliability itself. From an economic perspective, the reliability requirements for the design of new structures are generally higher than strictly necessary from a human safety perspective. This distinction is essential when considering the (environmental) costs of replacing a structure (Steenbergen & Vrouwenvelder, 2010; Stewart et al., 2001).

Inspections, structural assessments, and maintenance are essential for reliable bridges and viaducts. As the infrastructure ages and endures increased traffic loads and environmental challenges, versatile assessment methods are needed to address these evolving conditions. As a result, wear is often present, and it is difficult to tell if it impairs structural reliability. Fortunately, tests can be carried out on the structure to gather supplementary data. Tests on reinforced concrete structures commonly entail measuring the geometry, drilling cores or scanning reinforcement. Alternatively, load tests may be performed, and these may be roughly divided into three categories: diagnostic, proof load, and collapse tests. Diagnostic load tests quantify structural properties by measuring a structure's response to small to moderate loads. Typically, a finite element model of the structure is updated with the test results to better approximate the measured response. In proof load tests, larger loads are exerted to *prove* a structure has sufficient capacity to resist future loading situations. This type of testing is the subject of the current thesis and focuses on its probabilistic substantiation. Lastly, collapse tests may be performed to enhance the understanding of the structure's failure mechanism and the ultimate capacity. In reality, test types often overlap since the loads are applied in small increments and the test results may be used for different purposes (Alampalli et al., 2019; Lantsoght, 2019a).

1.2 Proof load testing

1.2.1 Historical development and international standards

Before international standardisation and calibrated safety factors were commonplace, load testing typically took place before bridges were opened to prove the bridge's safety to the general public. The loads applied during such tests lacked the theoretical substantiation that underpins current design and assessment procedures. In some countries, such as France, Italy and Switzerland, a load test is still required as part of the commissioning process for major bridges. However, these loads are typically small and are best described as diagnostic tests (Bolle et al., 2011; Lantsoght, 2019a). Today, load testing is increasingly recognised as a versatile structural assessment method, and its application and standardisation are becoming more prevalent (Alampalli et al., 2019).

The load applied to the structure during proof load testing is called the *target load*, and its magnitude is of primary interest. Similar to design loads, the target load is expected to be determined in accordance with relevant standards and aligned with normative reliability requirements. However, the assessment of existing structures

via proof load testing is not a standardised procedure in the Netherlands and many other countries. When national guidance is lacking, standards or guidelines from other countries can provide useful insight into accepted practices. In the USA, the Manual for Bridge Evaluation (MBE) (AASHTO, 2022) serves as a guideline for proof load testing. The target proof load is expressed in terms of the regular traffic load model and is magnified by a proof load factor (X_p). Its default value (1.4) was derived in a basic probabilistic analysis (Lichtenstein, 1993). Another relevant American standard is ACI 437.2M (ACI, 2013; 2023), which describes the requirements for load testing of existing concrete buildings, including loading protocols and acceptance criteria. The most recent edition (ACI 437.2M-22) also includes provisions for evaluating shear capacity.

Recently, the German committee for reinforced concrete has published a new version of its guideline for proof load tests on concrete structures (DAfStb, 2020). The guideline is mainly intended for buildings, but speaks in more general terms, such as structures and components. The magnitude of the proof load is expressed in a format that resembles the load effect in Equation (6.10) of EN 1990 (CEN, 2019). An interesting aspect of the guideline is the consideration of multiple similar components. When a limited number of equivalent components are tested, the remaining uncertainty associated with their slight differences is accounted for by increasing the magnitude of the test load (Marx et al., 2019).

1.2.2 State-of-the-art reliability interpretation

The use of load testing to assess the performance of a structure in relation to its structural reliability was already recognised in the 1980s, with pioneering work by Grigoriu and Hall (1984), Lin and Nowak (1984), and Rackwitz and Schrupp (1985). In these early works, the result of a successful proof load test is incorporated via a 'cut-off' of the resistance distribution function. Proof load testing is still an active field of research and continues to gain attention due to the growing need for versatile and accurate assessment methods (Lantsoght, 2019b; Lantsoght et al., 2017a). Improving procedures remains relevant to avoid replacing or upgrading bridges that are in satisfactory condition. Target load magnitudes should vary depending on the load rating, dead/live load ratios, degradation, bridge age, reference period and prior service loads (Faber et al., 2000).

A number of topics, such as degradation, changing reference periods and in-service proven strength (Wang et al., 2011), primarily relate to the time-dependent nature of structural reliability. An early description of the changing reliability in relation to proof load testing is found in Spaethe (1994). It is shown that during the proof load test, the reliability of the structure is low due to the relatively large applied load, but afterwards, the reliability is increased—if the test is successful. In the more recent work by Schacht et al. (2019) and Frangopol et al. (2019), the decrease of reliability with time in case of deterioration is also recognised. For bridges and viaducts, traffic loads are a primary source of time dependence, making an accurate statistical description essential. In Casas and Gómez (2013), proof load factors are derived as part of the large-scale Assessment and Rehabilitation of Central European Highway Structures (ARCHES) pro-

Figure 1.1 Pilot proof load tests in the Netherlands using the specialised BELFA loading vehicle on the Vlijmen-Oost viaduct in 2013 (left) and using a loading frame on the Vechtbrug in 2016 (right). Reprinted with permission from Lantsoght et al. (2017a) and De Vries et al. (2022).

ject. The study used recent traffic load data from five European countries, including the Netherlands, and offered the option to include bridge documentation.

Today, proof load testing is starting to be considered in the context of the more generic updating of structural reliability (Lantsoght, 2019b; Straub & Papaioannou, 2015). Applying the large loads commonly used in proof load testing is resource-intensive and imposes a risk on the structure, equipment, and personnel. To avoid excessively large loads, all relevant information about the structure should be considered, even when uncertain. Bayesian probability theory provides the opportunity to incorporate various sources of information into the updating process (Nishijima & Faber, 2007). The theory can provide a mathematical basis for updated distributions of the resistance after proof load testing (Yuefei et al., 2014). It can also provide decision support and identify valuable information to aid in the modelling and monitoring of structures (Kapoor, 2021; Schmidt et al., 2020; W.-H. Zhang et al., 2021). More details about state-of-the-art proof load testing in relation to structural reliability are provided in De Vries and Lantsoght (2022).

1.2.3 Towards standardised assessment procedures

To arrive at standardised assessment procedures for proof load testing, a combination of theory and practice is required. Even though there is ample experience with load testing internationally, country-specific factors, such as building regulations, bridge construction practices, traffic load models and load testing knowledge within the engineering community, play an important role. Several pilot proof load tests were performed to gain experience with the load testing assessment procedure in the Netherlands (Fennis et al., 2014; Koekkoek, Lantsoght, Yang, De Boer & Hordijk, 2016; Lantsoght et al., 2017b; 2017c; 2017d). In these tests, various options with regard to preparation, load application, and monitoring were explored (Figure 1.1).

From the pilot tests, lessons were learned about preparing sensor plans and applying various measurement types. Monitoring the structural response during testing is typically required to ensure the safety of personnel and the public in the vicinity of the

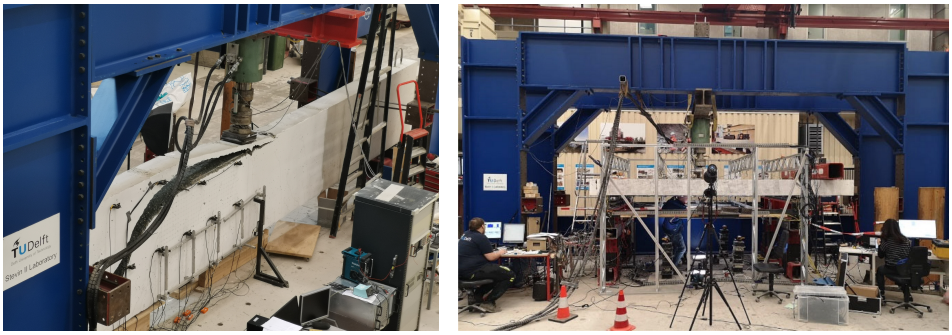


Figure 1.2 Laboratory testing of concrete strips (left) and slabs (right), displaying the application of the concentrated load and a large variety of sensors to monitor the specimen's response. Reprinted with permission from F. Zhang (2022) and Lantsoght (2023).

tested bridge or span. The loading protocol, including unloading steps, proved helpful in determining residual deformations and the occurrence of non-linear behaviour. Reaching the target load, as calculated before the execution of the load test, may not always be feasible due to signs of distress during load application. Even if the structure can withstand the target load, damage that would affect its remaining service life should be prevented. To judge whether distress or irreversible damage occurs, stop criteria may be evaluated. Stop criteria typically address unwanted structural behaviour but act on a measurable property (indicator). For example, excessive strains indicate that the reinforcement could be yielding; therefore, a stop criterion based on the strains is formulated. Various criteria are provided in the German guideline for proof load testing (DAfStb, 2000; 2020). However, in the pilot tests, several criteria proved difficult to apply in practice. For instance, stop criteria for steel strains are difficult to apply without removing the concrete cover. To address this issue, new criteria for flexure were proposed (Lantsoght et al., 2019).

By evaluating bridges and viaducts for which proof load testing would be desirable, it becomes clear that stop criteria are needed, particularly for the common concrete slab bridges. Despite their basic appearance, their structural behaviour is complex, and many are found to be shear-critical upon assessment. Defining the necessary stop criteria for slab bridges is a topic of current research (Christensen et al., 2022). The behaviour of concrete slabs under concentrated loads was recently studied in the laboratory using state-of-the-art sensing techniques, including Digital Image Correlation (DIC), fibre optic sensors, and acoustic emission sensors—in addition to traditional laser and LVDT measurements (Figure 1.2) (Zarate Garnica, Lantsoght & Yang, 2022; Zarate Garnica et al., 2024a; 2024b; F. Zhang, 2022).

Throughout the research on proof load testing, reliability-related issues arise. For example, if the reliability of the stop criteria is low, the test may be either terminated too early or too late. Since the latter results in the collapse of the structure, it is generally avoided. Therefore, stop criteria tend to be formulated conservatively—but one may question what safety margin is actually required. In the same way, a desire for target

loads arises that addresses specific structural and traffic situations. Aligning proof load testing procedures with internationally recognised standards, such as the reliability-based Eurocodes, can provide harmonisation while allowing flexibility to accommodate national requirements. Evidently, fundamental research on the probabilistic aspects of proof load testing and international collaboration is needed (ElBatanouny et al., 2019; Lantsoght, 2023; Schmidt et al., 2018).

1.3 Research goal and questions

This dissertation concerns fundamental research on structural reliability updating through proof load testing, particularly for reinforced concrete bridges and viaducts. In this probabilistic context, the aim is to prove that a bridge or viaduct is sufficiently reliable by withstanding the target load. A flexible probabilistic framework is proposed since the available information between structures varies, and the load test may be performed in various ways. The main research question is formulated as:

> *How can proof load testing establish the structural reliability of reinforced concrete bridges and viaducts?*

A number of challenges were identified in answering the main research question. To address these knowledge gaps, the following sub-questions, along with their motivations, are formulated:

1. *What influence do the functional lifetime, time-dependent effects, the moment at which the load test is performed and proven strength have on the target load?* By considering time dependence in the reliability analysis, all of these temporal aspects may be addressed. In this way, the reliability level can be accurately predicted as a function of time.

2. *How does the knowledge level (available information) of the structure change the way in which the target load can be calculated?* To account for the varying amounts and types of information available across different structures, a flexible approach is necessary. Possessing more details about the structure, on both loads and resistance, reduces the uncertainty and typically lowers the target load. A systematic way to combine data and subjective knowledge is required.

3. *How can in-situ monitoring data be utilised to directly update the structural reliability and indicate when to stop?* During a proof load test, it is common to perform measurements that give an indication of the structural performance, such as deflections, displacements and crack widths. Beyond evaluating stop criteria, in-situ monitoring offers valuable information about the resistance of the structure. This aspect links to the parallel research on stop criteria, emphasising its application in the broader context of reliability updating.

4. *How can spatial correlation and system reliability be addressed to overcome the practical limitations of testing a few cross-sections, spans, or even structures?* A

structure's reliability should be assessed at the system-level as it comprises multiple components. The properties of these components, such as cross-sections, structural elements, or spans, may be viewed as members of the same statistical population. Moreover, if multiple structures are built according to the same specifications, their performance could be related, providing additional knowledge.

A more detailed description of the context, the challenges and the suggested approach to address the above research questions may be found in De Vries et al. (2022) (Appendix A).

1.4 Research methodology

1.4.1 Fundamental research

This research centres around probabilistic methods to evaluate the structural reliability of reinforced concrete bridges and viaducts through proof load testing. As a first step, a literature study was carried out, which resulted in a state-of-the-art report on proof load testing and probabilistic approaches De Vries and Lantsoght (2022). Subsequently, the research questions were addressed in the order presented in the previous section. The phases of the research methodology, along with the thesis outline (Section 1.5), are schematically displayed in Figure 1.3.

The initial exploratory and conceptual phase of the research focused on the first two research questions. It aimed to find a flexible framework where proof load testing intersects with modern reliability concepts. To achieve this, proof load testing needed to be treated equivalently to other load types, with the emphasis shifted to the time-dependent nature of reliability and its underlying principle of conditional probability. Time-dependent analyses were carried out using Monte Carlo simulation and system reliability methods to create insight into the evolution of reliability. As time passes without structural failure, this evidence may be used to update our beliefs about the structure's strength and loading conditions, resulting in new reliability estimates. This process was more generally understood through Bayes' theorem, offering the required flexibility to address varying knowledge levels across structures and the subjective interpretation of probability. In this context, *informed* refers to the level of detail and certainty in the probabilistic description of the resistance. The first, fully informed analyses were compared with the outcomes of an uninformed approach solely based on load effects, and later extended to Bayesian analyses with low-informative prior resistance assumptions.

The second phase of the research addressed the remaining two questions by using the findings from the first phase. In the adopted Bayesian setting, the incorporation of in-situ monitoring data and laboratory measurements was studied to address the reliability of stop criteria. A relationship between observed crack widths and critical load levels was established by using measurement data previously obtained in the laboratory of TU Delft. Statistical inference of test results was used to identify a suitable indicator. Such indicators may entail a function of deflections, displacements, crack widths, strains, and so on. By combining the information obtained from measurements

with the survival of the proof load, realistic case studies revealed that the same level of reliability can be achieved with lower target loads. This transfer of knowledge from the laboratory to the structure under consideration, or between components in general, played a central role in the final topic concerning spatial variability and object similarity. By adopting a hierarchical Bayesian model, properties of components which are assumed to be members of the same statistical population can be inferred from prior knowledge and the censored data provided by the proof load test. The influence of the number of tested objects and population size is indicated by considering cases with varying degrees of correlation and variability. Using the developed Bayesian model, it is studied in what cases an increase in the target load can still demonstrate sufficient system-level reliability.

1.4.2 International perspective

The research questions and the presented methodology structure the fundamental research on the probabilistic aspects of proof load testing. However, as noted in Section 1.2.3, international collaboration is required to attain standardised assessment procedures. With this objective in mind, interim findings have been presented at the IABMAS 2022, ICASP14, and IABMAS 2024 conferences (De Vries et al., 2022; De Vries et al., 2023c; 2024). Further engagement with the scientific community was established through the publication of articles in the Transportation Research Record, Structure and Infrastructure Engineering, Engineering Structures and Structural Safety journals (De Vries et al., 2023a; 2023b; 2025b; 2026). In addition to notably improving the quality and direction of the research, international interaction also led to collaboration on reliability-based proof load testing in Italy (Addonizio et al., 2024). Preliminary findings have also been presented in the Transportation Research Board (USA) and *fib* COM3 TG3.1 (Europe) standardisation committees, as well as the Delft Reliability Exchange, to gather feedback and ensure alignment with future standards. Furthermore, international cooperation is fostered by the preparation of a joint technical report between IABMAS and *fib* COM3 TG3.2 on proof load testing.

1.5 Outline of the dissertation

This dissertation is structured around the research topics and questions formulated in Section 1.3. The first question is addressed in Chapter 2, and provides insight into how time dependence influences the assessment of existing structures and the effect of proof load testing. The varying state of information, as considered in the second research question, is examined through fully informed and uninformed procedures described in Section 2.3 and Chapter 3, to finally arrive at the Bayesian approach proposed in Chapter 4. With the adoption of the Bayesian method, the remaining two research questions regarding in-situ monitoring data and spatial variability may be addressed. The incorporation of monitoring data during the proof load test is described in Chapter 5, and the spatial correlation and object similarity are treated in Chapter 6. The dissertation ends with a discussion in Chapter 7, along with conclusions and re-

commendations for future research in Chapter 8. A visual representation of the journey through the topics and corresponding chapters is provided in Figure 1.3.

Since the outcomes of this research have already been published in scientific articles, each chapter only provides the crucial thoughts, results and conclusions. The dissertation is meant to synthesise the main findings and refer to the published articles for more details. The key publications are incorporated into the thesis as appendices (Appendix A to Appendix D), and the complete list of authored works may be found in the bibliography. The data and code underlying this dissertation have been prepared in line with the FAIR principles (Wilkinson et al., 2016), and are openly accessible through the 4TU.ResearchData repository (De Vries & Lantsoght, 2025a; 2025b).

Because of the concise nature of this dissertation, no elaborate background is provided on statistics, probabilistic methods and structural reliability. Although specific literature is cited where relevant, those unfamiliar with reliability concepts are referred to the excellent books by Ang and Tang (2006), Der Kiureghian (2022) and Madsen et al. (2006). An introduction to the Bayesian interpretation of probability may be found in Wasserman (2004) and its application to structural engineering in Benjamin and Cornell (2014) and Ditlevsen and Madsen (1996).

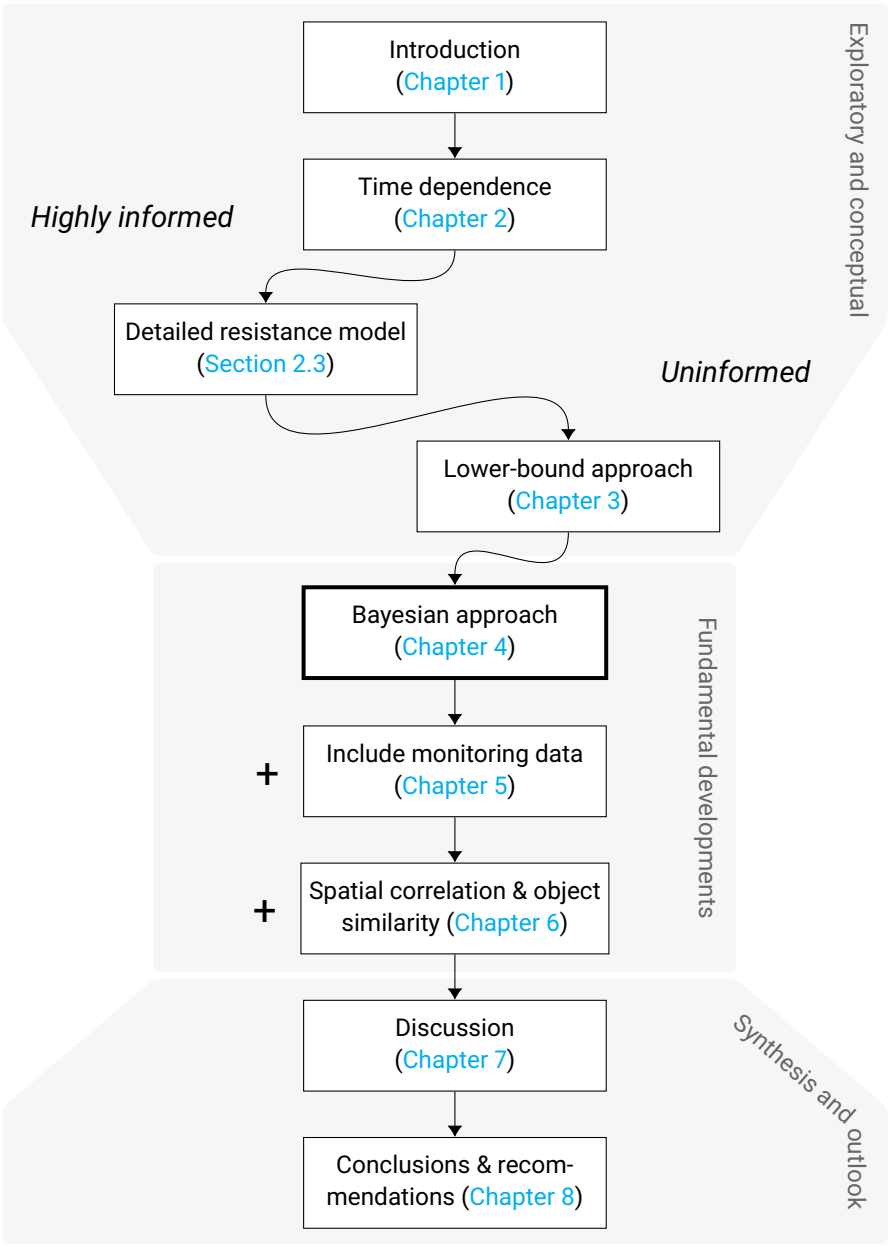


Figure 1.3 Flowchart of the methodology phases and topics within the dissertation, highlighting the transition from fully informed and uninformed to the Bayesian approach.

2

Time-dependent reliability assessment

Recognising that structural reliability is time-dependent due to variable loads, material degradation and changing traffic intensity, evaluating conditional annual reliability is proposed. By assessing annual reliability, assessments can accommodate the varying remaining service life between existing structures. Specific to proof load testing of bridges and viaducts, a limit state function is developed to incorporate traffic trends, progressive deterioration, and model uncertainty. A case study of a short-span concrete slab bridge shows a temporary dip in reliability during testing, followed by a substantial increase when the target load is survived. The suggested procedure enables strategic planning of proof load tests considering their timing and target loads. The method clearly demonstrates that reliability updates are not limited to structural resistance—all autocorrelated parameters are affected. Therefore, proof load testing should be considered in the generic context of reliability updating.

2.1 Annual reliability requirements

The loads acting on a structure are generally divided into permanent (G) and variable loads (Q). The distinction is made based on how often the load changes over time. For instance, the self-weight of a bridge is typically not going to change in time, although it is still a random variable—i.e. a structure's exact weight can only be established once it is built. In the case of variable loads, such as wind or traffic, the structural reliability becomes a function of time (Madsen et al., 2006). Additional time-dependent processes, such as load trends and material deterioration, add to this effect. Naturally, during a proof load test with large applied loads, the structural reliability is expected to be markedly lower than in normal operation (Spaethe, 1994).

Table 2.1 Annual reliability indices in line with current regulations in the Netherlands (De Vries et al., 2023d).

Consequence class	New structures	Renovation	Disapproval
CC1b	3.8	3.1	2.3
CC2	4.3	3.7	3.4
CC3	4.7	4.1	4.0

In general, the longer the reference period for the reliability analysis, the greater the probability of an undesired event occurring. Therefore, the desired remaining life span of the structure is of utmost importance for the reliability assessment. The target load levels used in proof load testing should reflect the projected life span of the structure. A fixed design reference period (e.g. 50 years) lacks the flexibility to accommodate the remaining functional life span of an existing structure. Moreover, the reliability level for new structures is generally higher than strictly required for human safety due to economic benefit. In De Vries et al. (2023d), annual reliability indices for the Netherlands are proposed in line with the current reliability requirements for regular reference periods (Table 2.1). Also internationally, annual reliability assessment is now recognised as a useful tool (ISO, 2015; Melhem et al., 2020).

The bridges and viaducts under consideration in this thesis are situated on Dutch highways, and as a result, their consequence class is CC3. Structures in the secondary road network would be classified as CC2. Since proof load testing is presented as a form of structural assessment, the disapproval level applies. Following the requirements in Table 2.1, the calculated annual reliability index should be equal to or larger than 4. The Dutch road authority Rijkswaterstaat also defines an intermediate 'usage' level between CC2 and CC3 (RWS, 2013) that corresponds with an annual reliability of 3.7. If the structure does not fulfil the requirement in any of the years of the projected life span, it should be renovated or replaced (NEN, 2011). Given the considerable variation in the remaining life of bridges and viaducts, it is necessary to evaluate their reliability appropriately. The proposed annual format enables this type of assessment, and the next section presents a conditional probability calculation method that can be used to demonstrate sufficient reliability.

2.2 Calculation method addressing time dependence

2.2.1 Conditional failure probability

Standard texts on reliability theory describe the proven strength from past loading followed by wear-out via the 'bathtub curve' of the failure rate (Jonkman et al., 2015). In a sense, every truck passing a bridge may be viewed as a test, contributing to the service-proven strength of the bridge (Wang et al., 2011). Instead of wear-out, deterioration and increased traffic loads will lead to lower structural reliability over time. The annual reliability is calculated under the condition that no failure has occurred in any of the

years before the year under consideration (Schmidt et al., 2020). By using the following events:

- A failure in the year i
- B failure in the years 1 to $i-1$
- B' no failure in the years 1 to $i-1$ (complement)

the conditional annual failure probability can be written as:

$$\mathrm{P}(A \mid B') = \frac{\mathrm{P}(A \cap B')}{\mathrm{P}(B')} = \frac{\mathrm{P}(A \cup B) - \mathrm{P}(B)}{1 - \mathrm{P}(B)} \tag{2.1}$$

The probability $\mathrm{P}(A \cup B)$ may be read as the cumulative failure probability up to and including the year i, whereas $\mathrm{P}(B)$ is the cumulative failure probability up to, but not including, the year i. These cumulative probabilities are easy to calculate when the statistical description of the load remains the same, but not when the description changes over time, e.g. due to increasing traffic loads or a proof load test. A Monte Carlo simulation can be used, but its application becomes challenging due to the large number of required samples with small failure probabilities.

Since conditional failure probabilities are related to the analysis of systems, established system reliability calculation methods can also be employed to calculate the annual reliability. To do so, first the reliability index and influence coefficients of each year need to be calculated, e.g. using FORM (Hasofer & Lind, 1974) or SORM (Breitung, 1984). The individual years are the system components in this calculation. Next, the cumulative failure probabilities can be calculated as OR-combined systems using the equivalent planes method (Roscoe et al., 2015). And finally, Equation (2.1) may be employed to determine the conditional annual failure probability. The standard transformation applies to obtain the corresponding (annual) reliability index:

$$\beta = \Phi^{-1}(1 - P_\mathrm{f}) = -\Phi^{-1}(P_\mathrm{f}) \tag{2.2}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal CDF. A more detailed description of system reliability literature and its application to proof load testing may be found in De Vries et al. (2022) (Appendix A) and De Vries et al. (2023b) (Appendix C).

2.2.2 Limit state functions

The limit state function for the probabilistic analysis of structural failure is formulated in terms of resistance (R) and the load effect (E), following JCSS (2015) and *fib* (2016). The load effect is split into the contributions from the dead load (G_DL), the superimposed dead load (G_SDL) and the variable load (Q). Both the resistance and load effect are associated with model uncertainty (θ_R and θ_E). Specific to the variable load is the time-invariant part of the variability (C_{0Q}). In addition, two random variables are added that account for the deterioration of the resistance (c_R) and trend in traffic load (c_Q):

$$Z = \theta_R c_R R - \theta_E (G_\mathrm{DL} + G_\mathrm{SDL} + c_Q C_{0Q} Q) \tag{2.3}$$

Except for the traffic load effect (Q), all random variables in the limit state function are *fully autocorrelated*. Although c_R and c_Q change with time, they are not stochastic processes since they reflect progressive deterioration and a gradual increase in traffic loading (Val & Stewart, 2019; Vu & Stewart, 2000).

The limit state function is subsequently adjusted to incorporate the proof load test event. When a proof load test is performed, an additional term (Q_{PL}) is included for the proof load effect in the limit state function:

$$Z = \theta_R c_R R - \theta_E [G_{\mathrm{DL}} + G_{\mathrm{SDL}} + \max(c_Q C_{0Q} Q, Q_{\mathrm{PL}})] \tag{2.4}$$

where the max-function is used to ensure that the regular traffic load is also considered for the year in which a proof load test is conducted. If a very low target load is used for proof load testing, it will thus have no effect. The adoption of the same model uncertainty for the load effect caused by traffic and by proof load testing is discussed in Section 2.4.

2.3 Case study

2.3.1 Description

The following case study demonstrates the time-dependent reliability assessment procedure for an existing structure subjected to a proof load (De Vries et al., 2023b) (Appendix C). The hypothetical structure under consideration is a concrete slab bridge with a relatively short span of $L = 10$ m. This type of bridge is prevalent in the Netherlands and also in many other countries (Christensen et al., 2022). It is assumed that the structure was built in 1960 and designed according to the prevailing standards of that time (KIVI, 1938; 1950). The traffic load used in its original design is inappropriate compared to today's high traffic intensity. But, the assumed values of material properties (e.g. steel and concrete strength) were quite conservative. As a result, old bridges and viaducts can still possess adequate structural strength to resist today's higher loads. To demonstrate the methodology, this case study will only consider the bending moment at midspan. In reality, the shear capacity of the slab near the supports and other bridge components will also require assessment.

In case the original bridge documentation, such as drawings and calculations, is still available, it may be used to infer the (prior) probabilistic description of the resistance. In this case, the bridge documentation is not available. Therefore, its design was 'reverse engineered' by using historic standards (Harrewijn et al., 2021). Conservatively, only the right-most lane is considered, as it is primarily used by trucks. This conservative approach does not take into account the distribution of forces that typically occurs across multiple lanes in a slab. Dynamic amplification is addressed through the uncertainty regarding the traffic load (C_{0Q}) with a mean value of 1.1; however, appropriate values vary from case to case. A top view and cross-section with the bottom reinforcement layout inferred from the historical standards are provided in Figure 2.1.

The bending moment resistance (R) is calculated from the equilibrium of normal forces in the cross-section when the reinforcement yields. Performing a cross-sectional

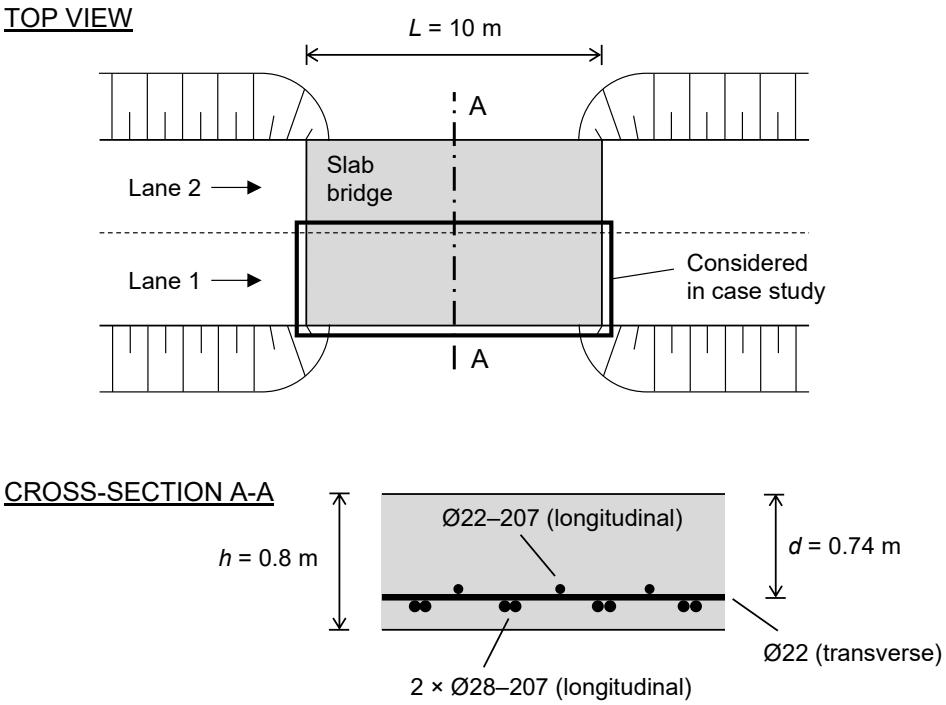


Figure 2.1 Layout of the case study slab bridge and the reverse engineered bottom reinforcement. Rebar size and spacing in mm. Reprinted with permission from De Vries et al. (2023b).

analysis with equivalent concrete stress block results in the following codified expression for the moment resistance (CEN, 2011):

$$R = A_s f_y \left(d - \frac{A_s f_y}{2 a_{cc} f_c b} \right) \tag{2.5}$$

where A_s is the cross-sectional area of the reinforcement over the width ($b = 3$ m), f_y is the yield strength of the reinforcement, and f_c is the concrete compressive strength. The concrete strength is reduced with a factor $a_{cc} = 0.85$ to account for long-term effects and possible unfavourable effects from the way the load is applied. It should be noted that the parameters in Equation (2.5) are random variables, and therefore, the customary design subscript 'd' is omitted. Their probabilistic description is provided in the next section.

2.3.2 Probabilistic model

The completion of the probabilistic model is achieved by assigning a distribution, mean value, coefficient of variation and autocorrelation coefficient to each of the random variables (Table 2.2). The chosen distribution types and parameter values are, where appropriate, based on JCSS (2015) and *fib* (2016). The description of the resistance fol-

Table 2.2 Random variables used in the limit state functions, including their description, distribution, coefficient of variation (COV) and Pearson autocorrelation coefficient.

Var.	Description	Distribution	Mean	COV	Auto-corr.
θ_R	Model uncertainty of the resistance	Lognormal	1	0.05	1
f_c	Concrete compressive strength (K250)	Lognormal	21.1 MPa	0.20	1
f_y	Reinforcement steel yield stress (QR24)	Lognormal	261 MPa	0.05	1
h	Height of the slab	Normal	0.8 m	0.02	1
a	Distance of the reinforcement to the surface ($a = h - d$)	Gamma	0.057 m	0.17	1
θ_E	Model uncertainty of the load effect	Lognormal	1	0.11	1
G_{DL}	Load effect of the dead load	Normal	721 kNm	0.05	1
G_{SDL}	Load effect of the superimposed dead load	Normal	101 kNm	0.1	1
C_{0Q}	Time-independent uncertainty of the variable load, including bias for dynamic load effect	Lognormal	1.1	0.1	1
Q	Load effect of the traffic load; annual maximum	Gumbel	1150 kNm	0.025	0
t_{R0}	Initiation time to deterioration in years	Lognormal	20	0.1	1
Δc_R	Degradation per year	Lognormal	0.0025	0.1	1
c_{Q0}	Starting value of the trend	Lognormal	0.78	0.1	1
Δc_Q	Increase of traffic load per year	Lognormal	0.004	0.1	1
Q_{PL}	Load effect of the proof load	Normal	1800 kNm & 2000 kNm	0.01	1

lows from the properties of material strength and geometry. The material properties are based on historical information about the K250 concrete and the smooth (not ribbed) QR24 rebars used at the time. The statistical description of the traffic load effect, i.e. the bending moment at midspan, is based on WIM data from 2015. Only the traffic in the right-most lane, where the trucks drive, has been analysed. To reflect the lower traffic loads in the 1960s, the corresponding trend factor (c_Q) is significantly lower at the beginning of the time-dependent analysis.

The area of the reinforcing steel (A_s) is not included as a random variable because its variability is typically very small. Also, note that the resistance model uncertainty (θ_R) is intended to account for any remaining uncertainty between modelling and reality. If corrosion of the reinforcement plays a significant role, the reduction of the reinforcement area should be included to update reliability predictions (Jacinto et al., 2015). A separate and more general deterioration parameter c_R is incorporated in this case

study. Specifically, the following relations hold for the time-dependent deterioration and load trend coefficients:

$$c_R(t) = \begin{cases} 1 & t \leq t_{R0} \\ 1 - \Delta c_R (t - t_{R0}) & t > t_{R0} \end{cases} \tag{2.6a}$$

$$c_Q(t) = c_{Q0} + \Delta c_Q \, t \tag{2.6b}$$

where the parameters are random variables (listed as well in Table 2.2). To demonstrate the method in Section 2.2, the mean values of the parameters were chosen in such a way that the annual reliability is insufficient around the year 2020. Next, an assessment of the structure should be conducted, and in this case, it will be done through proof load testing.

2.3.3 Results

With the completed model description, the conditional failure probability calculation procedure and limit state equations (Section 2.2) may now be employed to perform the time-dependent reliability analysis. This can be achieved through Monte Carlo simulation, but system reliability methods may also be used to reduce computational effort. In the latter case, reliability analyses are conducted for each year separately using FORM (Hasofer & Lind, 1974) or SORM (Breitung, 1984; Hohenbichler et al., 1987; Tvedt, 1983), after which the individual outcomes are incrementally combined (Gong & Zhou, 2017; Roscoe et al., 2015). Comparisons with Monte Carlo simulations indicated an acceptable difference of about a tenth in the annual reliability index.

The result of the first calculation, without any proof load test, is displayed in Figure 2.2. The base case considers the situation without traffic trends and degradation. In this case, the annual reliability increases gradually due to the proven strength of past traffic loads. The traffic trend and degradation are incorporated subsequently to dis-

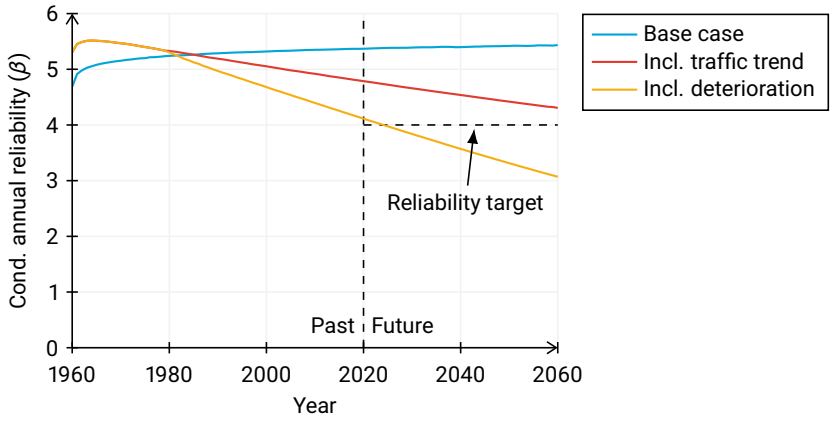


Figure 2.2 Development of the conditional annual reliability with time. The influence of the traffic load trend and deterioration of the resistance are shown separately.

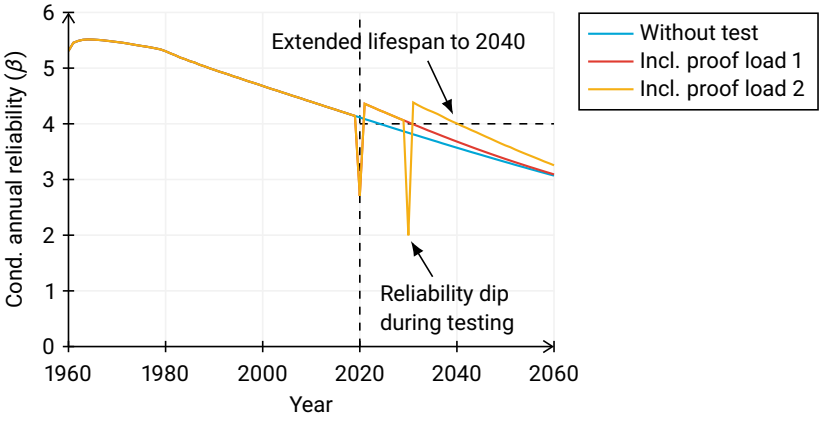


Figure 2.3 Effect of proof load testing on the annual reliability. The target loads are chosen such that the annual reliability remains sufficient in the next 10 years.

play their detrimental effect on the evolution of the annual reliability. Higher reliability is calculated in the first years when including the traffic trend because the adopted linear trend expresses a load reduction before 2015 and an increase afterwards.

Next, proof load testing is adopted to evaluate whether the bridge meets the required annual reliability (Figure 2.3). The first proof load test is performed in the year 2020 and has a target (mean) value of 1800 kNm. Then, in 2030, the second proof load test is performed with a higher target load effect of 2000 kNm. The annual reliability is markedly lower in the year the test is performed, but if the test is survived, the reliability in the following years is increased. The target loads have been determined in an iterative fashion such that the annual reliability remains above the target in the next 10 years. Alternatively, the higher target load could have been applied directly in 2020, also leading to sufficient reliability until 2040; however, the probability of failure during the test in 2020 would be higher. The proposed time-dependent calculation procedure can thus be used to determine the ideal strategy, balancing risks in testing with the remaining functional lifespan of the structure.

It is observed that the reliability indices during proof load testing are below the acceptable target reliability for normal operation. Hence, further safety measures are necessary during the load test, such as closing off the area underneath the viaduct or bridge. In addition, instrumenting the structure and evaluating the stop criteria after each load cycle can avert damage and collapse. Careful monitoring increases reliability during the test, but the target load should still be attained to demonstrate sufficient reliability. Lower test loads may suffice when in-situ measurements can be used to infer the resistance and update reliability predictions (Chapter 5).

2.4 Gained insight

When variable loads act on a structure, such as traffic loads, structural reliability inherently becomes a function of time. In addition, degradation of resistance and trends in

loading also lead to time-varying reliability. Annual reliability indices were derived to offer the flexibility needed for the varying lifespans of existing structures. The proposed conditional probability calculation method enables assessment of the structure's reliability at any point in time, under the condition of no past failure. Although the reliability calculation can be performed using Monte Carlo simulation, computationally less demanding alternatives such as FORM and SORM, applied within a system-reliability procedure, also give sufficiently accurate results. Limit state functions were formulated based on resistance and load effects, including model uncertainty and additional factors that account for deterioration and traffic trends.

The case study on a short-span concrete slab bridge built in 1960 illustrated the application of a time-dependent reliability analysis. It is shown that the decrease of conditional reliability may follow from progressive deterioration and a gradual increase in traffic loads. The reliability temporarily decreases significantly during a proof load test, but if the test is survived, the conditional reliability improves for the remainder of the lifespan. The time-dependent analysis allows for strategic test planning—ensuring annual reliability requirements are met throughout the projected lifespan. A detailed resistance model enables the failure probability calculation during the test and, combined with load-test monitoring, facilitates risk mitigation for personnel and equipment.

The key insight following from studying the time dependence of reliability is that a generally applicable conditional probability calculation method eliminates the need for manual alteration of the resistance distribution following a proof load test. In the proposed formulation, a proof load test is plainly incorporated as a momentarily high variable load. As with regular variable loads, the absence of structural failure over time serves as evidence to update our beliefs about the structure's strength and loading conditions. Consequently, not just the resistance distribution is updated, but the distributions of all autocorrelated random variables, including the model uncertainties. Proof load testing should therefore be considered in the context of structural reliability updating, allowing for statistical inference during and following the proof load test.

3

Lower-bound approach based on the load effect

If limited information on a structure is available or significant concerns exist regarding its condition, proof load testing may be used to establish a lower bound to the structural reliability. A conservative approach based solely on the load effect serves as the background for current standards and guidelines on proof load testing, particularly the Manual for Bridge Evaluation (MBE) used in the USA. In this procedure, the proof load is expressed by the regular traffic load model magnified by the target proof load factor. Of particular interest is the reliability level achieved as the target proof load factor is varied. In this chapter, traffic data and reliability requirements from the Netherlands are used to demonstrate reasonably constant proof load factors across various span lengths for both bending and shear. Although the lower-bound approach is easily derived and practical, several conceptual challenges remain.

3.1 Objectivity and standardisation

In many cases, the level of information is insufficient to perform informative analyses regarding the evolution of structural reliability as bridges age, as illustrated in the previous chapter. Older bridges typically lack the necessary documentation, and their resistance may have deteriorated over time. In practical applications, there is a strong desire for objectivity concerning the available information and condition of the structure (Casas & Gómez, 2013). In this regard, proof load testing can play a key role in verifying the current capacity and condition of the structure. Using design loads from normative standards as target loads might appear logical, but their contextual application differs considerably. Design loads are calibrated for new structures with high safety margins, economic considerations and long lifespans. In addition, their derivation assumes a

simultaneous occurrence of incidentally low structural resistance and high load effect (Madsen et al., 2006). This situation does not reflect the actual conditions encountered during proof load testing of existing structures, where the resistance and permanent loads are no longer hypothetical but actual values (realisations). Clearly, there is a need for a specific, standardised approach to define appropriate target load (Section 1.2.1) levels that suit the proof load testing context.

In this chapter, a lower-bound approach is described which makes use of the key consideration that the true resistance of the structure is equal to, or larger than, the load effect achieved during the proof load test: $R \geq G + Q_{\mathrm{PL}}$. It will be shown that this information can be used to formulate a method that is only dependent on the load effects of the traffic and the proof load test. Following design practice, the target proof load effect is expressed by the characteristic value of the traffic load model (Q_{k}) magnified by a partial safety factor, i.e. the proof load factor (γ_{PL}):

$$Q_{\mathrm{PL}} = \gamma_{\mathrm{PL}} Q_{\mathrm{k}} \tag{3.1}$$

The same rationale serves as the background for current standards and guidelines on proof load testing, particularly the Manual for Bridge Evaluation (MBE) used in the USA (AASHTO, 2022). In the MBE, the factor γ_{PL} corresponds to the parameter X_{p}, with a default value of 1.4, and may be adjusted to reflect bridge-specific circumstances. The definition of the characteristic traffic load effect (Q_{k}) depends on the context and may refer to either a statistically derived value or one calculated using a codified load model, with the latter adopted in this chapter. Further details about the background to the MBE method (Lichtenstein, 1993) and the proposed improvements, leading to the up-to-date formulation presented here, may be found in the original article on which this chapter is based: (De Vries et al., 2023a) (Appendix B).

3.2 Up-to-date formulation

To arrive at an up-to-date formulation of a lower-bound approach assuming no resistance information, a fully probabilistic treatment of all variables is required. In contrast to Lichtenstein (1993), it is unnecessary to make deterministic assumptions or explicitly differentiate between, for example, mean and characteristic (nominal) values. Additionally, model uncertainties should be introduced to acknowledge that our engineering knowledge is incomplete and account for unfortunate differences. The proposed formulation is similar to the method used to determine optimal proof load levels in Appendix E of Kapoor (2021). To complete the up-to-date approach, modernising the traffic load description and adopting the appropriate load model is required. In Section 3.2.2, it is described how Dutch traffic data is utilised to obtain a suitable statistical description of the load effect.

3.2.1 Probabilistic model

After a successful proof load test, it is known that the resistance must be equal to or larger than the load effect following from permanent loads and the applied target load

($R \geq G + Q_{\mathrm{PL}}$). With this fundamental ingredient, the limit state function may be re-written to exclude the resistance and permanent loads. In essence, the probability of failure of the structural part, or cross-section, is directly reformulated into the probability that a future traffic load effect (including dynamic amplification) exceeds the load effect produced during the proof load test:

$$\begin{aligned} Z &= R - (G + Q) \\ &= (G + Q_{\mathrm{PL}}) - (G + Q) \\ &= Q_{\mathrm{PL}} - Q \end{aligned} \tag{3.2}$$

where the resistance (R) and permanent loads (G) are eliminated, and the load effect survived during the proof load test (Q_{PL}) and the traffic load effect (Q) remain. Since the resistance term has been removed, potential degradation cannot be explicitly accounted for and must be compensated by requiring a higher target reliability.

The dynamic amplification of the traffic load effect (impact) varies significantly with span length and specific bridge configuration (Deng et al., 2015). The probabilistic analysis used to determine target load factors is therefore based on traffic loads with no, or only minimal, dynamic contribution, such as WIM data. Target loads obtained from this procedure should consequently be multiplied by an appropriate dynamic amplification factor. Since the resulting target loads are comparable to design loads in terms of extremity, design procedures that account for dynamic effects (González et al., 2009) are deemed suitable here as well.

Missing in the limit state function of Equation (3.2) are model uncertainties. Our understanding of the translation from applied loads, in a test or from actual traffic, to the load effect is limited. The degree of uncertainty depends on the level of sophistication incorporated in the analytical or FEM model. Additional uncertainty stems from the statistical modelling of the load effect—i.e. the assumed distribution functions and statistical inference process. The variability of the traffic load effect may be split into a time-invariant (C_{0Q}) and time-variant part (Q) (*fib*, 2016). By including the model uncertainties and splitting the traffic load uncertainty, the limit state function becomes:

$$Z = \theta_{E,\mathrm{PL}} Q_{\mathrm{PL}} - \theta_E C_{0Q} Q \tag{3.3}$$

where $\theta_{E,\mathrm{PL}}$ and θ_E are model uncertainties associated with the load effect.

Table 3.1 Overview of variables in the limit state function for the lower-bound approach, considering only variable load effects and their model uncertainties.

Var.	Description	Distribution	Mean	COV
$\theta_{E,\mathrm{PL}}$	Model uncertainty of the proof load effect	Lognormal	1	0.1
Q_{PL}	Load effect achieved by proof load	-	(varies)	0
θ_E	Model uncertainty of the traffic load effect	Lognormal	1	0.11
C_{0Q}	Time-invariant uncertainty of the traffic load	Lognormal	1	0.1
Q	Annual maximum of the traffic load effect	Gumbel	(varies)	(varies)

An overview of the variables in the limit state function is provided in Table 3.1. Their statistical properties are based on general recommendations for probabilistic modelling (*fib*, 2016; JCSS, 2015). The model uncertainty regarding the load effect produced in the proof load test ($\theta_{E,\mathrm{PL}}$) is based on the model uncertainty associated with the traffic load. Because the conditions are more controlled during a test, a lower variation coefficient may seem more appropriate. However, when viewed as a resistance parameter, it should also cover the uncertainty in selecting the most critical locations to test. This issue is alleviated when using a moving vehicle for the test, but some uncertainty remains regarding the transverse location and axle configuration. For simplicity, the proof load effect (Q_{PL}) is assumed to be a deterministic value, although the exact applied load will naturally vary slightly. In addition, the correlation between the model uncertainty of the proof load effect produced during the test and regular traffic has been neglected. These aspects are challenging to quantify, and the simplifications offset each other in terms of calculated reliability.

3.2.2 Load effect analysis using WIM data

Weigh-in-motion (WIM) data are well-suited to obtain an accurate statistical representation of the traffic load effect. In the Netherlands, WIM recording stations are positioned at several traffic-intensive highway locations. Traffic simulations have been performed to obtain the maximum bending moment at midspan and the maximum shear force near the supports of a simply-supported span using WIM data from 2015. These load effects derived from WIM data exhibit minimal or no dynamic load amplification. Over a period of one year, various block maxima may be determined: hourly, daily, weekly and so on. Given the difference in traffic between weekdays and weekends, a week is an appropriate cycle, yielding 52 data points. Subsequently, a Gumbel distribution is fitted to the data, with an exceedance probability threshold ($S = 0.25$) to capture the right tail. Figure 3.1 shows the fitted distribution to the data points of

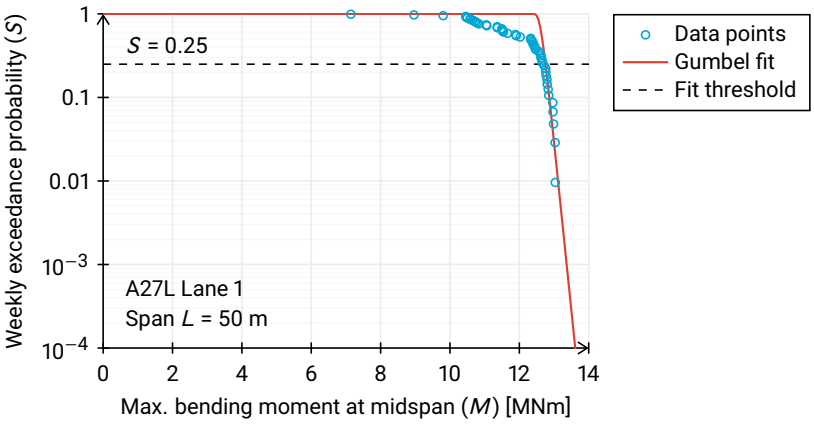


Figure 3.1 Gumbel fit of the load effect data points for the maximum bending moment at midspan of a simply-supported span.

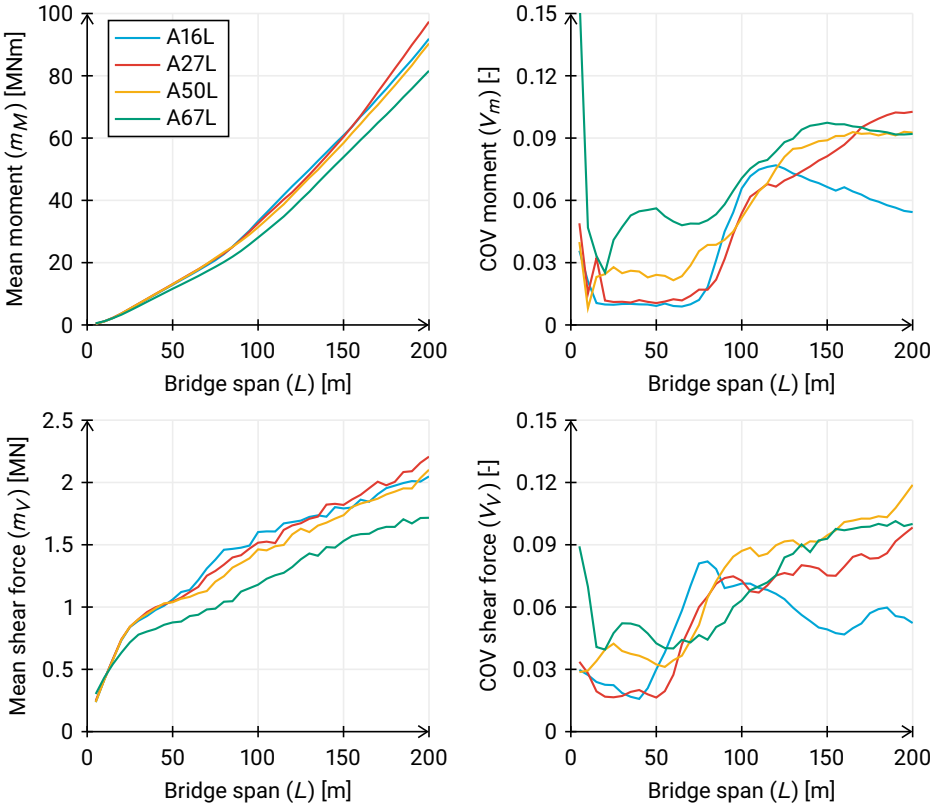


Figure 3.2 Parameters of fitted Gumbel distributions for the annual maximum load effect of the bending moment at midspan and shear force near the supports.

the maximum bending moment of Dutch highway A27L lane 1, the rightmost lane, occupied mainly by trucks.

Because the weekly maxima are sufficiently uncorrelated, the Gumbel distribution may be converted to annual maxima by shifting the location parameter via $\mu_a = \mu_w + \beta \ln(52)$ where β is the scale parameter and 52 is the number of weeks in a year. The mean and coefficient of variation of the distribution are obtained as $m = \mu + \beta\gamma$ and $V = (\beta\pi/\sqrt{6})/m$ where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant (Gumbel, 1954). Distributions have been fitted for various WIM datasets and span lengths, as displayed in Figure 3.2. The analysed roads show a comparable trend in mean and coefficient of variation with span length. In a reliability analysis, the average of the four different roads is used (plus model uncertainty θ_E, see Table 3.1).

3.3 Calculated proof load factors

To study the relationship between span length and target proof load factor, calculations are made with the probabilistic model (Section 3.2.1) and traffic data from the Neth-

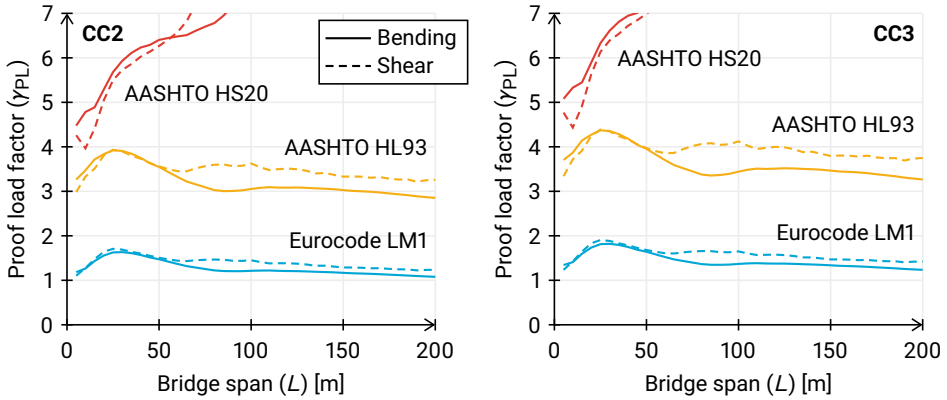


Figure 3.3 Relation between the span length and target proof load factor considering unfactored load models in the bending and shear, for the disapproval of CC2 and CC3 structures (i.e. target annual reliability $\beta = 3.4$ and $\beta = 4$).

erlands (Section 3.2.2). Per span length, two iterative probabilistic analyses are performed: one considering the bending moment at midspan, and the other considering the shear force near the supports. Consequence classes 2 and 3 of EN 1990 (CEN, 2019) are considered with existing-structure annual reliability target indices of 3.4 and 4.0, respectively (Chapter 2). The distributed load of the Eurocode LM1 and the AASHTO HL93 traffic load model are applied over a lane width of 3 m. The proof load factors following from the reliability analyses are displayed in Figure 3.3.

It is observed that the calculated target proof load factor is considerably larger when using the AASHTO HS20 and HL93 load models in comparison to Eurocode LM1. This follows from the relatively high unfactored load effect following from LM1. A quick comparison of the axle loads signifies the difference: 300 kN for LM1 versus 145 kN (32 kip) for HS20 and HL93. The traffic load from the Netherlands is relatively high compared to other countries (O'Brien et al., 2012) and meshes with the high loads of Eurocode LM1, leading to moderate values of the proof load factor. Another observation is the continuously increasing proof load factor with span length when the HS20 load model is used, an effect attributed to the single-truck schematisation. The issue is overcome by the HL93 load model, which also includes a distributed lane load. For both the Eurocode LM1 and the HL93 load model, an almost constant factor is obtained over various spans. Only around $L = 30$ m a relatively large factor is required. This may be explained by the presence of long and heavy vehicles (an oversize load for which an exemption must usually be requested) that are not accurately represented by the load model. Further research is required to decide whether and how they should be excluded from the WIM dataset. In addition, vehicles travelling at a speed lower than 40 km/h are not present in the dataset. The effect of congestion is expected to result in higher proof load factors for larger bridge spans.

3.4 Reflection on the method

The rationale behind the lower-bound approach, i.e. the resistance is larger or equal to the self-weight and the applied proof load effect, is both clear and practical. The proposed up-to-date probabilistic formulation results in acceptable proof load factors with reliability requirements suitable for existing structures. The relatively straightforward modelling approach is transparent and suitable in a conservative code-making setting. The target load may be expressed through a proof load factor applied to the customary design load model, such as Eurocode LM1, with its magnitude directly linked to the desired level of reliability. To account for dynamic effects, a suitable dynamic amplification factor must be applied separately. The analyses show that the proof load factor remains reasonably constant across different span lengths. Still, it is preferable to let the factor depend on span or influence length, rather than adopting a single conservative value.

While a method based on variable loads alone is appealing because no assumptions on the resistance need to be made, several challenges remain. As shown previously (Chapter 2), structural reliability is inherently time-dependent. Demonstrating sufficient reliability at the time of testing does not guarantee the structure will remain reliable for decades to come. Testing to prove reliability levels above the minimum may be required to compensate for potential degradation over the structure's remaining service life. With the high proof loading levels following from the conservative lower-bound approach, the probability of failure during testing becomes a concern. However, without resistance information, this probability cannot be calculated. In addition, the bridge or viaduct must be viewed as a system comprising multiple components and critical cross-sections, each contributing to the failure probability and requiring assessment (Chapter 6).

The required knowledge regarding potential resistance degradation and the identification of critical components and cross-sections directly contradicts the primary appeal of the lower-bound method—namely, that it avoids resistance assumptions. This requirement highlights that in a realistic proof loading scenario, some information about the structure's resistance is inevitably required. Moreover, highly valuable information can be obtained by monitoring the structure's response during the test (Chapter 5). In this context, Bayesian methods can provide the means to incorporate all these data, and will be discussed in the next chapter.

4

Bayesian approach to address varying knowledge levels

Even if only limited information about a structure's resistance is available, that information can still be incorporated into structural assessments through proof load testing. In the proposed Bayesian methodology, all available information regarding the structure is incorporated, even when it is subject to uncertainty. This chapter demonstrates how basic information about a structure, combined with proof load tests and past traffic loads, can be used to establish the reliability of a structure. A sensitivity study on prior distribution types quantifies the impact of low-informative resistance formulations. Their influence is examined in a time-dependent Bayesian analysis, which motivates the use of higher target loads and reveals an increased failure probability during proof load testing due to limited prior knowledge. Ultimately, the Bayesian methodology forms a basis for developing flexible approaches to proof load testing that can incorporate monitoring data and account for spatial correlations.

4.1 Incomplete knowledge in reliability assessment

As an assessment method for existing structures, proof load testing is typically applied when structural knowledge is incomplete or lacking. The amount and quality of information available before testing can vary significantly between structures. For the most accurate estimates of structural reliability, all relevant information should be incorporated. The benefit of adopting a Bayesian approach is that it allows for the inclusion of information subject to uncertainty, together with a prior belief about that uncertainty. This places proof load testing within the broader framework of structural reliability updating, where multiple information sources are combined, often sequen-

tially, to improve reliability estimates progressively (Benjamin & Cornell, 2014; Straub & Papaioannou, 2015).

The adoption of Bayesian methods represents a true paradigm shift, as mechanical resistance models are no longer viewed as essential representations of structural behaviour but are instead interpreted as sources of information. Still, knowledge of structural mechanics, including the calculation of load effects and the identification of critical failure modes, remains required. This approach enables the description of resistance in a purely statistical manner, e.g. derived from laboratory tests (Chapter 5), but partial knowledge, for example regarding system behaviour and spatial correlation, can also be incorporated (Chapter 6). If little is known about the resistance, a weakly or low-informative prior may be formulated based on basic, physically justified information about the structure (Ditlevsen & Vrouwenvelder, 1994). Bayesian methods provide a middle ground between the highly informative setting, in which precise structural information is available, and the conservative lower-bound estimation considered in the previous chapters. More details about the proposed Bayesian approach to address varying knowledge levels in proof load testing may be found in De Vries et al. (2022) (Appendix A) and De Vries et al. (2023b) (Appendix C).

4.2 Bayesian inference following testing

4.2.1 Bayesian inference

In many applications, it is necessary to infer model parameter values from a limited set of observations or data. Doing so provides insights into the workings of our abstractions of reality (models) by aiding their development, allowing for calibration, and quantifying their uncertainties. This process of statistical inference, aimed at improving prediction for unobserved data, has long been employed across many application fields (Geisser, 1993). Bayesian inference is the process of performing statistical inference via Bayes' theorem, which has a broader application scope (Jeffreys, 1961):

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\,\mathrm{P}(A)}{\mathrm{P}(B)} \tag{4.1}$$

where A and B are events and $\mathrm{P}(B) > 0$. Events are subsets of the sample space, defined as collections of outcomes to which a probability is assigned, and a single outcome may be part of various events. In a statistical inference context, the event for which the probability is established may be understood as the hypothesis ($A = H$)—for example, a parameter of the model taking on a specific value—and the conditioning event represents the data that has become available ($B = D$). The data consists of one or more possible outcomes of the model observed in real life and may take on many forms (e.g. a numeric value, a flag indicating failure, etc.). Because of this generic nature, various types of data (or evidence) can be combined.

In Bayesian inference, beliefs about model parameters ($\boldsymbol{\theta}$) are typically expressed through distributions. These distributions may be discrete or continuous, and describe the range of plausible parameter values along with their associated probabilities. Prior

distributions express domain knowledge or previous findings before incorporation of the data ($\mathbf{x}$). To evaluate how likely the observed data are given certain combinations of parameter values, a likelihood function is invoked. The probability mass or density functions of the assumed distribution are a common choice, but in essence, any function expressing the data-parameter relationship suffices. The posterior distribution, reflecting both prior beliefs and knowledge gained from the data, is expressed as (Box & Tiao, 1973):

$$p(\theta \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \theta)\, p(\theta)}{p(\mathbf{x})} \quad \Leftrightarrow \quad p(\theta \mid \mathbf{x}) \propto p(\mathbf{x} \mid \theta)\, p(\theta) \tag{4.2}$$

where $p(\mathbf{x} \mid \theta)$ indicates the likelihood, $p(\theta)$ the prior distribution, and $p(\mathbf{x})$ the marginal likelihood. As the likelihood essentially acts as a normalisation constant, one is rarely interested in its particular value. Therefore, the notation with a 'proportional to' symbol ($\propto$) is generally preferred.

A common illustrative case involves a parameter vector comprising the mean and standard deviation of a normal distribution, $\theta = (\mu, \sigma)$. Before the analysis, each of these parameters is assigned a distribution separately, or jointly, expressing prior beliefs. The likelihood function then corresponds to the probability density function of the normal distribution, but the formulation of the prior, $p(\theta) = p(\mu)\, p(\sigma)$, is less straightforward, as will be discussed next.

4.2.2 Prior probability

An important part of the Bayesian method is the specification of the prior probability, which expresses the belief about an uncertain quantity before any evidence is considered. If subjective influences are undesired, objective or uninformative priors may be adopted. These priors, sometimes called unbiased, diffuse, or flat, represent our ignorance about a parameter. Unfortunately, no single distribution eliminates subjectivity in all cases, as its formulation depends on the parameter and inference problem (Figure 4.1). The main issue is that many priors are not transformation invariant: if the mathematical formulation is chosen differently, a prior can become more informative. Several methods have been put forward to obtain uninformative priors. For instance, Jeffreys' prior is defined as $p(\theta) \propto \sqrt{|I(\theta)|}$, where $I(\cdot)$ is the Fisher information. However, if θ contains a pure location parameter, Jeffreys proposed a modified rule based on translation invariance, yielding a constant prior for such a parameter, e.g. $p(\mu) \propto 1$ (Jeffreys, 1961). Alternative non-informative priors include those derived from Kullback-Leibler divergence and the Haldane prior (Kass & Wasserman, 1996).

For a normally distributed quantity, the non-informative priors correspond to $p(\mu) \propto 1$ and $p(\sigma) \propto 1/\sigma$. The resulting posterior predictive distribution (Section 4.2.3) has a convenient closed form and is commonly used in statistical inference (Section 5.2). The main strength of the Bayesian method, however, lies in its ability to incorporate prior knowledge and to represent statistical uncertainty across a wide range of distribution types and use cases.

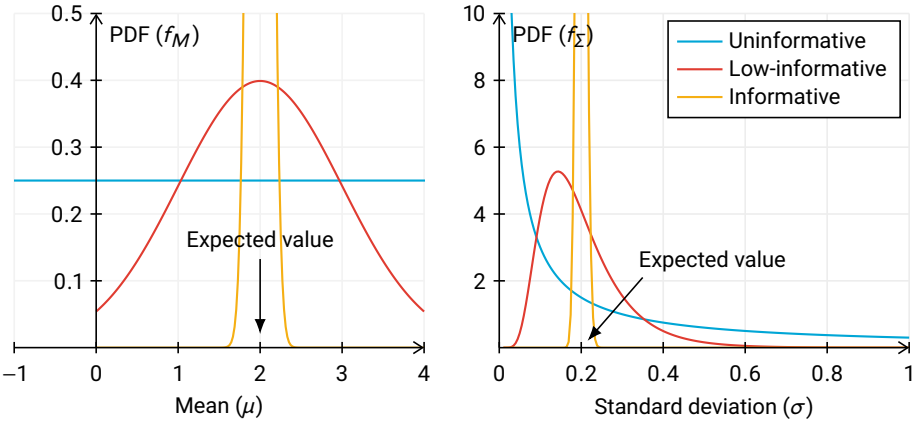


Figure 4.1 Prior distributions with varying degrees of belief for the mean and standard deviation of a normally distributed variable. The low-informative priors have a coefficient of variation (COV) of 0.5, while the informative priors have a COV of 0.05.

At the other end of the spectrum, for well-known quantities, strongly informative priors can be used. These priors incorporate existing domain knowledge, such as results from previous experiments, a detailed understanding of the underlying mechanics, or reflect precisely controlled environments. Typically, these are strongly peaked distributions with a small coefficient of variation. However, care must be taken to ensure that the prior is applicable to the current context, and it should be recognised that overly restrictive priors can bias results—as they can outweigh the information contained in the data. Conversely, in the sparse-information regime, uninformative priors chosen for mathematical convenience can describe physically unrealistic or even impossible situations. Therefore, even when significant knowledge is lacking, low-informative priors are generally preferred over uninformative formulations (Ditlevsen & Vrouwenvelder, 1994).

4.2.3 Posterior predictive distribution

Once the prior distribution $p(\theta)$ has been updated on the basis of data $\mathbf{x}$ to yield the posterior distribution $p(\theta \mid \mathbf{x})$, the resulting distribution can be used in various ways. In some applications, it is desired to directly update the original probability distribution of random variable X. To obtain this posterior predictive distribution, or Bayes' distribution, the uncertainty regarding the parameters is integrated into the probability distribution via:

$$p(x \mid \mathbf{x}) = \int_{\Theta} p(x \mid \theta)\, p(\theta \mid \mathbf{x})\, \mathrm{d}\theta \tag{4.3}$$

where $p(x \mid \cdot)$ can refer to the probability density or cumulative distribution function of X (Benjamin & Cornell, 2014). Typically, numerical methods are required to evaluate the integral, and in many cases it is convenient to combine the Bayesian analysis, yielding the posterior distribution, with the integration of uncertainties (Section 4.3).

4.2.4 Sample testing versus proof load testing

Although related, inferences from sample testing and proof load testing are not equivalent. Both approaches rely on observed data to update probabilistic models, but their objectives and implementation differ substantially in a structural reliability context:

- **Sample testing**: The objective is to infer the properties of a population from a representative subset of specimens. The analysis often employs a hierarchical Bayesian model, in which the test results inform the parameters of the population distribution, while the distribution type remains unchanged. Sample testing typically focuses on a single structural component, after which designs incorporating such components may use the statistically derived characterisation, e.g. see Annex D of EN 1990 (CEN, 2019).

- **Proof load testing**: This approach is targeted at assessing the performance of a specific structure or component. Hierarchical modelling is of limited relevance here, except with partially-informed priors and cases where results from multiple (correlated) components are to be combined. Updating is performed at the distribution level, and as a result, the shape of distributions typically changes. In a proof load test, the structural system is subjected to a test load, and all relevant distributions are updated based on the observed performance.

The distinction between the two approaches can blur when rather than loading specimens to failure in sample testing, a maximum test load is specified. Such tests generate right-censored data, which can reduce testing effort and allow the tested specimens to be placed in service for their intended purpose. (In some production processes, acceptance tests are performed in which all products are tested in this way, as part of quality control.) However, when tested elements are subsequently put into use, they are typically subjected to new boundary conditions. Compared to proof load testing, these new operating conditions limit the value of the test, since their performance as part of the system is not evaluated.

4.3 Bayesian reliability updating method

4.3.1 Principle and sequential updating

The Bayesian reliability updating method extends the principles of Bayesian inference to the evaluation of structural reliability. In this context, the objective is to determine an updated reliability, given new evidence such as test results, inspections, or the outcome of a proof load test. No mechanical model is strictly required, although such models can provide valuable prior information. A key advantage of the Bayesian approach is the ability to sequentially update reliability estimates by considering additional data and using the posterior distribution from one stage as the prior distribution for the next. Once the desired reliability is achieved, data collection can stop— provided the evidence collected thus far offers an unbiased view (Ditlevsen & Madsen, 1996; Straub & Papaioannou, 2015).

Conceptually, the reliability calculation procedure is related to the posterior predictive distribution (Section 4.2.3). To obtain the failure probability, the limit state function, $g(\cdot)$, is evaluated over the distribution of the basic random variables ($\mathbf{X}$), weighted by the posterior distribution of the Bayesian model parameters ($\boldsymbol{\theta}$):

$$P_\mathrm{f} = \int_\Theta \int_\mathcal{X} \mathbb{1}[g(\mathbf{X}) < 0]\, p(\mathbf{X} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathbf{x})\, \mathrm{d}\mathbf{X}\, \mathrm{d}\boldsymbol{\theta} \tag{4.4}$$

where $\mathbb{1}[\cdot]$ is the indicator function, and $\mathbf{x}$ denotes the data. The corresponding reliability index is obtained as $\beta = -\Phi^{-1}(P_\mathrm{f})$ by definition, where $\Phi(\cdot)$ indicates the standard normal cumulative distribution function. Owing to its predictive nature, the outcome is an estimate reflecting statistical and model uncertainties, not a true failure probability (which is difficult to establish for civil engineering structures).

The double integral in Equation (4.4) reflects that the basic variables $\mathbf{X}$ depend on the Bayesian model parameters $\boldsymbol{\theta}$, which is important for the order in which random variables are sampled. However, the distinction between $\mathbf{X}$ and $\boldsymbol{\theta}$ disappears if the updated distributions are not only model parameters, but also basic variables—as is the case in proof load testing. In this context, it is convenient to remove the distinction, such that $\boldsymbol{\theta}$ carries the entire state of the system, including any unconditioned variables:

$$P_\mathrm{f} = \int_\Theta \mathbb{1}[g(\boldsymbol{\theta}) < 0]\, p(\boldsymbol{\theta} \mid \mathbf{x})\, \mathrm{d}\boldsymbol{\theta} \tag{4.5}$$

This integral motivates numerical implementations that perform Bayesian updating and calculate failure probabilities in a single pass, without explicitly evaluating or storing posterior distributions (Section 4.3.3).

4.3.2 Application to proof load testing

To apply Bayesian reliability updating to the proof load testing situation, first, a description of the reliability problem is required. In line with the formulations in Chapters 2 and 3, the limit state function for a bridge or viaduct subject to traffic loads can be expressed as:

$$Z = g(\mathbf{X}) = \theta_R R - \theta_E (G_\mathrm{DL} + G_\mathrm{SDL} + C_{0Q} Q) \tag{4.6}$$

where θ_R is the uncertainty in the resistance model, R the structural resistance, θ_E the model uncertainty of the load-effect calculation, G_DL the dead load effect, G_SDL the superimposed dead load effect, C_{0Q} the time-invariant part of the live load effect, and Q the time-variant (traffic) load effect. The resistance model uncertainty (θ_R) requires additional thought when employing a low-informative prior for the resistance, as R represents aleatory variability and epistemic uncertainty. In this setting, θ_R may be redundant, as its epistemic contribution is typically negligible compared to R. However, retaining θ_R can be justified in the light of the assumed probabilistic schematisation, especially when R becomes more informative.

The low-informative prior for resistance can be based on basic engineering knowledge. Even in the absence of detailed information, it is usually possible to establish an approximate resistance. Materials used, geometry, and structural layout provide suf-

ficient details to estimate permanent load effects, while inspection data and material tests may refine the resistance estimation. For bridges and viaducts, it is reasonable to expect that the resistance substantially exceeds annual traffic load effects (in the absence of known issues). This basic knowledge allows for formulating a low-informative distribution for R with a physically meaningful mean value. If used directly, such a prior leads to an unacceptably low reliability index, as the probability mass of low resistance values dominates. In combination with the information from a survived proof load test, however, the low-value probability mass is eliminated, and the updated reliability can reach acceptable levels.

In statistical inference problems, the likelihood is typically expressed through a probability density function; however, survival in a proof load test is represented differently. Here, an indicator function (or potential) is applied, which discriminates between states agreeing and disagreeing with the test outcome (Abril-Pla et al., 2023; Lauritzen et al., 1990). The following limit state function for proof load testing is used to evaluate likelihood $p(Z_{\mathrm{PL}} \geq 0 \mid \boldsymbol{\theta})$:

$$Z_{\mathrm{PL}} = g_{\mathrm{PL}}(\mathbf{X}) = \theta_R R - [\theta_E (G_{\mathrm{DL}} + G_{\mathrm{SDL}}) + \theta_{E,\mathrm{PL}} Q_{\mathrm{PL}}] \tag{4.7}$$

where Q_{PL} is the applied proof load effect, and $\theta_{E,\mathrm{PL}}$ the associated model uncertainty. As a result of the updating process, the distributions of all variables in the limit state function of Equation (4.7) are updated—although the distributions of the test-related variables will not be used in the subsequent evaluation of the posterior reliability via Equation (4.6). The update may also consider survival under past traffic loading, as the assessment concerns an existing structure. Extending the data to incorporate one or several years of observed traffic load typically improves the posterior reliability. Ideally, time dependence related to resistance degradation and traffic load trend is accounted for explicitly through additional factors such as c_R and c_Q (Chapter 2). However, the most significant increase in conditional reliability follows from the first year of demonstrated in-service strength (Section 4.5).

4.3.3 Calculation methods

Conceptually, the Bayesian reliability updating procedure consists of two parts: (1) perform Bayesian inference (Section 4.2) and, given the posterior distribution, (2) calculate the updated reliability (Section 4.3.1). In structural reliability, posterior distributions quickly become difficult to obtain analytically due to the large number of random variables involved. Numerical methods are therefore required to evaluate the joint posterior distribution. Two common approaches are:

- **Markov-chain Monte Carlo (MCMC)**: In this popular sampling procedure, a Markov chain is constructed to generate samples from the posterior distribution. The Metropolis-Hastings (Hastings, 1970; Metropolis, 1953) and Gibbs sampling algorithms (Geman & Geman, 1984) may be used to produce such a sequence, or chain. Because the chain is built gradually, a burn-in period is required to discard initial samples that do not yet represent the joint posterior distribution.

- **Bayesian Monte Carlo (BMC)**: In this sampling method, the prior distributions are sampled directly (Dilks et al., 1992; Hornberger & Spear, 1980). The likelihood of the data is evaluated for each parameter value and used as a weight (Kloek & Van Dijk, 1978). If an indicator function is applied (e.g. in proof load testing), samples inconsistent with observed performance are discarded. The remaining samples describe the joint posterior distribution.

The main benefit of MCMC over BMC is its flexibility to incorporate uninformative priors, which are typically improper and cannot be sampled directly. This is possible because the sampling algorithm uses proposal distributions, with accepted samples forming the posterior. These distributions, however, must be set up and adjusted (automatically) based on acceptance rates. BMC, in contrast, is simpler but requires proper priors and sufficient posterior density to ensure effective sampling. Although concerns have been raised regarding the efficiency of BMC (Qian et al., 2003), many samples are quickly discarded using an indicator function as the likelihood. In addition, all procedures described in this dissertation are implemented in C++, a language designed for high-performance computing (Stroustrup, 2013), thereby enabling the use of large sample sizes.

Structural reliability is often evaluated using efficient iterative procedures, such as the first- and second-order reliability methods (Breitung, 1984; Hasofer & Lind, 1974). In Bayesian updating, these methods can be applied in hierarchical models where only parameters are updated. If the basic random variables themselves are altered, their application is less straightforward. In specific cases, the updating expressions can be reformulated to allow approximations with standard reliability methods (see Section 5.3.5). More often, numerical methods such as Monte Carlo simulation are again required to calculate the updated reliability. It thus makes sense to combine the inference and failure probability calculation in a single procedure: given posterior samples, additional predictive samples are generated for evaluating the limit state function. In this way, reliability updating is achieved in a single calculation, without the need for explicitly storing the posterior distribution.

4.4 Prior distribution sensitivity study

To examine the sensitivity to the assumed prior distribution, additional analyses are performed on the structure described in Section 2.3. The simply-supported slab bridge spans 10 m, carries a single lane of traffic, and its bending moment at midspan is considered. From a Bayesian perspective, the previous analysis was highly informative with respect to the probabilistic description of the resistance. Instead of using a mechanical resistance model, a low-informative prior distribution based on the basic information provided by the statistical description of the traffic load is used. Resistance deterioration and load effect trend are not considered as sensitivity to the assumed

prior distribution is the main focus, and therefore, a concise formulation based on a residual resistance ($\hat{R}$) can be adopted:

$$\begin{aligned} Z &= (\theta_E G_{\mathrm{DL}} + \theta_E G_{\mathrm{SDL}} + \hat{R}) - \theta_E (G_{\mathrm{DL}} + G_{\mathrm{SDL}} + C_{0Q} Q) \\ &= \hat{R} - \theta_E C_{0Q} Q \end{aligned} \tag{4.8}$$

where the definition of the random variables is described in Section 4.3.2. Following the same residual resistance formulation, the limit state function for the proof load testing situation becomes:

$$Z_{\mathrm{PL}} = \hat{R} - \theta_{E,\mathrm{PL}} Q_{\mathrm{PL}} \tag{4.9}$$

For simplicity, it is assumed that the same model uncertainty holds for the proof load testing and regular traffic load situation ($\theta_{E,\mathrm{PL}} = \theta_E$), effectively eliminating these variables from the reliability problem. For the remaining random variables considering the load effect (C_{0Q}, Q, and Q_{PL}), the parameter values in Table 2.2 apply.

Three different types of prior distributions are employed to compare posterior reliability: normal, uniform and triangular (see Table 4.1 and Figure 4.2). The mean value of the annual maximum traffic load effect derived from WIM data is used as the primary parameter to formulate the prior distributions. For the normal prior distribution, the mean value is equal to the mean traffic load effect (1150 kNm, see Table 4.1). In addition, a second normal prior is considered with the mean value equal to 1.5 times the annual traffic load effect ($1.5 \times 1150 = 1725$ kNm). Referring to Equation (2.5) and Table 2.2, the residual resistance is about $4100 - 721 - 101 = 3278$ kNm, which is roughly a factor of 3 higher than the mean traffic load effect and highlights the cautious choice of prior distributions. For the upper bound of the uniform distribution, the safety margin associated with the non-deteriorated structure is adopted as the upper bound parameter, corresponding to $3 \times 1150 = 3450$ kNm. A triangular distribution, with the same bounds as the uniform distribution, is also included to (conservatively) express a stronger belief in lower resistance values.

Reliability analyses for each prior distribution type (normal, uniform, triangular) were performed using the Bayesian Monte Carlo method (Section 4.3.3). The posterior distributions of $\hat{R}$ given the survival of a target load of 1800 kNm are shown in Figure 4.3. The annual reliability indices after a successful proof load test are provided in Table 4.1. With values ranging from 3.95 to 4.29, these prior distributions may indeed be regarded as low-informative, reflecting that mainly the lowest resistance values influence the res-

Table 4.1 Prior distribution parametrisation for $\hat{R}$ and results of reliability analyses, including those of the lower-bound approach.

Prior distribution	Parameter values	Annual reliability index
Normal	$\mu = 1150$ kNm, $V = 0.5$	3.95
Normal (factor 1.5)	$\mu = 1725$ kNm, $V = 0.5$	4.17
Uniform	$a = 0$, $b = 3450$ kNm	4.29
Triangular	$a = 0$, $b = 3450$ kNm, $c = 0$	4.14
Lower bound	-	3.43

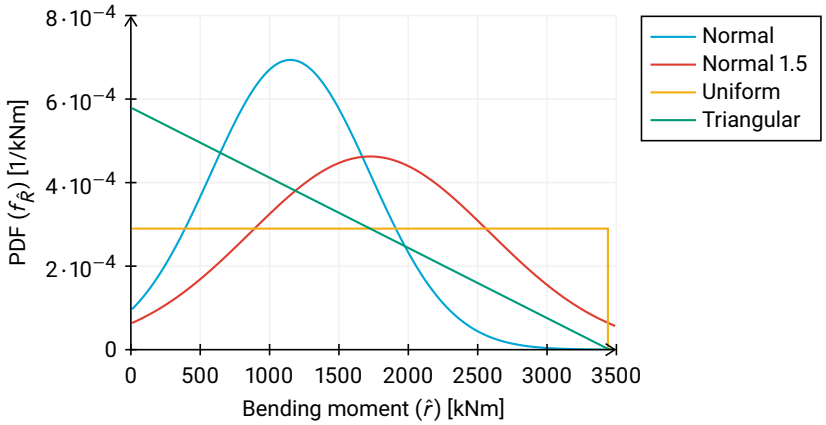


Figure 4.2 Low-informative prior distributions considered for the resistance, based on a mean annual maximum traffic load effect of 1150 kNm.

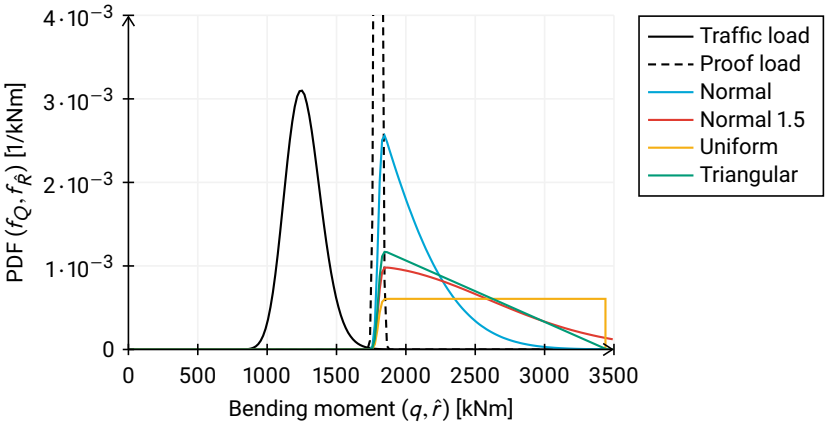


Figure 4.3 Annual maximum traffic load ($C_{0Q}Q$), proof load (Q_{PL}) and residual resistance ($\hat{R}$) posterior distributions following from using different prior distributions.

ults. Increasing the COV of the load (Q) and varying the right-tail shape between light (Weibull) and heavy (Fréchet) produced the same relative differences, indicating no apparent sensitivity to such alterations. The lower-bound reliability calculation was also included for comparison, and results in a smaller index due to its conservative nature. From a Bayesian perspective, a result derived from a low-informative prior represents the most credible estimate of the reliability index, as the lower-bound approximation does not accurately reflect physical reality. The presented results thus illustrate the impact of such a methodological choice.

4.5 Time-dependent analysis with low-informative prior

The Bayesian approach with a low-informative prior for the resistance can be integrated in a time-dependent reliability analysis to incorporate (and visualise) effects such as resistance deterioration and a trend in traffic loads. Primarily, the reliability analysis is performed as outlined in Chapter 2, but with a broad-banded prior distribution reflecting the epistemic uncertainty. To compare the evolution of the reliability using such a low-informative prior with an informed approach, a modification of the case study described in Section 2.3 is performed. A broad-banded normal prior distribution is adopted, with the mean value equal to 1.5 times the permanent and traffic loads, i.e. $1.5 \times (721 + 101 + 1150) = 2958$ kNm, and a variation coefficient of 0.5. As demonstrated in the previous section, this prior description provides a middle ground between the considered options. In contrast to the previous section, this calculation includes permanent loads because resistance deterioration is considered (c_R applies to R, not $\hat{R} = R - G$).

The result of the time-dependent Bayesian analysis with two proof load tests is provided in Figure 4.4. The informative time-dependent analysis result (Section 2.3) is included for comparison. The conditional annual reliability is markedly lower for the low-informative case when compared to the informative case, particularly before the first proof load test is performed. The in-service proven strength from surviving past traffic loads alone does not achieve the target annual reliability ($\beta = 4$). This indicates that either more information needs to be collected to form a more informative resistance distribution, or the target loads need to be increased. To reach the desired reliability in 2020–2040, the target loads need to be increased from 1800 and 2000 kNm to 2000 and 2100 kNm. When increased, the reliability level during the first test decreases even further ($\beta \approx 0.3$, $P_f \approx 0.38$), clearly highlighting the lack of knowledge.

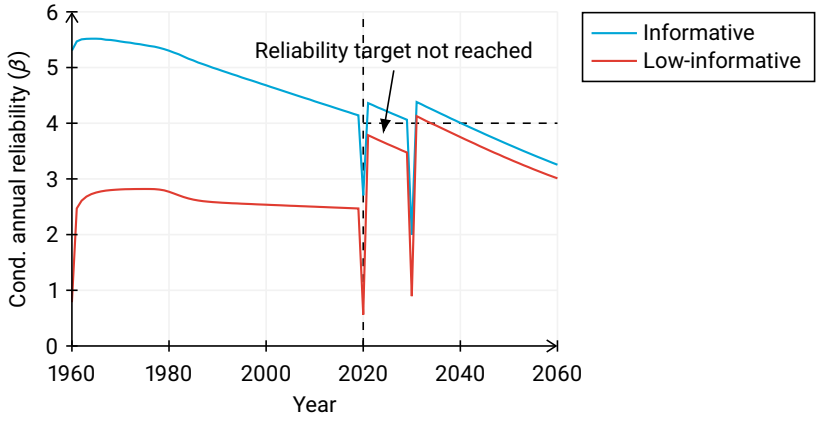


Figure 4.4 Conditional annual reliability over time, assuming an informative and low-informative prior distribution for the resistance. For both cases the target loads are 1800 kNm in 2020 and 2000 kNm in 2030.

4.6 Concluding remarks

This chapter has shown that the adoption of Bayesian methods provides a versatile methodology for proof load testing under varying levels of knowledge. By combining prior information with test outcomes and in-service proven strength, structural reliability estimates become increasingly accurate. The Bayesian approach accommodates both informative and low-informative settings, offering an expansion of conventional reliability assessment methods where epistemic and aleatory uncertainty are equally important. A sensitivity study demonstrated that reliability estimates are largely consistent given different low-informative prior distributions. However, a time-dependent analysis highlights the need for higher target loads when proof load testing under limited knowledge, compared to earlier informed analyses. Additionally, the reliability values during proof load testing were shown to be very low, underscoring the importance of gathering detailed information before testing is undertaken.

The primary value of this chapter lies not in the general use of low-informative priors, but rather in the formulation of Bayesian reliability updating for proof load testing. From a Bayesian viewpoint, mechanical resistance models are treated as sources of information rather than required inputs. This concept supports a broad range of applications for the proposed methodology. The following chapters build on this foundation by employing Bayesian methods to incorporate monitoring data (Chapter 5) and to address various failure modes and spatial correlations (Chapter 6)—utilising priors with varying levels of informativeness. Each application is accompanied by dedicated calculation procedures and flowcharts, illustrating how the approach can evolve into a powerful tool for structural assessment.

5

Incorporating monitoring data into the assessment

To prevent excessive test loads and their potential damage, it is desirable to incorporate as much information as possible in an assessment via proof load testing. In this chapter, in-situ monitoring and laboratory testing data are combined in a Bayesian update of the structural reliability after each successful load increment. Two case studies are presented where laboratory testing and analytical modelling provide ample evidence to justify test load reductions of 20% and 25%. The proposed method offers a systematic framework to combine information from distinct sources and address the uncertainties in resistance, loads and measurements. Nonetheless, the representativeness of the data in terms of structural similarity and the influence of measurement noise remain important factors. Despite these challenges, incorporating monitoring data during proof load testing can reduce test loads and inform decisions on whether to continue or abort the test to prevent structural damage.

5.1 In-situ measurements and laboratory data

This thesis aims to develop a comprehensive structural reliability updating framework in which as much information as possible can be incorporated, also during the proof load test. To keep track of structural performance during the test, numerous measurement devices are typically installed, such as lasers, LVDTs, strain gauges and cameras. During the test, the proof load is applied in controlled increments to minimise potential damage. Within each of these load steps, the structural performance may be evaluated based on indicators such as displacements and crack widths. While not immediately indicative of overall structural health, such indicators can be interpreted in light of structural behaviour observed in previous laboratory experiments and analyt-

ical resistance models. For example, sectional analysis can help to establish a strain stop criterion that prevents unwanted damage during testing (Zarate Garnica & Lantsoght, 2021). In addition to evaluating stop criteria, in-situ measurements can provide valuable insight into the structure's reliability during the test. This chapter presents a novel reliability updating method that integrates this information into the assessment via proof load testing. More detailed descriptions of the proposed method and the case studies are provided in De Vries et al. (2024) and De Vries et al. (2025b) (Appendix D).

5.2 Statistical inference

To probabilistically interpret the information from indicators based on structural responses, the laboratory testing results should be post-processed in such a way that they can be used within a reliability analysis. As the number of specimens tested is usually limited due to cost, time, and material availability constraints, the data cannot be expected to represent the inherent variability precisely. The process of distilling a meaningful description from limited data, acknowledging both natural and statistical uncertainty, is called *statistical inference* (Wasserman, 2004).

Suppose n observations of a quantity of interest X are available, for example concrete strength, reinforcement yield stress or resistance ratio. These data can be used to formulate probabilistic statements about a future observation X_{n+1}. Its (posterior) predictive distribution should also reflect the statistical uncertainty arising from the limited sample size. Assuming normality and unknown mean (no prior information), two cases are typically distinguished: one where the standard deviation is assumed to be known, and one in which it is unknown (no prior information) and must be estimated from the data, respectively (Box & Tiao, 1973; Geisser, 1993):

$$X_{n+1} = \bar{x} + \sigma\sqrt{1+1/n}U \sim \mathcal{N}\left(\bar{x}, \sigma^2(1+1/n)\right) \tag{5.1a}$$

$$X_{n+1} = \bar{x} + s\sqrt{1+1/n}T \sim t_{n-1}\left(\bar{x}, s^2(1+1/n)\right) \tag{5.1b}$$

where $\bar{x}$ is the sample mean, U is a standard normally distributed variable, and σ is the (known) population standard deviation. If unknown, the sample standard deviation s may be calculated from the data. This necessitates the use of T, a random variable following Student's t-distribution with $\nu = n-1$ degrees of freedom (Gosset, 1908). Besides the variance inflation by factor $1+1/n$, more uncertainty is introduced with the t-distribution because it is typically wider than the normal distribution. In some cases, n is artificially increased when (prior) information suggests a similar standard deviation as calculated from the dataset. Information about the distribution type can also be incorporated, for example, by assuming a lognormal distribution for material properties (CEN, 2019; JCSS, 2015).

Rather than relying on Equations (5.1a) and (5.1b), explicitly performing Bayesian inference enables the incorporation of statistical uncertainty, prior knowledge, and censored observations. In a Bayesian inference context, usage of Equations (5.1a) and (5.1b) is equivalent to adopting a non-informative Jeffreys prior for the parameters of the normal distribution (Gelman et al., 2013; Jeffreys, 1961). In Section 2.2.2 of De

Vries et al. (2025b) (Appendix D), a direct, fiducial derivation was presented to motivate the variance inflation and use of the t-distribution (Fisher, 1956; Geisser, 1993; Hannig, 2009). However, the fiducial approach has been criticised (Savage, 1961; Zabell, 1992). For example, conditioning on the observed data implies that the pivotal quantities (U and T) are no longer standard normal and t-distributed, requiring the introduction of independent copies. The intended pivoting operation amounts to adopting a right-Haar prior within a Bayesian framework, which coincides with Jeffreys' prior for location-scale distribution families (Eaton & Sudderth, 2004). Such non-informative priors may, however, be overly generic and fail to reflect actual prior knowledge, leading to conservative posteriors. Bayesian inference is therefore preferable, as it provides a coherent and flexible framework. Bayesian inference and the way in which structural reliability may be updated on the basis of new information were described in the previous chapter (Chapter 4).

5.3 Reliability updating method

5.3.1 General principle

After each load increment in the proof load test, the estimation of the resistance on the basis of measurements forms the first source of information (Figure 5.1, point 1). The proposed method relates the observed in-situ response, through the use of indicator values, to the response observed in laboratory tests on similar structural elements or derived from suitable analytical models. For example, strains derived from measurements at the bottom of a beam or slab can serve as suitable indicators for structural performance (Lantsoght et al., 2019). By utilising the same indicators in both laboratory and in-situ testing, it is possible to determine the resistance distribution via statistical inference (Section 5.2).

If a structure can withstand a specific proof load, it also means that its resistance is equal to or greater than the load effect during the test. This (gradual) truncation of

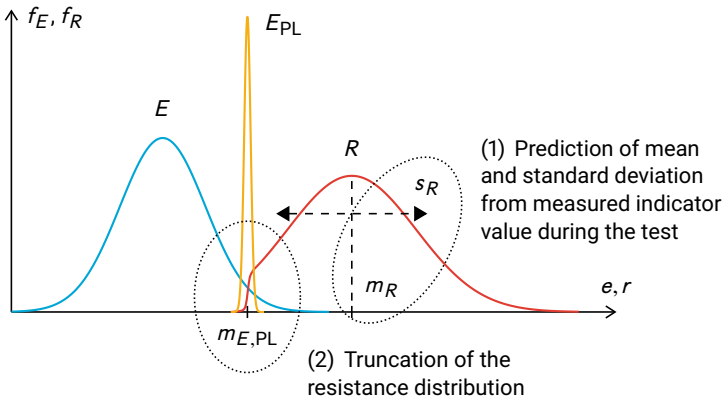


Figure 5.1 General principle of updating the resistance distribution from predicted mean and standard deviation, followed by truncation.

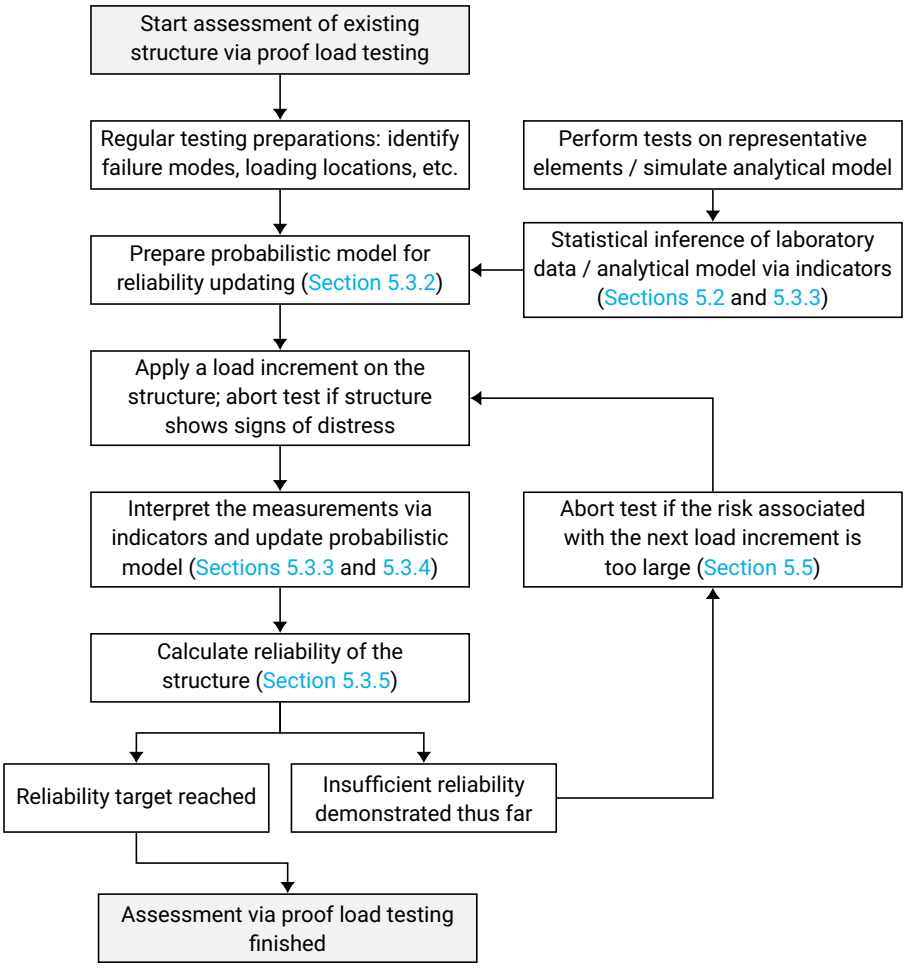


Figure 5.2 Flowchart outlining the steps in the proposed proof load testing assessment method.

the resistance is the second source of information (Figure 5.1, point 2). The information from the two sources is processed in the presented order, allowing the structure's reliability to be assessed after each load increment. Generally, as the applied load increases, the structural reliability also tends to increase because of the proven ability to withstand large loads. A flowchart describing the proposed proof load testing assessment method is provided in Figure 5.2. It should be noted that the presented method does not consider time-dependent effects, such as resistance deterioration or traffic load trends (Chapter 2). Nonetheless, the Bayesian formulation is well-suited to incorporate these effects as an advancement.

5.3.2 Probabilistic model

The first limit state function considers the situation in which the regular traffic loads act on the bridge. The function is the same as adopted previously (Chapters 2 and 4) and aligns with JCSS (2015) and *fib* (2016):

$$Z = \theta_R R - \theta_E (G_{\mathrm{DL}} + G_{\mathrm{SDL}} + C_{0Q} Q) \tag{5.2}$$

where θ_R is the uncertainty associated with resistance calculation, R is the resistance, θ_E is the model uncertainty of the load effect calculation, G_{DL} is the dead load effect, G_{SDL} is the superimposed dead load effect, C_{0Q} is the time-invariant part of the live load effect, and Q is the time-variant part of the live load effect (i.e. traffic load effect). The reliability of the bridge prior to proof load testing could be evaluated if distributions were assigned to all random variables, including R. Because of the low information state, the distribution of R would exhibit significant variability and thus result in low reliability. Rather than employing a conventional structural resistance model, the resistance will be estimated by combining in-situ measurements with insights gained from laboratory experiments or analytical modelling. This also affects how the resistance model uncertainty should be quantified (Section 5.3.4).

Each time the structure withstands a new load level in a proof load test, the distribution function of R can be updated to reflect the information obtained via measurements. The revised distribution can be expressed as the product of a resistance ratio (X) and the proof load effect (E_{PL}) produced during the last successful load test cycle ($R = XE_{\mathrm{PL}}$). Given a certain observed indicator value, the distribution of X may be obtained—as determined before proof load testing (Section 5.2). The limit state function describing the proof load testing situation, in which the regular traffic is absent, is expressed as:

$$Z_{\mathrm{PL}} = \theta_R X E_{\mathrm{PL}} - E_{\mathrm{PL}} = (\theta_R X - 1) E_{\mathrm{PL}} \propto \theta_R X - 1 \tag{5.3}$$

where $E_{\mathrm{PL}} = \theta_E (G_{\mathrm{DL}} + G_{\mathrm{SDL}}) + \theta_{E,\mathrm{PL}} Q_{\mathrm{PL}}$, with Q_{PL} denoting the load effect created by the proof load and $\theta_{E,\mathrm{PL}}$ its corresponding model uncertainty. By assuming that the proof load is withstood, it directly follows that $Z_{\mathrm{PL}} > 0$ and thus $\theta_R X > 1$. The conditionality can be satisfied in a Bayesian updating process by obtaining the joint posterior distribution of $\boldsymbol{\theta} = (\theta_R, X)$ as:

$$p(\boldsymbol{\theta} \mid Z_{\mathrm{PL}} > 0) \propto p(Z_{\mathrm{PL}} > 0 \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{5.4}$$

where the likelihood $p(Z_{\mathrm{PL}} > 0 \mid \boldsymbol{\theta})$ acts as an indicator function, or potential (Abril-Pla et al., 2023; Lauritzen et al., 1990), and $p(\boldsymbol{\theta})$ is the prior probability (Chapter 4). After this update, the marginal distributions should not be sampled independently because the interdependence of variable combinations significantly influences the outcomes. Returning to the original traffic load situation, Equation (5.2), with $R = XE_{\mathrm{PL}}$ and denoting the updated variables as θ'_R and X' gives:

$$Z = \theta'_R X' [\theta_E (G_{\mathrm{DL}} + G_{\mathrm{SDL}}) + \theta_{E,\mathrm{PL}} Q_{\mathrm{PL}}] - \theta_E (G_{\mathrm{DL}} + G_{\mathrm{SDL}} + C_{0Q} Q) \tag{5.5}$$

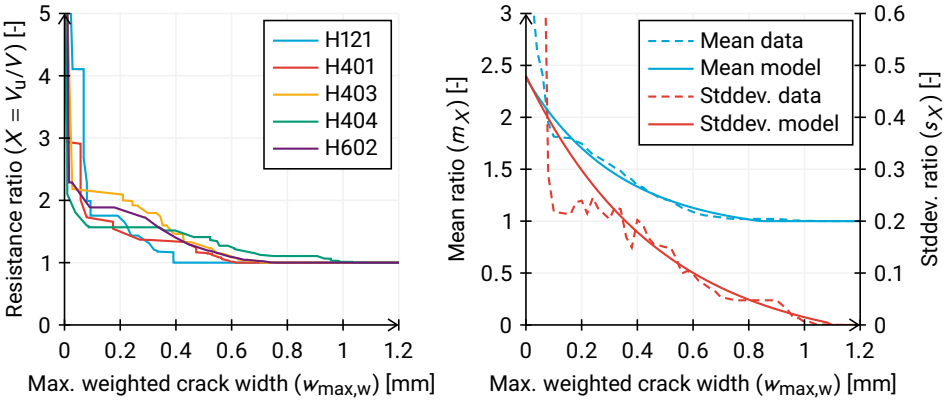


Figure 5.3 Shear resistance ratio versus maximum weighted nominal crack width data (left) and statistical post-processing with an exponential model (right).

and can be used to evaluate the structural reliability after a successful test cycle. The chosen probabilistic formulation accounts for model uncertainties in regular load effects (θ_E) and proof load effects ($\theta_{E,\mathrm{PL}}$), and their correlation.

5.3.3 Distribution of the resistance ratio

When laboratory data are available, the relationship between measurements or indicator values and the resistance ratio ($X = R/E$) can be inferred from these tests. The laboratory tests should be conducted on similar elements and in a configuration comparable to the in-situ proof load test. Each specimen has a resistance (R) that corresponds to the load effect at the moment that the limit state is reached (failure). During each load step, the indicator value is observed and the current load effect (E) is calculated, ultimately producing a resistance ratio curve for each specimen. Once all resistance curves have been obtained, statistical inference is used to obtain their probabilistic description (Section 5.2). To illustrate the procedure, Figure 5.3 displays the resistance ratio curves and the statistical post-processing belonging to the case study described later in Section 5.4.1.

In cases where laboratory measurements are unavailable, analytical models can be utilised as an alternative. Rather than solely calculating the (design) resistance, the aim shifts to determining the distribution of the resistance ratio as the indicator value increases. A regular resistance model can be developed, but the parameters need to be random variables. Typically, a simulation procedure is used to integrate the probability space and derive the statistical distribution of the resistance ratio. Accounting for statistical (small-sample) uncertainty is unnecessary in this procedure, but additional resistance model uncertainty is introduced.

5.3.4 Resistance model and measurement uncertainty

In the described probabilistic framework, θ_R accounts for uncertainty in the resistance calculation introduced by the proposed procedure. This uncertainty is small when the laboratory specimens closely resemble the actual structure or when the accuracy of the analytical model has been validated. Conversely, significant uncertainty is expected if the structures differ substantially, the mechanical model is overly simplistic, measurement errors are significant, or there are inconsistencies in data post-processing. The application of assessment methods in real-world scenarios can highlight and quantify the difference between laboratory and in-situ observations (Šýkora & Holický, 2013).

Measurement errors can significantly influence results, especially with small values, such as minor crack widths or small strains. As the magnitude of the values increases, the relative impact of measurement errors typically decreases. In addition, the error also depends on the parameter being measured and whether it can be directly measured or must be inferred. For example, larger measurement errors are expected when estimating crack widths compared to measuring strains directly. The moderate load values in the proposed method will typically result in small indicator values, and thus, the large uncertainty should be appropriately accounted for. In Bayesian inference, measurement noise can be addressed within the inference procedure (Bishop, 2006; MacKay, 1991). But other analytical approaches can also enhance model accuracy by incorporating physically expected trends, thereby providing a nuanced understanding of the data (see, for instance, Section 5.4.1).

5.3.5 Calculation procedure

In order to compute the structural reliability after a successful proof load test, the knowledge of surviving the applied load needs to be incorporated, see Equation (5.4). To obtain the joint posterior distribution $p(\theta \mid Z_{\mathrm{PL}} > 0)$, several calculation methods can be employed, such as the Bayesian Monte Carlo Method (BMC) (Dilks et al., 1992; Hornberger & Spear, 1980) or Markov chain Monte Carlo (MCMC) (Geman & Geman, 1984; Hastings, 1970; Metropolis, 1953). In this chapter, the MCMC method is adopted because of its versatility. Owing to its Monte Carlo nature, the chain's current state is directly used to evaluate the reliability of the structure with the posterior distribution.

A computationally attractive alternative is using the SORM (Breitung, 1984; Hohenbichler et al., 1987), which can approximately account for the non-linearity present in the limit state function. Because the survival condition, $Z_{\mathrm{PL}} > 0$, cannot be incorporated directly, an approximation must be made. This may be achieved by introducing a substitute random variable $\hat{X}$ with the combined variance of θ_R and X, i.e. $\mathrm{Var}(\hat{X}) = \mathrm{Var}(\theta_R) + \mathrm{Var}(X)$, which can subsequently be left-truncated to the set $[1, \infty)$ to impose the survival of the proof load test.

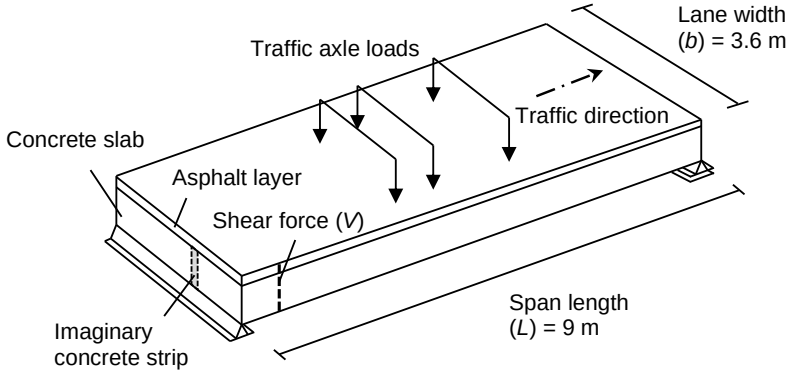


Figure 5.4 Hypothetical reinforced concrete slab used as a case study. Reprinted with permission from De Vries et al. (2025b).

5

5.4 Applications of the method

5.4.1 Shear resistance supported by laboratory tests

In order to illustrate the practical application of the proposed method, the reliability of a hypothetical shear-critical reinforced concrete slab bridge is considered. The case exemplifies older Dutch slab bridges that lack shear reinforcement. For simplicity, the slab is designed to match the exact width of a single traffic lane (Figure 5.4). Deep beams representing sections of such a slab were tested in the laboratory to evaluate their shear resistance (Yang et al., 2021). The strips, or deep beams, tested in the laboratory had a length of 9 m, a width of 0.3 m and a height of 1.2 m. They were subjected to a load via a single jack positioned near the midpoint of the span, leading to shear failure near the supports (Figure 1.2, left). Given the specified lane width of 3.6 m, the slab comprises 3.6 / 0.3 = 12 strips. The reliability of the bridge will be assessed using the resistance data and the measurements obtained from a proof load test.

Digital image correlation (DIC) (Jones & Iadicola, 2018) was used to monitor and analyse the development of cracks in the concrete during the tests. DIC allows for the determination of the nominal crack width at various locations across the length of the beam (Zarate Garnica, De Vries & Lantsoght, 2022). To assign greater importance to the cracks near the supports, a weighted crack width calculation is introduced, multiplying the nominal crack width by the factor Vd/M, where V and M refer to the shear force and bending moment at the specific location, and d is the effective depth. Statistical inference of the measurement data resulted in the sample mean and standard deviation as a function of the indicator ($w_{\max,w}$). For nominal crack widths less than 0.08 mm, noticeably high mean and standard deviation values were obtained. These high values are likely due to noise affecting the DIC measurements at very small displacements. An

Table 5.1 Calculated reliability indices given different levels of loading, during and after proof load testing (PLT).

WIM characteristic load factor ($m_{Q,\mathrm{PL}}/Q_{\mathrm{k,WIM}}$)	LM1 characteristic load factor ($m_{Q,\mathrm{PL}}/Q_{\mathrm{k,LM1}}$)	Reliability during PLT	Annual reliability after successful PLT	
			Proposed method	Lower-bound
1.0	0.78	1.80	2.99	0.53
1.2	0.93	1.79	3.77	1.89
1.4	1.09	1.78	4.55	3.04
1.6	1.24	1.71	5.23	4.03
1.8	1.40	1.61	5.83	4.85

exponential model was used to address the erratic behaviour in the sample mean and standard deviation of the resistance ratio (Figure 5.3):

$$m_X(w_{\mathrm{max,w}}) = \begin{cases} 1.5\exp(-3\,w_{\mathrm{max,w}}) + 0.88 & w_{\mathrm{max,w}} < 0.85 \text{ mm} \\ 1 & w_{\mathrm{max,w}} \geq 0.85 \text{ mm} \end{cases} \tag{5.6a}$$

$$s_X(w_{\mathrm{max,w}}) = \begin{cases} 0.53\exp(-2.1\,w_{\mathrm{max,w}}) - 0.05 & w_{\mathrm{max,w}} < 1.1 \text{ mm} \\ 0 & w_{\mathrm{max,w}} \geq 1.1 \text{ mm} \end{cases} \tag{5.6b}$$

Meaningful target loads for proof load testing may be defined as the characteristic traffic load multiplied by a certain factor. For a given target load, the laboratory test data were used to estimate plausible indicator values; in real-world applications, this intermediate step would not be necessary. The probabilistic model specific to the considered structure, including an overview of the random variables and their parameter values, may be found in De Vries et al. (2025b).

The calculated reliability indices, both during and after a successful test, are provided for each load level in Table 5.1. The results clearly show that the proposed method provides higher annual reliability values after a successful test than the lower-bound approach (Chapter 3) utilising the same probabilistic description. Its added value is particularly significant when the test loads are fairly low. Around the annual reliability index of 4, the minimum for CC3 structures in the Netherlands, the required target load is about 20% lower when using the proposed method ($m_{Q,\mathrm{PL}}/Q_{\mathrm{k,WIM}}$ value of 1.26 versus 1.59). The reliability values during testing should not be compared to regular requirements (Section 2.1), as public safety measures are typically in place. The reliability index of $\beta = 1.8$ during the test suggests that the structure is, in fact, likely to withstand the applied load ($P_{\mathrm{f}} = \Phi(-1.8) \approx 0.04$). The additional information provided by monitoring data offers a clear advantage over low-informative resistance assumptions, which generally indicate lower reliability during testing (Section 4.5).

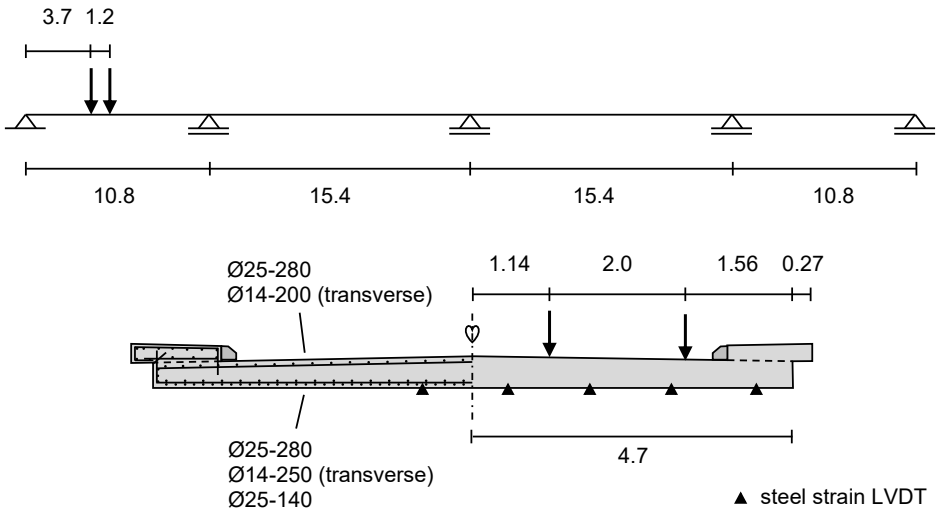


Figure 5.5 Schematic side view (top) and cross-section of the De Beek viaduct with the bending proof load test location indicated by the arrows (bottom). Reinforcement diameters and spacings provided in mm, other measurements provided in m. Reprinted with permission from De Vries et al. (2025b).

5.4.2 Bending resistance supported by a calculation model

In the second case study, proof load levels relating to the bending resistance of the reinforced concrete slab viaduct De Beek will be examined. The viaduct from 1963 passes over the A67 highway near Ommel in the Netherlands, but the viaduct itself is part of the secondary road network. The cross-section height varies parabolically along the longitudinal direction from 0.47 m to 0.87 m, measured in the heart of the deck. The bridge is constructed as a continuous slab and thus experiences support moments above its middle three supports (Figure 5.5). A pilot proof load test of the first span has already been performed for both bending and shear mechanisms, but only bending will be considered in this case study. Thus, in contrast to the previous case study, in-situ test results and measurements are available (Koekkoek, Lantsoght, Yang & Hordijk, 2016). However, laboratory tests on similar bridge decks have not been performed and will be substituted by an analytical model.

Since no laboratory tests have been performed on representative specimens, an analytical model is used to interpret in situ measurements. Using earlier measurements, a modified cross-sectional analysis model was calibrated for the current viaduct. While such analysis is often presented as a moment-curvature (M-κ) diagram, the bottom steel strain (ε_s) is utilised in this case study because it was measured during the test. Latin hypercube sampling (Helton & Davis, 2003; MacKay, 1991) was used to simulate responses according to the probabilistic resistance model, as it better represents the output distribution with small sample sizes. Unlike the previous case study, the statistics of the resistance ratio X can be directly obtained from the curves (Figure 5.6).

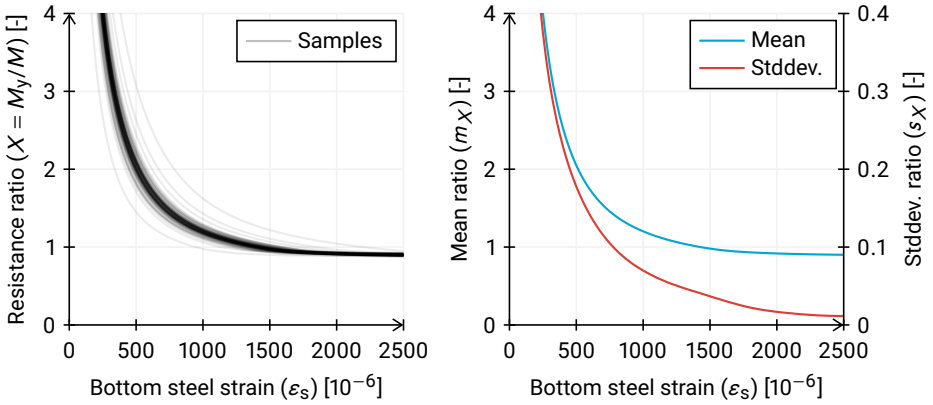


Figure 5.6 Bending resistance ratio versus steel strain following from LHS (left), and the calculated mean and standard deviation (right).

Table 5.2 Calculated reliability indices given different levels of loading, during and after proof load testing (PLT).

WIM characteristic load factor ($m_{Q,\mathrm{PL}}/Q_{\mathrm{k,WIM}}$)	LM1 characteristic load factor ($m_{Q,\mathrm{PL}}/Q_{\mathrm{k,LM1}}$)	Reliability during PLT	Annual reliability after successful PLT	
			Proposed method	Lower-bound
0.67	0.46	4.78	3.58	−1.70
1.12	0.76	3.18	3.99	1.95
1.57	1.06	1.79	5.00	4.12
1.91	1.29	0.91	5.78	5.27
1.96	1.33	0.80	5.92	5.40

The probabilistic model described in Section 5.3 is now employed to perform the reliability analysis. The indicator is the steel strain in the bottom reinforcement near the edge of the deck (ε_s). Because it is based on a stochastic model rather than data, the resistance ratio (X) is described by a normal distribution rather than the Student's t-distribution. For each of the load levels used in the proof load test, the mean and coefficient of variation of the resistance ratio are calculated (Table 5.2). Also, in this application, using indicator data yields markedly higher reliability indices at low load levels compared to the lower-bound method. Around the minimum annual reliability index of 4, the required target load is about 25% lower ($m_{Q,\mathrm{PL}}/Q_{\mathrm{k,WIM}}$ value of 1.12 versus 1.55).

5.5 Considerations for future application

The proposed proof load testing method involves updating the resistance distribution using two sources of information: (1) the observed in-situ response, which is related to the resistance via indicator values and associated resistance ratios, and (2) the survival of the applied load during the proof load test. Although the prediction of the resistance

on the basis of the in-situ response is associated with considerable uncertainty, this information is still valuable and can be accounted for probabilistically. In this Bayesian methodology, the failure probability during the test is calculated using the most current data—offering a risk-optimal alternative to predetermined stop criteria.

The application of the proposed method was demonstrated through two case studies. Each of the applications presented specific challenges with respect to the treatment of the resistance model and measurement uncertainty. The first case study highlighted the influence of representativeness, small sample sizes and measurement noise in laboratory data, while the second showed how probabilistic mechanical models can provide an effective alternative when such data are unavailable. The coefficient of variation used for the model uncertainty of the resistance in both case studies ($V_{\theta R} = 0.15$) is higher than commonly adopted values (Sýkora et al., 2015). The larger value reflects the considerable uncertainty associated with the representativeness of laboratory tests in the first case study, and the use of a basic mechanical model in the second. Future research and applications can further refine the framework, providing valuable insights into suitable model uncertainties and measurement techniques.

With the proposed method, proof load testing can become more cost-effective and less time-consuming, minimising traffic disruption during bridge testing. The two case studies demonstrated that proof load reductions of 20% and 25% are feasible compared to the lower-bound approach (Chapter 3). The method enhances proof load testing by providing accurate insights into structural reliability during and after testing, thereby supporting optimal asset management decisions. Large-scale implementation would benefit from research into generally applicable indicators, resistance ratio curves, identification of application criteria (representativeness), stochastic FEM modelling, and extension of the method to account for time-dependent effects such as resistance deterioration and increasing traffic loads.

6

Addressing spatial correlation and system reliability

For structures with many components, critical sections, and multiple possible failure mechanisms, it is challenging to demonstrate sufficient reliability with only a limited number of tests. This chapter proposes a system-reliability updating framework that addresses these challenges. A hierarchical Bayesian model is developed in which outcomes from a small number of tests can be used to infer the reliability of the entire—provided higher test loads are applied. The method is implemented through a Bayesian Monte Carlo procedure that accounts for spatial correlation in resistance and load effects. Two case studies illustrate the approach: one derives transfer factors quantifying the required increase in target loads, and the other identifies optimal testing strategies and vehicle configurations. The proposed methodology enables comprehensive proof load testing procedures in which the number of tests and target loads is minimised.

6.1 System reliability and practical testing limitations

Although proof load testing can be a versatile method for structural assessments, conceptual and practical challenges arise when a structure is viewed as a system of related components. These include physical elements such as the bridge deck and supports, but also cross-sections and connections, which may be subjected to different loading situations. A single component may be involved in multiple potential failure mechanisms, depending on the loading scenario, each of which must be evaluated. In practice, the number of tests is often restricted by practical constraints. This raises the question: *to what extent can a limited number of load tests demonstrate the reliability of the entire structure?* The proof load testing situation differs significantly from code-based

structural design, where elements are typically verified at the component level. Yet, the reliability of a system can vary considerably depending on its configuration and (spatial) correlation between components (Ditlevsen, 1979). In addition, questions persist about whether structural design and assessments should adopt a system-reliability perspective and what reliability levels would be appropriate (Diamantidis et al., 2025).

Addressing system performance and spatial correlations is required to integrate proof load testing into comprehensive reliability assessment procedures. This chapter begins by outlining the necessary background on representativeness of tests, natural variability in shear and bending resistance, and the use of random fields to represent spatial variations. Next, a Bayesian system reliability updating method addressing spatial correlation and component similarity is presented. The method employs a hierarchical model with a combination of informative and low-informative priors to represent the expected failure modes. A more detailed description of the proposed method and its applications may be found in De Vries et al. (2026) (Appendix E).

6.2 Background

6.2.1 Representativeness of tests and transfer factors

A key concern in proof load testing is whether the tested components, cross-sections or spans can adequately represent the performance of similar components within the same or comparable structures. Selecting representative components and determining the number to test is crucial for the conclusions drawn about untested components. For example, one span may be chosen to represent an entire bridge, potentially yielding conservative results if it has unfavourable geometry, detailing or condition (Zhou & Guzda, 2020). Where components exhibit similar demand-to-capacity ratios, a more systematic approach can provide a statistical basis for addressing this uncertainty.

While proof load testing typically focuses on the performance of an individual component, (spatial) variability between components or cross-sections introduces uncertainty regarding their performance. The translation of this additional statistical uncertainty into characteristic or design values is addressed in Annex D of EN 1990 (CEN, 2019). Specifically for proof load testing, a transfer factor (γ_t) is introduced in the most recent German guideline for load tests on concrete structures (DAfStb, 2020; Marx et al., 2019). The factor increases the target load to compensate for the added uncertainty of not testing all components in the assumed series system configuration:

$$Q^*_{PL} = \gamma_t Q_{PL} \tag{6.1}$$

where γ_t is the transfer factor, and Q_{PL} is the target load that applies to one component. The value of the transfer factor depends on the number of components tested (n), the total number of components (N), and the variation coefficient of the resistance (V_R).

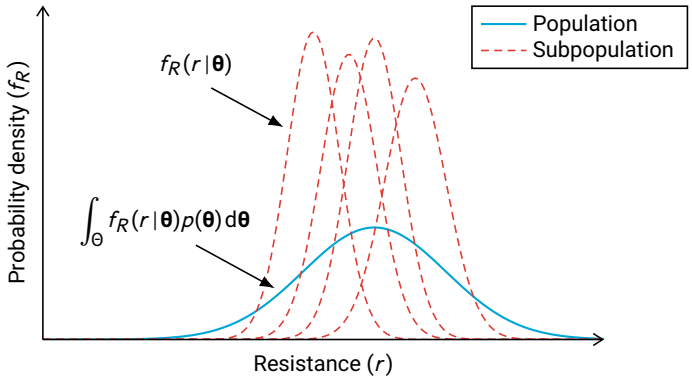


Figure 6.1 Hierarchical representation of resistance uncertainty: subpopulations capture the natural variability, while the population reflects both epistemic and aleatory. The Bayesian notations of Chapter 4 apply; vector $\boldsymbol{\theta}$ contains the model parameters.

6.2.2 Natural variability of the resistance

Structural members exhibit variability in their resistance due to natural, mostly random, differences in composition, manufacturing processes, degradation, and interaction with other structural components. It is the variability that would persist if a manufacturer were asked to produce the same structural component multiple times. This type of uncertainty is also referred to as aleatory, or inherent, variability. It is considered separate from model, or epistemic, uncertainty, which is attributed to a lack of knowledge. In many applications, incorrect separation of aleatory and epistemic uncertainty does not significantly influence the results—provided both are incorporated. However, the separation becomes important when some form of scaling is performed, e.g. spatially (system reliability) or in a time-variant context (Der Kiureghian, 2022).

Within a structure, resistances are typically dependent when viewed within the overall population, for instance, due to a shared manufacturer, source materials or environmental factors. Such dependencies reduce uncertainty when extrapolating test results to untested components (Grigoriu & Hall, 1984; JCSS, 2015). A simple way to model them is through basic correlation values and equicorrelation matrices. Alternatively, hierarchical Bayesian modelling allows these dependencies to be represented more naturally (Nishijima & Faber, 2007). This approach has also been used to establish a *track record* of structural performance from in-service proven strength of buildings sharing specific floor detailing (De Vries et al., 2025a). Hierarchical Bayesian modelling (Section 6.3.2) enables separation of uncertainties: in this application, high-level parameters capture epistemic uncertainty, while lower-level variables describe natural variability. Figure 6.1 illustrates this: the solid line is the population (prior predictive) distribution, combining epistemic and aleatory uncertainty, while dashed lines show subpopulation (conditional) distributions, capturing only natural variability. A subpopulation resistance may be assigned a random field to model spatial variability (Section 6.2.3) or used to produce realisations of resistances within the structure.

Table 6.1 Shear resistance variation coefficients (V_R) following from repeat tests of RC beams *without* shear reinforcement.

Reference	Test specimens	Number of tests	COV (V_R)
Bentz and Buckley (2005)	SBB2.1 – SBB2.3	3	0.03
	SBB3.1 – SBB3.3	3	0.03
Bernander (1957)	A1, A2, A4, A5	4	0.06
Chana (1981)	2.1a - 2.3b	6	0.04
	3.3a, 3.3b, D3	3	0.11
	5.1a-5.2b	4	0.09
Collins and Kuchma (1999)	B100, B100B, B100-R	3	0.10
Hallgren (1994)	B90 SB5, SB6, SB9, SB10	4	0.05
	B91 SD1 – SD6	6	0.06
Leonhardt and Walther (1962)	D3/1, D3/2l, D3/2r	3	0.06
	D4/1, D4/2l, D4/2r	3	0.04
Lubell et al. (2004)	AT-3/N1 – AT-3/T2	4	0.04
Regan (1971)	R1, R3, R7, R29	4	0.10
Sarkhosh et al. (2015)	S1B1 – S1B6	6	0.05
	S3B1 – S5B5	10	0.04
Sherwood (2008)	S-10N1+2, S-20N1+2, S50N1+2	6	0.03

Table 6.2 Shear resistance variation coefficients (V_R) following from repeat tests of RC beams *with* shear reinforcement.

Reference	Test specimens	Number of tests	COV (V_R)
Ahmad et al. (1996)	NHW-3, NHW-3a, NHW-4	3	0.07[a]
Bernhardt and Fynboe (1986)	S7, S8	4	0.08[a]
Kong and Rangan (1997)	S1	6	0.12
Krefeld and Thurston (1966)	29a-2, 29b-2, 29f-2	3	0.06[b]
Özden (1967)	T6, T7, T9	3	0.06

[a] Slightly varying shear reinforcement.
[b] Specimens with very low concrete compressive strength omitted.

To determine suitable variation coefficients for subpopulations intended to represent the natural variability, relevant *repeat tests* were identified in the literature. Such tests are commonly included in extensive testing campaigns to verify monitoring equipment and testing procedures, and are expected to display differences primarily due to natural variability. The DAfStb-ACI dataset contains numerous test series, both with and without shear reinforcement (ACI-DAfStb, 2025; Reineck et al., 2013; 2014). For the first time, variation coefficients are derived from these data to quantify the natural variability. An overview of the suitable test series and calculated variation coeffi-

cients is provided in Tables 6.1 and 6.2. Due to the lack of series in which all specimens were tested under identical parameters, series with slightly varying shear reinforcement are also included. Most, if not all, specimens are purpose-made for laboratory testing, as these studies focused on mechanical behaviour rather than assessing existing structures. Taking this contextual difference into account and observing no systematic difference in the variation coefficients between members with and without shear reinforcement, a single representative value of $V_R = 0.10$ is adopted for shear resistance. To complete the statistical description, spatial variability must also be considered and will be addressed next.

6.2.3 Spatial correlation and random fields

To complete the required background for the treatment of spatial variability and system effects, the modelling of spatial correlation through stochastic processes, or random fields, is introduced. Stochastic processes and fields mathematically describe random variables with dependencies, allowing their integration into probabilistic models. Typically, one-dimensional time-dependent processes are referred to as stochastic processes, while descriptions involving spatial or multidimensional variability are referred to as random fields. These fields can represent the spatial correlation of loads, material properties, or directly their resulting resistance (Section 6.3.4). Gaussian processes model such phenomena by assuming that the observations are normally distributed, and fully characterised by a mean and a covariance function (Vanmarcke, 1983). Although some processes are better captured by non-normal distributions, Gaussian processes remain instrumental, as their realisations can always be related through transformation, for example, using the inverse-transform method (Papoulis & Pillai, 2002).

Autocorrelation functions, or kernels, describe the correlation between two points separated by a given lag or distance. The type of kernel and correlation length depend on the material under consideration. The exponential and Gaussian kernels represent two limiting cases in terms of smoothness. Figure 6.2 shows the normalised autocorrel-

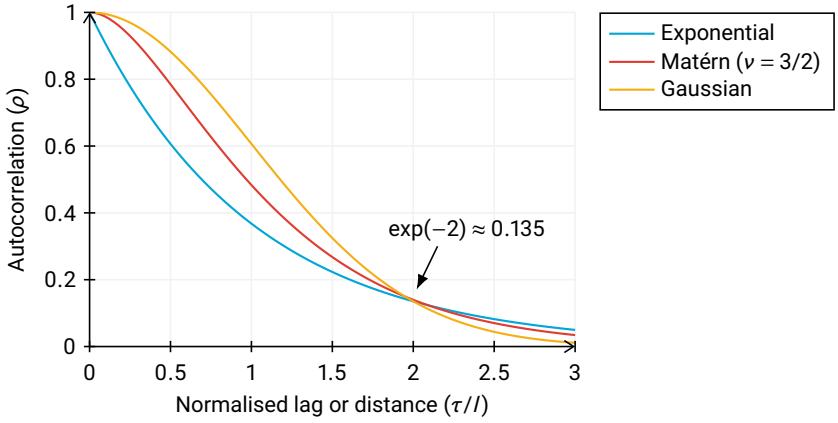


Figure 6.2 Comparison of the exponential, Matérn and Gaussian autocorrelation functions.

ation functions and highlights the intermediate smoothness of the Matérn kernel with $\nu = 3/2$. Although other kernels exist, the three described here are typical for structural reliability analyses, and their corresponding expressions are provided in Appendix E. It is worth noting that the Gaussian kernel appears in the literature with and without a factor 2 in the denominator describing the correlation length—thus reported lengths must be interpreted with care (Botte et al., 2023; Geyer et al., 2023; JCSS, 2015).

6.3 System-reliability evaluation method

6.3.1 Reliability updating given system performance

The Bayesian reliability updating procedures incorporating proof load testing, as described in Chapter 4, may also be performed within a system reliability context. A series system is assumed, meaning that system failure occurs if any component fails. In the current context, a component describes a specific structural element, cross-section, and failure mode. The analysis accounts for the survival of multiple components under a given proof load effect. Each component is characterised by its own limit state function, involving component-specific random variables; however, inter-component correlations are typically present. For notational convenience, the first n components in an otherwise arbitrary enumeration of all N system components are those tested. If the survival of n components is used as evidence to assess the failure probability of a system comprising N components, the system's failure probability is expressed as:

$$P_{\mathrm{f,sys}} = \mathrm{P}\Big(\bigcup_{i=1}^{N} Z_i < 0 \,\Big|\, \bigcap_{j=1}^{n} Z_{\mathrm{PL},j} \geq 0 \Big) \tag{6.2}$$

where the limit state function for traffic (Z_i) and the function representing the proof load testing situation ($Z_{\mathrm{PL},j}$) are as defined in Section 4.3.2. When applied in a system-reliability context, these functions and their random variables may be annotated with the corresponding subscript (i and j), to indicate component-specific quantities. The failure probability of a particular component i, given a successful proof load test involving n components, is expressed as:

$$P_{\mathrm{f},i} = \mathrm{P}\Big(Z_i < 0 \,\Big|\, \bigcap_{j=1}^{n} Z_{\mathrm{PL},j} \geq 0 \Big) \tag{6.3}$$

The component-level evaluation is relevant, as it aligns with codified assessment procedures where, in principle, structural elements are verified individually. A component's posterior reliability depends strongly on whether it belongs to the subset of tested components. Tested components typically exhibit high posterior reliability, as proof load tests directly demonstrate their resistance. Untested components are more critical, as they rely solely on demonstrated system performance. Consequently, the survival of past traffic loads provides useful additional evidence, as detailed in Section 6.3.3.

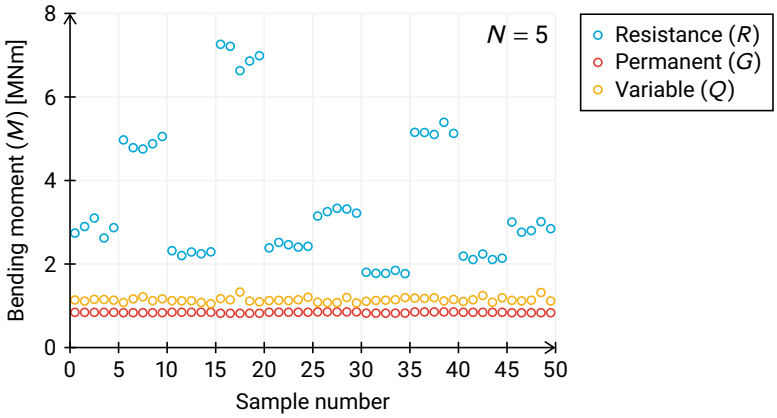


Figure 6.3 Samples obtained via a hierarchical model for bending-moment resistance, together with permanent and variable load effects drawn from regular distributions.

6.3.2 Hierarchical Bayesian model

As detailed knowledge of the structure is not always available and deterioration can significantly affect resistance, low-informative priors can be formulated using basic information about the structure and traffic loads (Chapter 4). A hierarchical Bayesian model enables separate treatment of the typically high uncertainty surrounding the mean resistance and the more refined knowledge regarding its coefficient of variation. If represented by a single broad-banded distribution, these nuances would be lost. It should, however, be recognised that if deterioration continues, a successful proof-load test does not guarantee future reliability. This issue can be addressed by incorporating time-dependence (Chapter 2), or by targeting a higher reliability.

In the proposed hierarchical Bayesian model, the resistance (R) of the considered component is assumed to follow a lognormal distribution, parametrised through a mean and coefficient of variation (De Vries et al., 2025a):

$$R \sim \text{Lognormal}(m_R, V_R) \tag{6.4}$$

The mean and coefficient of variation themselves are modelled as random variables as well, also following lognormal distributions:

$$m_R \sim \text{Lognormal}(m_{m_R}, V_{m_R}), \quad V_R \sim \text{Lognormal}(m_{V_R}, V_{V_R}) \tag{6.5}$$

The hyperparameters $m_{m_R}, V_{m_R}, m_{V_R}$, and V_{V_R} are assigned appropriate values based on expert judgement and case-specific information. The parametrisation via mean and coefficient of variation does introduce a dependency between the parameters of the underlying normal distribution. However, this dependency is not considered problematic, as it reflects domain knowledge, and the Bayesian updating process introduces such dependencies as well. An example of realisations produced by the hierarchical sampling scheme is provided in Figure 6.3, where 10 structures with $N = 5$ components

were generated, resulting in 50 samples. The example illustrates how distinct groups of resistances form, with limited variability present within each group.

6.3.3 Calculation procedure

Various methods can be used to calculate the posterior distribution, incorporating observed system performance and the resulting reliability of the system or its components (Section 6.3.1). In this chapter, the Bayesian Monte Carlo method is employed, where the accepted (posterior) samples may be directly used to evaluate the reliability of the structure (Section 4.3.3). This simplifies the calculation, such that effectively just one integral needs to be evaluated. For example, the failure probability of component i, given successful proof load testing of n components, may be evaluated via:

$$P_{\mathrm{f},i} = \int_{\Theta} \mathbb{1}[g_i(\theta) < 0]\, p\Big(\theta \;\Big|\; \bigcap_{j=1}^{n} g_{\mathrm{PL},j}(\theta) \geq 0\Big)\, \mathrm{d}\theta \tag{6.6}$$

where $\mathbb{1}[\cdot]$ is the indicator function, $p(\theta \mid \cdot)$ is the posterior distribution, and $\theta = (\theta_R, m_R, V_R, R_1, R_2, \ldots)$ carries the entire state of the system, including unconditioned predictive variables such as the future traffic loads. If the system failure probability is desired, the indicator function would consider the minimum of all limit state functions $g_i(\cdot)$. To acquire a more realistic posterior distribution, the survival of one year of traffic load may also be included. This inclusion can be viewed as sequential Bayesian updating (Gelman et al., 2013), in which the posterior obtained after accounting for the survival of past traffic loads serves as the prior for the survival update of the proof load test. A flowchart outlining the Bayesian Monte Carlo calculation procedure, assuming the same statistical description for all components, is provided in Figure 6.4.

6.3.4 Simulation of correlated variables and random fields

When evaluating system reliability, it is necessary to consider the correlation between quantities used in the probabilistic description of a structure. As described in Section 6.2.3, these quantities may be modelled as random fields but must be discretised for use in evaluating local limit state functions. This applies to resistance, permanent loads, and variable (traffic) loads. Typically, the (statistical) model uncertainties $(\theta_R, \theta_E, C_{0Q})$ can be assumed to be fully correlated across components, as they stem from a common schematisation or modelling approach (Melchers & Beck, 2018).

The correlation between resistances depends on the failure mechanism considered. Regarding shear resistance, the material properties of concrete govern and may be assumed to be uncorrelated, except when cross-sections are very close. For bending resistance, the degree of correlation depends primarily on whether the same reinforcement bars are present, as their strength is the dominant influence (Section 6.5). The

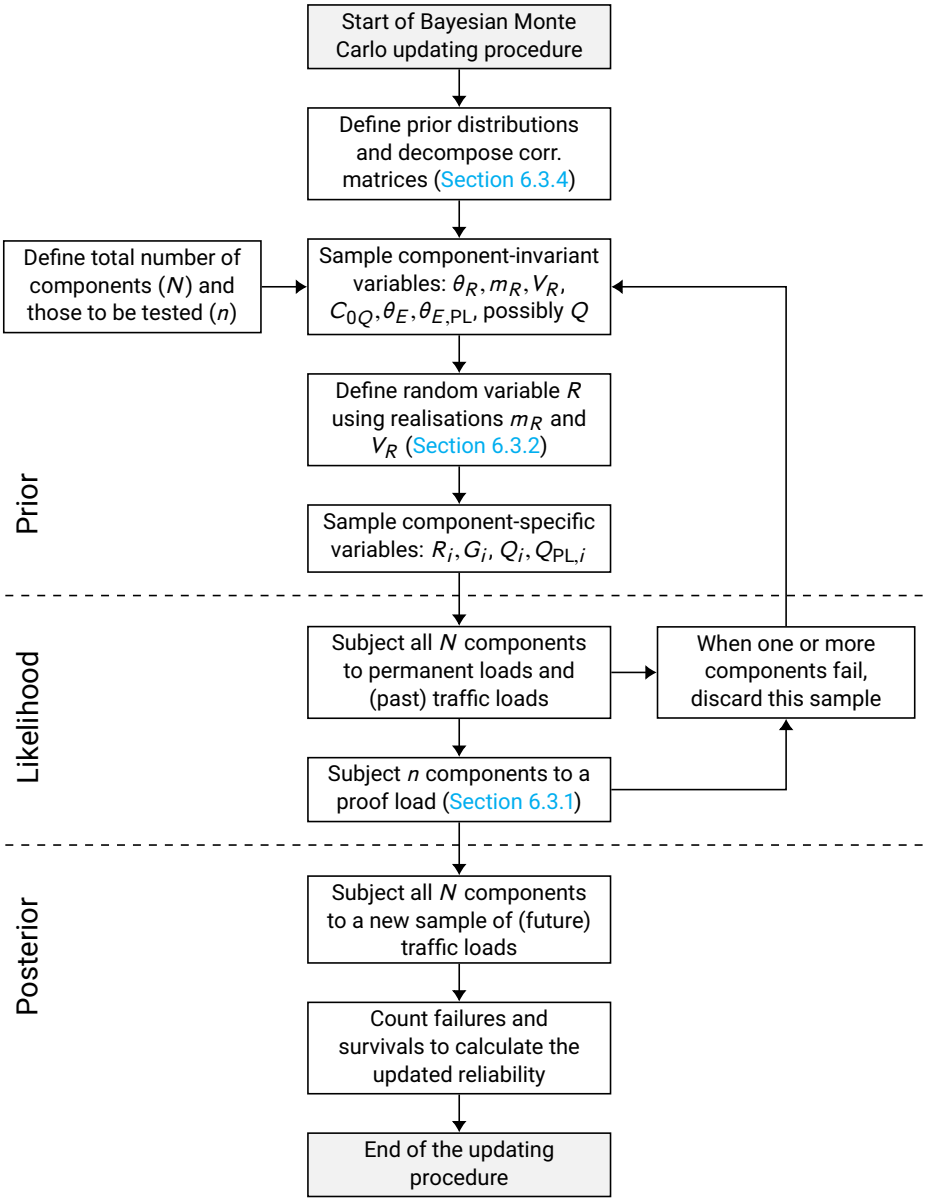


Figure 6.4 Flowchart outlining the steps in the proposed Bayesian Monte Carlo procedure to update the system reliability.

discretised correlations, derived from the evaluation of autocorrelation functions or otherwise, can be represented using correlation matrices, for example:

$$\mathbf{R}^{(2)} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}, \quad \mathbf{R}^{(3)} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \tag{6.7}$$

where ρ_{ij} denote the Pearson correlation coefficients between variables. Correlation matrices are symmetric, reflecting the symmetry of autocorrelation functions. To sample correlated variables, a correlation matrix may be decomposed such that multiplication by a vector of independent standard normally distributed variables, $\mathbf{u}$, yields the desired correlated vector $\mathbf{u}_c$. If the correlation matrix is positive definite, it can be decomposed using the Cholesky decomposition (Golub & Van Loan, 2013):

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T, \quad \mathbf{u}_c = \mathbf{L}\mathbf{u} \tag{6.8}$$

where $\mathbf{L}$ is the lower-triangular matrix resulting from the decomposition.

6.3.5 Load correlation decay with increasing reference period

In analysing the load effects derived from weigh-in-motion (WIM) data (FHWA, 2018), it was found that the calculated Pearson correlation coefficients between month maxima were systematically lower than those between week maxima. A similar reduction was observed when comparing weekly with daily maxima, and this trend persisted across different locations and load effects (bending moment and shear). These observations are consistent with extreme value theory, which states that dependencies between block maxima generally weaken with increasing block size (Russell & Huang, 2021; Sibuya, 1959).

For structural reliability calculations, reference periods longer than just weeks or months are typically considered. Therefore, it is necessary to predict how correlation decays as the reference period increases. To this end, Monte Carlo simulations were conducted to compute the remaining correlation. These simulations employed the generalised extreme value distribution, with varied parameters—although these variations did not change the result. A graph demonstrating the decay of correlation with increasing reference period was produced by considering multiple initial correlation coefficients and various block sizes (Figure 6.5).

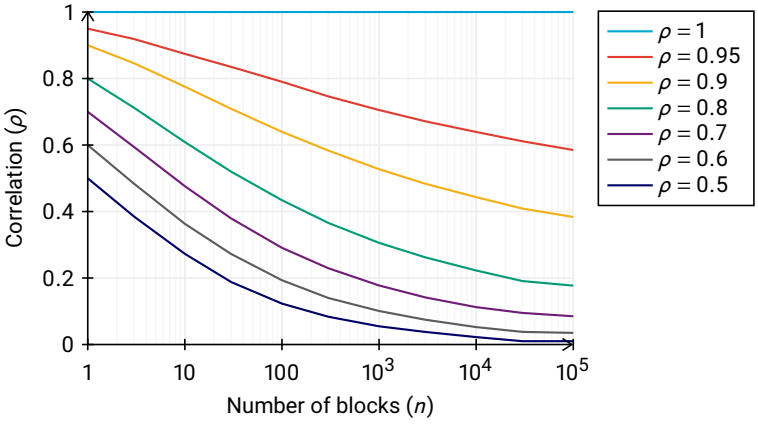


Figure 6.5 Decay of the Pearson correlation coefficient between normalised block maxima as a function of the number of blocks, for various initial correlation values.

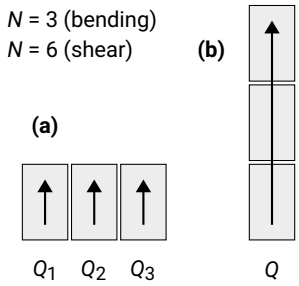


Figure 6.6 Schematic top view of slab bridges composed of three spans, with arrows indicating the traffic on spans situated (a) side-by-side (uncorrelated) and (b) one after the other (correlated).

6.4 Slab bridges with a varying number of spans

6.4.1 Description and objectives

To illustrate the application of the proposed method, a case study on simply-supported slab bridges is presented. Such bridges and viaducts, often with short spans, are common in the Netherlands and many other countries (Christensen et al., 2022; Lantsoght et al., 2017a). In this case study, bridges composed of multiple separate slabs are considered. These configurations are less common than continuous-span bridges (Section 6.5), but provide a clear basis. The reinforced concrete slabs are 10 m long and 3.5 m wide, representing a conservative single-lane scenario without load sharing between lanes. Given the simply-supported configuration, the maximum bending moment typically occurs at mid-span, while the maximum shear force is expected near the supports. When the critical cross-sections are treated as components, two shear-critical sections exist per span, which is relevant for defining the number of components (N) and sample size (n).

Two variable-load configurations are considered in the following analysis: one in which the load effect is independent across components, and the other in which the load effects are dependent (i.e. fully correlated). Applied to the simply-supported slab bridges, these configurations correspond to a bridge with its spans situated (a) side-by-side and (b) one after the other (Figure 6.6). Conservatively, failure of a single component constitutes system failure in both situations. If the traffic load effect originates from multiple lanes, the correlation will generally decrease as vehicles travel at different speeds and change lanes. The situations considered in the following analyses, i.e. independent and fully correlated, represent the two extremes. The full definition of the probabilistic model and description of correlations is provided in De Vries et al. (2026) (Appendix E).

6.4.2 Results

It is found that the variable-load dependency, i.e. spans situated side-by-side or one after the other, has a negligible influence on the reliability index for both failure mech-

Table 6.3 Transfer factors calculated for the simply-supported case study, considering bending failure.

Number of comp. (N)	Sample size (n)						
	1	2	3	4	5	6	8
1	1	-	-	-	-	-	-
2	1.13	1	-	-	-	-	-
3	1.17	1.07	1	-	-	-	-
5	1.21	1.11	1.07	1.03	1	-	-
8	1.23	1.14	1.11	1.08	1.05	1.03	1
10	1.25	1.16	1.12	1.09	1.07	1.05	1.02
20	1.28	1.19	1.15	1.13	1.11	1.10	1.07

Table 6.4 Transfer factors calculated for the simply-supported case study, considering shear failure.

Number of comp. (N)	Sample size (n)						
	1	2	3	4	5	6	8
1	1	-	-	-	-	-	-
2	1.50	1	-	-	-	-	-
3	1.58	1.30	1	-	-	-	-
5	1.66	1.40	1.27	1.16	1	-	-
8	1.73	1.47	1.34	1.27	1.21	1.15	1
10	1.76	1.49	1.38	1.30	1.25	1.20	1.12
20	1.83	1.56	1.44	1.38	1.33	1.29	1.24

anisms (maximum difference of 0.01). Therefore, the resulting transfer factors are presented in Tables 6.3 and 6.4 for bending and shear, respectively, without further distinction with respect to traffic-load dependency. As expected, transfer factors for shear are generally higher, reflecting the larger variability associated with this failure mechanism. In the considered case study, each span includes two critical cross-sections, effectively doubling the number of components (N). As a result, the transfer factor increases rapidly when the number of test positions is limited, illustrating the reduced representativeness of the test results.

For reference, the results are compared with values from the German guideline (DAfStb, 2020), which provides transfer factors for generic situations, focusing solely on resistance. For bending, the factors calculated in this case study are slightly lower than those reported in Table 2 of the German guideline. For shear, the values are comparable if the presence of shear reinforcement is assumed. In this study, no significant difference in the variation coefficient was found between cases with or without shear reinforcement (Section 6.2.2). Although the numerical values are similar, the underlying methodology differs substantially. Notably, the current approach explicitly considers the survival of one year of traffic loading, which significantly reduces the resulting transfer factors. Furthermore, this case study focuses on short-span slabs, their specific traffic loading, and the associated permanent-to-variable load ratio. The presented results are thus bridge and configuration-dependent.

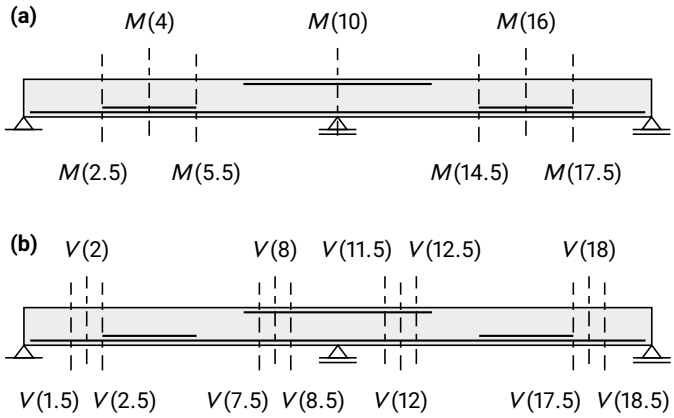


Figure 6.7 Schematic side view of the continuous slab bridge, displaying the main reinforcement and the (a) bending and (b) shear cross-sections of interest.

6.5 Multiple cross-sections and failure modes

6.5.1 Description and objectives

This second case study presents a more complex application, assessing a continuous slab bridge while accounting for spatial variability across multiple cross-sections and failure modes. Unlike the previous case study, no generalisation in terms of the number of components (N) and tested samples (n) is applicable. The objective is to quantify system-level reliability for this specific bridge configuration in bending and shear through proof load testing, considering uncertainties in material properties, reinforcement layout, and traffic loading. By treating its cross-sections as a series system, optimal testing strategies are developed, including the configuration of a dedicated load-testing vehicle. A hypothetical concrete slab bridge with two equal spans of 10 meters is considered, comprising a single traffic lane. The critical cross-sections are defined at multiple positions, considering bending moments and shear forces (Figure 6.7). The mid-span cross-sections at $x = 4$ m and 16 m are known to contain additional reinforcement (as could be determined from scans), while the rebar properties (diameter, grade, and condition) remain uncertain. The full description of the probabilistic model and definition of spatial correlations is provided in De Vries et al. (2026) (Appendix E).

6.5.2 Optimised locations and target loads

In a classical testing approach, each critical section would be considered individually through specific load placement and corresponding target load. The Eurocode LM1 tandem, comprising two axle loads of 300 kN spaced 1.2 m apart, is commonly used in Europe to mimic heavy vehicles (Lantsoght, 2019a; 2019b). Other axle configurations and spacings can also be used to achieve the target load effects. Fewer tests can be

Table 6.5 Reliability indices for the continuous slab bridge following optimised location testing and vehicle runs. Reliability index β_0 after one year of traffic loads; β_i after individual tests, β_{all} after surviving all tests, β_{tip} for the tipper truck, and β_{tra} for the semi-low trailer.

Cross-section	β_0	β_1	β_2	β_3	β_{all}	β_{tip}	β_{tra}
$M(2.5)$	2.7	**4.8**	2.7	4.3	5.0	4.2	4.6
$M(4)$	2.6	**4.3**	2.6	3.9	4.2	4.0	4.5
$M(5.5)$	2.8	**4.8**	2.9	4.4	5.0	4.3	4.7
$M(10)$	2.6	2.6	**4.2**	2.6	4.1	4.4	4.1
$M(14.5)$	2.8	4.4	2.9	**4.8**	5.0	4.2	4.8
$M(16)$	2.6	3.9	2.6	**4.3**	4.2	4.0	4.5
$M(17.5)$	2.7	4.3	2.7	**4.8**	5.0	4.1	4.7
$V(1.5)$	2.5	**4.1**	2.6	3.3	4.0	4.2	4.0
$V(2)$	3.1	**4.9**	3.2	4.0	4.9	4.7	4.3
$V(2.5)$	3.9	**4.7**	4.0	4.9	5.0	> 5	4.8
$V(7.5)$	3.7	3.7	**> 5**	3.7	> 5	> 5	> 5
$V(8)$	3.2	3.2	**4.8**	3.2	4.7	4.8	4.8
$V(8.5)$	2.8	2.8	**4.1**	2.8	4.0	4.3	4.7
$V(11.5)$	2.8	2.8	**4.1**	2.8	4.0	4.6	4.8
$V(12)$	3.2	3.2	**4.8**	3.2	4.7	> 5	4.9
$V(12.5)$	3.7	3.7	**> 5**	3.7	> 5	> 5	> 5
$V(17.5)$	3.9	4.6	4.0	**4.7**	5.0	> 5	4.8
$V(18)$	3.1	4.0	3.2	**4.9**	4.9	4.5	4.3
$V(18.5)$	2.5	3.3	2.6	**4.1**	4.0	4.2	4.0
System	1.7	2.3	1.9	2.3	3.6	3.6	3.7

achieved by adopting higher target loads, although this increases the probability of failure or damage during testing.

For complex bridge configurations, the minimum number of trucks and load cases that simultaneously meet the target load requirements of all critical sections can be determined via optimisation (Zheng et al., 2023). Applying such a strategy for the current case study, an optimum at (2.5, 3.7) m was identified that can verify both the span moment and shear resistance near the left support. Compared to testing each cross-section, a modest increase of the tandem load factor from 1.45 to 1.55 is required (γ_t = 1.07) to yield sufficient reliability in both failure modes. Additionally, two tandems at (6.3, 7.5) m and (12.5, 13.7) m can be used to verify support moment and shear resistance on either side of the support. Here, increasing the load factor from 1.25 to 1.35 (γ_t = 1.08) eliminates the need for separate shear tests. The third test is conducted on the second span, with testing locations mirrored to the first span. Thus, with just three tests and slightly higher proof loads, sufficient reliability can be demonstrated (Table 6.5).

6.5.3 Moving vehicle results

Instead of complex test set-ups involving loading frames, hydraulic jacks and counterweights, it would be preferable to use a single load testing vehicle. As the vehicle gradually moves over the bridge, many loading positions are examined, and several passes can be made to capture loading eccentricities. A simulation can then determine the resulting load effect envelope for the relevant cross-sections. Two suitable vehicle configurations have been identified:

- **Tipper truck**: Commonly used on construction sites for transporting materials like sand or gravel. For the current case study, the preferred truck has two tandems, each with axle loads of 1.3 × 300 kN (40 t). Tandem spacing is 1.3 m, with 6 m between tandems.
- **Semi-low trailer**: To reduce axle loads, a semi-low trailer can also be used. The ideal set-up has eight axles, each loaded at 0.8 × 300 kN (25 t), spaced at 1.5 m.

The results for both configurations are included in Table 6.5, showing that the tipper truck achieves higher reliability values in verifying shear capacity, whereas the trailer is more effective for bending. Since all positions are tested, the system reliability is slightly higher than under a separate testing strategy. However, the semi-low trailer's mass exceeds 206 t, considerably higher than the tipper truck's 167 t, and this could lead to overloading of the supports—which may be just as critical regarding their resistance as the deck slab.

The probability of failure during testing can also be calculated using the derived probabilistic model, but with low-informative distributions it is generally high. Formulating more informative resistance priors can significantly lower this failure probability (Section 4.5).

6.6 Discussion and conclusions

This chapter proposed a method for assessing the reliability of bridges through proof load testing, explicitly incorporating spatial correlation and system behaviour. The suggested Bayesian procedure utilised a hierarchical model with a combination of informative and low-informative priors to represent the expected failure modes. From the analyses, it was found that for slab bridges with a varying number of spans, the increase in target loads required to compensate for untested components can be systematically represented by transfer factors. Despite methodological differences, the calculated factors are comparable to those in the German guideline for proof load testing (DAfStb, 2020). For a continuous slab bridge, an optimised testing scheme involving just three test locations and moderately increased loads achieved the desired reliability. Testing strategies using moving vehicles were also investigated, leading to tipper truck and semi-low trailer configurations suitable for proof load testing.

It should be noted that, if components are too dissimilar, there comes a point where no magnitude of the transfer factor can demonstrate sufficient reliability. In such cases, their variability is too substantial and appropriate subpopulations should be identified

to reduce epistemic uncertainty. At the other end of the spectrum are very similar components. For example, if cross-sections in concrete members are closely spaced, correlations should be introduced using random field descriptions (Section 6.2.3). However, the resistances and load effects will be very highly correlated. Then, the extra cross-sections will not necessarily lead to additional accuracy, as the random dimension does not increase significantly—only resulting in extra computational effort (Ghanem & Spanos, 1991; Vanmarcke, 1983).

The load levels in Section 6.5 were selected to ensure that each cross-section meets individual reliability requirements. As a result, the overall system reliability was lower than normative targets, since all components contribute to failure in the assumed series model. This procedure aligns with current practice, where components are verified individually. However, system-level reliability assessment can provide more accurate results, particularly within risk-based approaches addressing safety, economy, and the environment (Celati et al., 2025; Diamantidis et al., 2025).

The transfer factors described in this chapter are conceptually straightforward and easy to apply. In practice, however, structural systems cannot always be decomposed into independent components within a system reliability context. Furthermore, information about resistance degradation, increasing traffic demands, as well as laboratory and in-situ data (Chapters 2 and 5) should also be considered. An integral Bayesian procedure accommodates such extensions—as demonstrated by incorporating one year of in-service-proven strength—thereby enabling flexible and accurate reliability evaluations.

7

Discussion

7.1 Reflection on results

From Chapter 2, it follows that the annual reliability assessment framework is highly suitable for evaluating the safety of existing infrastructure. Annual reliability does not imply that assessments should occur each year, but rather that the *temporal evolution of reliability* becomes visible, allowing for strategic intervention. Such an approach requires explicit incorporation of traffic load trends and deterioration models for resistance. These aspects are integral to the time-dependent probabilistic analyses, which must therefore receive due attention in the development of future assessment procedures. Additionally, Bayesian updating procedures can be applied to infer the current deterioration state and its expected development over time.

While Bayesian procedures provide a systematic and mathematically sound framework (Chapter 4), inept usage can lead to incorrect results. Their inherent complexity makes it difficult to detect such issues, which are primarily related to the formulation of prior distributions and computational convergence. An important part of the Bayesian method is the specification of the prior probability, which expresses belief about an uncertain quantity before any evidence is considered. This requirement has led to criticism of Bayesian analysis, since results may depend on the choice of prior. Throughout this dissertation, input to the Bayesian methods has been discussed in detail for the proof load testing application; nevertheless, different researchers may express different views and beliefs. This should not discourage the use of Bayesian methods, as modelling preferences exist in solving any complex problem, and sensitivity to prior assumptions can be systematically examined.

In this research, advanced probabilistic methods, including Bayesian approaches, are applied extensively. This approach contrasts with current proof load testing practice, which typically relies on conservative mechanical models, design loads, and truncated resistance distributions. The implicit use of such advanced probabilistic methods can lead to a distorted view of the added value of proof load testing. In many cases, a probabilistic model that incorporates detailed descriptions of bridge-specific loads

and resistance characteristics already provides substantial improvements in structural assessments. Moreover, the extensive body of experimental data on resistance and its modelling can serve to refine structural models further without requiring full knowledge of the underlying mechanics. In this context, the proposed Bayesian procedures, which continuously update assessments with in-situ performance data (Chapter 5), occupy an intermediate position between diagnostic testing and proof load testing.

7.2 Practical aspects regarding implementation

This dissertation has focused on the fundamental research developments needed to advance the practice. As practising engineers are not expected to implement the Bayesian procedures directly, their application is demonstrated and clarified through case studies. Two applications of the methodology have been presented, resulting in specific procedures to (1) incorporate laboratory testing or probabilistic surrogate models in combination with monitoring during proof load testing (Chapter 5), and (2) account for spatial uncertainty and system reliability aspects, thereby providing guidance on optimal testing locations and vehicle configurations (Chapter 6). Through the applications and case studies, the practical value of the proposed procedures becomes clear, indicating that the gap between research and practice is smaller than it may initially appear to be.

Although somewhat understated in the present study, the lower-bound approach (Chapter 3) can already add significant value to current practice that relies on standardised factors. To be most effective, this method requires an accurate statistical description of the traffic load effect. Such descriptions, depending on span length, are readily provided for simply-supported, single-lane highway bridges. In other cases, a suitable description of the load effect can be derived using WIM data and traffic load simulations. In the Netherlands and other countries, such data are readily available and continue to grow in volume. Compared to the more advanced Bayesian procedures, bridge-specific implementation of the lower-bound approach offers excellent value, as the other procedures also require an accurate statistical description of load effects. However, more information is typically available before and during testing, and incorporating it requires a Bayesian approach.

Based on analyses incorporating spatial variability and system reliability aspects, two viable load-testing vehicle configurations were described in Chapter 6: a tipper truck and a semi-low trailer. It should be reiterated that their axle loads and gross vehicle weight are exceptionally high and cannot be permitted to drive to the test location in this condition. Additional effort will be required to prepare such vehicles locally for load testing, which may involve the use of cranes or other ballast-moving equipment. The stated axle loads follow from adopting a low-informative prior; if the data collection and monitoring procedures described in Chapter 5 are implemented, these axle loads could be reduced even further. In all cases, a cautious approach in which the loads are gradually incremented remains essential.

8

Conclusions and recommendations

8.1 Conclusions

This dissertation has investigated proof load testing as a tool for reliability-based assessment of bridges and viaducts, with emphasis on time dependence, varying knowledge levels, monitoring data and spatial correlation. In this section, the main findings are brought together by answering the research questions introduced in Section 1.3. The four sub-questions are first addressed individually, as they concern the separate challenges that were identified in answering the main research question: *How can proof load testing establish the structural reliability of reinforced concrete bridges and viaducts?* Having addressed these constituent aspects, a synthesis follows on how proof load testing can indeed establish structural reliability.

The conclusions of this dissertation follow the research phases presented in Figure 1.3, which illustrates the progression from exploratory analyses to methodological developments, and finally synthesis and outlook. In answering the research subquestions, the relations to the findings of the individual chapters are emphasised:

1. *What influence do the functional lifetime, time-dependent effects, the moment at which the load test is performed and proven strength have on the target load?*

   The temporal nature of structural reliability and the influence of proof load tests, both during and after the test, were examined in Chapter 2. Annual reliability analyses showed how reliability evolves over the service life of a structure. After a successful proof load test, reliability increases significantly; however, the reduction in reliability during the test highlights the downside of high target loads and the need for safety measures. The functional lifetime differs between structures, and target loads reflect this variation by being higher for longer lifetimes. Over time, in-service proven strength, demonstrated by surviving past loads and the proof load, improves reliability. However, deterioration and increasing traffic loads ultimately lead to a gradual decrease in reliability. For the first time in proof load testing, the evolution of reliability was systematically quantified and used

to strategically determine the timing of tests: higher loads can satisfy long-term reliability but increase the probability of failure during testing. These analyses highlighted the effects of proof load testing on annual reliability, the value of in-service proven strength and the decline in reliability due to deterioration and traffic growth.

2. *How does the knowledge level (available information) of the structure change the way in which the target load can be calculated?*

   Throughout this dissertation, various levels of informativeness were considered. The time-dependent analyses in Chapter 2 were performed under a highly informative assumption, where detailed information about the resistance is available. At the other extreme, Chapter 3 explored a lower-bound approach in which the resistance is entirely removed from the formulation. This conservative procedure underpins international guidelines, but its effectiveness depends on accurate statistical descriptions of the traffic load effects, and does not fully exploit the potential of in-situ measurements. An up-to-date, fully probabilistic formulation was derived, demonstrating how WIM data can be used to derive configuration-dependent target load factors utilising existing load models. Recognising that neither extreme accurately reflects the actual state of knowledge in many assessments, Bayesian methods are proposed to incorporate any available information, however uncertain (Chapter 4). From a Bayesian perspective, mechanical resistance models are treated as sources of information rather than required inputs, enabling the methodology to accomondate a wide range of knowledge levels. Consequently, the target load is no longer a fixed value derived from a prior assumed state of knowledge, but rather a variable that evolves as information is gained. The primary value of the proposed methodology lies in viewing proof load testing within a broader Bayesian reliability updating context.

3. *How can in-situ monitoring data be utilised to directly update the structural reliability and indicate when to stop?*

   As proposed in Chapter 5, in-situ monitoring data can be utilised in the reliability assessment by linking observed structural responses to outcomes from laboratory tests or stochastically simulated data from analytical models. In a novel procedure, the resistance distribution is updated based on the observed in-situ response after each load increment. In turn, this information is combined with the survival of the applied load to update the structural reliability. By incorporating both laboratory and in-situ monitoring data into the Bayesian updating process, the knowledge of the actual structural resistance is significantly improved. This updating scheme therefore provides a rational basis for deciding when the test may be stopped: either the target reliability has been reached, or the reliability predicted for the next loading step indicates an unacceptably high probability of failure. As a result, target loads were shown to decrease up to 25% in the studied cases compared to the lower-bound approach—while the failure probability under test loads is realistically quantified. These outcomes demonstrate the value

of combining diverse data sources and the large amount of information that can be gathered during a proof load test, even with moderate loads.

4. *How can spatial correlation and system reliability be addressed to overcome the practical limitations of testing a few cross-sections, spans, or even structures?*

   Spatial correlation and system reliability aspects can be addressed by treating the structure as a system of components, whose overall reliability is assessed. The components may be distinct structural elements, as well as different cross-sections, and concern multiple failure mechanisms. For structures with multiple critical sections or failure modes, proof load tests at only a few locations raise conceptual difficulties, as not all components are explicitly evaluated. The hierarchical Bayesian model in Chapter 6 separates epistemic and aleatory uncertainty, and can be set up to capture correlations—thereby allowing limited tests to inform the reliability of untested components. This novel approach enables a reduction of the number of tests by increasing target loads (via transfer factors) and the development of optimal testing strategies, including the configuration of suitable load testing vehicles. Applications involving multiple structures have not been explicitly addressed; however, the methodology can be extended to such cases, provided that assumptions are carefully reviewed and adjusted where necessary.

The above answers form the basis for addressing the main research question: *How can proof load testing establish the structural reliability of reinforced concrete bridges and viaducts?* Altogether, this research demonstrates that proof load testing should no longer be viewed as a standalone assessment tool, but as a key element within a Bayesian reliability updating framework. When combined with advanced probabilistic modelling, monitoring, explicit consideration of time-dependence, and system reliability considerations, proof load testing provides a *uniquely accurate assessment* method for establishing the reliability of existing infrastructure. Its distinctive value lies in observing actual structural behaviour under high loads and integrating these observations with mechanical models of varying sophistication.

In the broader context of structural assessments, this research bridges the gap between analytical predictions and observed in-situ performance. Although the application of Bayesian methods in the realm of proof load testing is not new, this work is the first to integrate in-situ monitoring for evidence-based performance prediction—effectively replacing stop criteria that lack a clear reliability basis. In parallel, the novel hierarchical Bayesian model conceptually unifies the practice of single-component proof load and multi-component sample testing. This advancement in probabilistic modelling directly provides practical benefits: it enables the optimisation of test locations, target loads and test vehicle layouts. Overall, the effort required to involve reliability and load-testing expertise in applying the proposed procedures will often be outweighed by the substantial economic benefits gained from extending service life and avoiding unnecessary replacement of bridges and viaducts.

8.2 Recommendations for future research

As with any scientifically studied topic, further research is always possible across many aspects. Yet the most relevant questions often emerge from the practical application of a technology. Future research will require such direction, as certain elements warrant more attention than others, depending on the context. Such critical judgment is particularly relevant for the Bayesian approach, which provides a highly flexible formulation. Two applications of this methodology have been presented herein, leading to tailored procedures for incorporating in-situ measurements and accounting for spatial variability. Their use in practice will highlight future research needs, leading to continued refinement and broader applicability. The following paragraphs offer more specific recommendations, though they reflect only the current state of research and are by no means exhaustive.

To enhance the application area of the suggested procedures, statistical characterisation of traffic load effects is needed for bridge types beyond simply-supported single-lane structures. Future studies can utilise the readily available WIM data and consider longer spans and multi-lane effects. In some cases, it may be worthwhile to consider local traffic conditions, for instance, using Bridge-WIM measurements. Accurate statistical characterisation of traffic loads benefits all of the procedures suggested in this dissertation, particularly the lower-bound approach (Chapter 3), which relies entirely on variable load effects.

The methodology developed for incorporating laboratory and monitoring data (Chapter 5) is versatile and can be extended to other forms of in-situ performance measurements. Promising applications include acoustic emission, strain and displacement fields, and long-term monitoring equipment. Each of these will require specific post-processing procedures for laboratory data and point to future needs for empirical testing. Attention should also be paid to practical aspects possibly hindering their application, such as sensor costs, durability, and signal quality. For structures occurring in large numbers, it may be worthwhile to establish databases that relate resistance ratios to indicator values. In specific cases, tables could be compiled with pre-computed reliability outcomes based on assumed observations. During testing, the corresponding observational 'path' would be identified, eliminating the need for Bayesian calculations during the test.

The hierarchical Bayesian modelling approach adopted in Chapter 6 to address spatial correlation and system reliability relies on the correct identification of component populations. Further research is warranted to determine which characteristics make components sufficiently similar to be represented as one type. These findings will be crucial for extending the methodology to broader families of structures. In particular, the epistemic uncertainty across different structures and structural types requires attention. Instead of proof load testing each structure of a repeated design, visual inspections and extensive checklists could be created to establish condition and representativeness. Another promising avenue for research is the incorporation of time-dependent aspects, along with laboratory and monitoring data, into the proposed system reliability evaluation procedure.

Finally, to maximise the impact of the proposed methodology, international collaboration is essential. Throughout this research, interaction has been pursued with IABMAS and *fib*, and efforts continue to document progress and harmonise testing procedures. As ageing infrastructure and uncertainties in structural resistance pose global challenges, local advances in proof load testing can yield international benefits. Such collaboration enhances research and accelerates codification, but also carries the risk of oversimplification in attempts to reach agreement. As stated in the previous section, proof load testing can provide uniquely accurate assessments when employed in a structural reliability updating context—yet, this strength may be undermined if reduced to one-size-fits-all tables. Future developments should therefore focus on harmonising procedures while preserving the flexibility needed to account for application-specific conditions.

Acknowledgements

The author wishes to express his gratitude and sincere appreciation to the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat) for financing the research work. Special thanks are due to the *doctoral committee* and *user committee* members, whose guidance and critical questions have been invaluable in shaping this work.

Bibliography

As an author

Addonizio, G., De Vries, R., Lantsoght, E. O. L. & Losanno, D. (2024). Reliability of I-girder PC bridges through proof load testing: Preliminary results. *Proceedings of the 13th International Conference on Bridge Maintenance, Safety and Management (IABMAS), Copenhagen, Denmark*, 306–314. https://doi.org/10.1201/9781003483755-32

De Vries, R. & Lantsoght, E. O. L. (2025a). *Code underlying the PhD thesis: Structural reliability updating through proof load testing*. Dataset, 4TU.ResearchData. Delft, The Netherlands. https://doi.org/10.4121/59968977-7d4a-4b98-be3c-dc9360b931e1

De Vries, R. & Lantsoght, E. O. L. (2025b). *Data underlying the PhD thesis: Structural reliability updating through proof load testing*. Dataset, 4TU.ResearchData. Delft, The Netherlands. https://doi.org/10.4121/f14b0777-b06d-43a9-8282-e751b9e84555

De Vries, R. & Lantsoght, E. O. L. (2022). *State-of-the-art report on probabilistic substantiation of proof load testing*. Stevin Report 25.5-21-01, Delft University of Technology. Delft, The Netherlands.

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M. & Fennis, S. A. A. M. (2022). Reliability assessment of existing reinforced concrete bridges and viaducts through proof load testing. *Proceedings of the 11th International Conference on Bridge Maintenance, Safety and Management (IABMAS), Barcelona, Spain*, 467–475. https://doi.org/10.1201/9781003322641-54 (Appendix A).

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M. & Naaktgeboren, M. (2023a). Proof load testing method by the American association of state highway and transportation officials and suggestions for improvement. *Transportation Research Record*, *2677*(11), 245–257. https://doi.org/10.1177/03611981231165026 (Appendix B).

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M. & Fennis, S. A. A. M. (2023b). Time-dependent reliability assessment of existing concrete bridges with varying knowledge levels by proof load testing. *Structure and Infrastructure Engineering: IABMAS 2022 Special Issue*, *20*(7-8), 1053–1067. https://doi.org/10.1080/15732479.2023.2280712 (Appendix C).

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M. & Naaktgeboren, M. (2023c). Proof load testing method by AASHTO and suggestions for improvement. *Proceedings of the 14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP), Dublin, Ireland*. http://hdl.handle.net/2262/103200

De Vries, R., Steenbergen, R. D. J. M. & Maljaars, J. (2023d). Annual reliability requirements for bridges and viaducts. *Heron*, *68*(2), 91–122. https://heronjournal.nl/68-2/1.html

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M., Hendriks, M. A. N. & Naaktgeboren, M. (2024). Structural reliability updating using monitoring data from in-situ load testing and laboratory test results. *Proceedings of the 13th International Conference on Bridge Maintenance, Safety and Management (IABMAS), Copenhagen, Denmark*, 409–417. https://doi.org/10.1201/9781003483755-44

De Vries, R., Steenbergen, R. D. J. M. & Vrouwenvelder, A. C. W. M. (2025a). Bayesian structural reliability updating using a population track record. *Reliability Engineering & System Safety*, *225*, 110644. https://doi.org/10.1016/j.ress.2024.110644

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M., Hendriks, M. A. N. & Naaktgeboren, M. (2025b). Structural reliability updating on the basis of proof load testing and monitoring data. *Engineering Structures*, *330*, 119863. https://doi.org/10.1016/j.engstruct.2025.119863 (Appendix D)

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M., Hendriks, M. A. N. & Fennis, S. A. A. M. (2026). Addressing spatial correlation and system reliability in proof load testing of reinforced concrete bridges and viaducts. *Structural Safety, Under review*. (Appendix E).

Other works

AASHTO. (2022). *Manual for bridge evaluation – 3rd edition, with 2020 and 2022 interim revisions*. Standard, American Association of State Highway and Transportation Officials. Washington, D.C., USA.

Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T. & Zinkov, R. (2023). PyMC: A modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*, *9*, e1516. https://doi.org/10.7717/peerj-cs.1516

ACI. (2013). *Code requirements for load testing of existing concrete structures*. Standard 437.2M, American Concrete Institute, Farmington Hills. Michigan, USA.

ACI. (2023). *Load testing of concrete structures – Code and commentary*. Standard 437.2M, American Concrete Institute, Farmington Hills. Michigan, USA.

ACI-DAfStb. (2025). "Shear databases" for structural concrete members [Accessed 12 March 2025].

Ahmad, S. H., Xie, Y. & Yu, T. (1996). Effectiveness of shear reinforcement for normal and high-strength concrete beams. *Structural Engineering Review*, *8*(1), 65–78.

Alampalli, S., Frangopol, D. M., Grimson, J., Kosnik, D., Halling, M., Lantsoght, E. O. L. & Zhou, Y. E. (2019). *Primer on bridge load testing*. Transportation Research Circular E-C257, Transportation Research Board, Washington, D.C., USA. https://www.trb.org/Publications/Blurbs/179887.aspx

Ang, A. H.-S. & Tang, W. H. (2006). *Probability concepts in engineering: Emphasis on applications to civil and environmental engineering* (2nd edition). Wiley.

Benjamin, J. R. & Cornell, C. A. (2014). *Probability, statistics, and decision for civil engineers* (Illustrated edition). Dover.

Bentz, E. C. & Buckley, S. (2005). Repeating a classic set of experiments on size effect in shear of members without stirrups. *ACI Structural Journal*, *102*(6), 832–838. https://doi.org/10.14359/14791

Bernander, K. (1957). An investigation of the shear strength of concrete beams without stirrups or diagonal bars. *Proceedings of the RILEM Symposium, Stockholm, Sweden*, *1*, 211–214.

Bernhardt, C. J. & Fynboe, C. C. (1986). High strength concrete beams. *Nordic Concrete Research*, *5*, 19–26.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer. https://link.springer.com/book/9780387310732

Bolle, G., Schacht, G. & Marx, S. (2011). Loading tests of existing concrete structures – Historical development and present practise. *Proceedings of the 10th fib Symposium, Prague, Czech Republic*.

Botte, W., Vereecken, E. & Caspeele, R. (2023). Random field modelling of spatial variability in concrete – a review. *Structure and Infrastructure Engineering, 21*, 1047–1060. https://doi.org/10.1080/15732479.2023.2248102

Box, G. & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley.

Breitung, K. (1984). Asymptotic approximation for multinormal integrals. *Journal of Engineering Mechanics, 110*(3), 357–366. https://doi.org/10.1061/(ASCE)0733-9399(1984)110:3(357)

Casas, J. R. & Gómez, J. D. (2013). Load rating of highway bridges by proof-loading. *KSCE Journal of Civil Engineering, 17*(3), 556–567. https://doi.org/10.1007/s12205-013-0007-8

Celati, S., Natali, A., Salvatore, W., Björnsson, I. & Thöns, S. (2025). Spatial and time-dependent reliability analysis for post-tensioned concrete decks subjected to multiple failure modes. *Structural Safety, 117*, 102634. https://doi.org/10.1016/j.strusafe.2025.102634

CEN. (2011). *Eurocode 2: Design of concrete structures – Part 1-1: General rules and rules for buildings*. Standard EN 1992-1-1+C2:2011, European Committee for Standardisation (CEN). Brussels, Belgium.

CEN. (2019). *Eurocode 0: Basis of structural design*. Standard EN 1990+A1+A1/C2:2019, European Committee for Standardisation (CEN). Brussels, Belgium.

Chana, P. S. (1981). *Some aspects of modelling the behaviour of reinforced concrete under shear loading*. Technical report 543, Cement and Concrete Association. Wexham Springs, UK.

Christensen, C. O., Zhang, F., Zarate Garnica, G. I., Lantsoght, E. O. L., Goltermann, P. & Smith, J. W. (2022). Identification of stop criteria for large-scale laboratory slab tests using digital image correlation and acoustic emission. *Infrastructures, 7*(3). https://doi.org/10.3390/infrastructures7030036

Collins, M. P. & Kuchma, D. (1999). How safe are our large, lightly reinforced concrete beams, slabs and footings? *ACI Structural Journal, 96*(4), 482–490. https://doi.org/10.14359/684

DAfStb. (2000). *DAfStb-Richtlinie: Belastungsversuche an Betonbauwerken (DAfStb guidline: Load tests on concrete structures)*. Guideline, Deutscher Ausschuss für Stahlbeton (German Committee for Reinforced Concrete). Berlin, Germany.

DAfStb. (2020). *DAfStb-Richtlinie: Belastungsversuche an Betonbauwerken (DAfStb guidline: Load tests on concrete structures)*. Guideline, Deutscher Ausschuss für Stahlbeton (German Committee for Reinforced Concrete). Berlin, Germany.

Deng, L., Yu, Y., Zou, Q. & Cai, C. S. (2015). State-of-the-art review of dynamic impact factors of highway bridges. *Journal of Bridge Engineering, 20*(5), 04014080. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000672

Der Kiureghian, A. (2022). *Structural and system reliability*. Cambridge University Press. https://doi.org/10.1017/9781108991889

Diamantidis, D., Tanner, P., Holický, M., Madsen, H. O. & Sykora, M. (2025). On reliability assessment of existing structures. *Structural Safety, 113*, 102452. https://doi.org/10.1016/j.strusafe.2024.102452

Dilks, D. W., Canale, R. P. & Meier, P. G. (1992). Development of Bayesian Monte Carlo techniques for water quality model uncertainty. *Ecological Modelling, 62*(1), 149–162. https://doi.org/10.1016/0304-3800(92)90087-U

Ditlevsen, O. (1979). Narrow reliability bounds for structural systems. *Journal of Structural Mechanics, 7*(4), 453–472. https://doi.org/10.1080/03601217908905329

Ditlevsen, O. & Madsen, H. O. (1996). *Structural reliability methods*. John Wiley & Sons.

Ditlevsen, O. & Vrouwenvelder, A. (1994). "Objective" low informative priors for Bayesian inference from totally censored Gaussian data. *Structural Safety, 16*(3), 175–188. https://doi.org/10.1016/0167-4730(94)00020-Q

Eaton, M. L. & Sudderth, W. D. (2004). Properties of right Haar predictive inference. *Sankhyā: The Indian Journal of Statistics*, *66*(3), 487–512.

ElBatanouny, M., Schacht, G. & Bolle, G. (2019). *History of load testing of bridges*. Book chapter, Load testing of bridges: Current practice and diagnostic load testing, Editor E. O. L. Lantsoght. https://doi.org/10.1201/9780429265426-3

fib. (2016). *Partial factor methods for existing concrete structures*. Bulletin 80, International Federation for Structural Concrete (*fib*), Lausanne, Switzerland. https://doi.org/10.35789/fib.BULL.0080

Faber, M. H. (2000). Reliability based assessment of existing structures. *Progress in Structural Engineering and Materials*, *2*(2), 247–253. https://doi.org/10.1002/1528-2716(200004/06)2:2<247::AID-PSE31>3.0.CO;2-H

Faber, M. H., V., V. D. & Stewart, M. G. (2000). Proof load testing for bridge assessment and upgrading. *Engineering Structures*, *22*, 1677–1689. https://doi.org/10.1016/S0141-0296(99)00111-X

Fennis, S. A. A. M., Hordijk, D. A., Hemert, P. H. A. & De Boer, A. (2014). Proefbelasting viaduct Vlijmen-Oost: Onderzoek naar beoordelingsmethode voor bestaande betonconstructies (Proof load test viaduct Vlijmen-Oost: Research into assessment of existing concrete structures). *Cement*, *66*(5), 40–45. https://www.cementonline.nl/artikel/proefbelasting-viaduct-vlijmen-oost

FHWA. (2018). *Weigh-in-motion pocket guide – Part 1: WIM technology, data acquisition, and procurement guide*. FHWA-PL-18-015, Federal Highway Administration (FHWA). Washington D.C., USA.

Fisher, R. A. (1956). *Statistical methods and scientific inference*. Oliver; Boyd.

Frangopol, D. M., Yang, D. Y., Lantsoght, E. O. L. & Steenbergen, R. D. J. M. (2019). *Reliability-based analysis and life-cycle management of load tests*. Book chapter, Load testing of bridges: Proof load testing and the future of load testing, Editor E. O. L. Lantsoght. https://doi.org/10.1201/9780429265969-9

Geisser, S. (1993). *Predictive inference: An introduction*. Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013). *Bayesian data analysis* (3rd edition). Chapman & Hall / CRC Press. https://doi.org/10.1201/b16018

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596

Geyer, S., Papaioannou, I. & Straub, D. (2023). Spatial modeling of concrete strength based on data. *Structural Safety*, *103*, 102345. https://doi.org/10.1016/j.strusafe.2023.102345

Ghanem, R. G. & Spanos, P. D. (1991). *Stochastic finite elements: A spectral approach*. Springer. https://doi.org/10.1007/978-1-4612-3094-6

Golub, G. H. & Van Loan, C. F. (2013). *Matrix computations* (4th edition). Johns Hopkins University Press.

Gong, C. & Zhou, W. (2017). Improvement of equivalent component approach for reliability analyses of series systems. *Structural Safety*, *68*, 65–72. https://doi.org/10.1016/j.strusafe.2017.06.0015

González, A., Žnidarič, A., Casas, J. R. & O'Brien, E. J. (2009). *Recommendations on dynamic amplification allowance in assessment of bridges*. Technical report, ARCHES Deliverable D10, FEHRL, European Commission. Brussels, Belgium. http://hdl.handle.net/10197/6238

Gosset, W. S. (1908). The probable error of a mean. *Biometrika*, *6*(1), 1–25. https://doi.org/10.1093/biomet/6.1.1

Grigoriu, M. & Hall, W. B. (1984). Probabilistic models for proof load testing. *Journal of Structural Engineering*, *110*(2), 260–274. https://doi.org/10.1061/(ASCE)0733-9445(1984)110:2(260)

Gumbel, E. J. (1954). *Statistical theory of extreme values and some practical applications* (Vol. 33). National Bureau of Standards.

Hallgren, M. (1994). *Shear tests on reinforced high and normal strength concrete beams without stirrups*. KTH Royal Institute of Technology. Stockholm, Sweden.

Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica, 19*(2), 491–544.

Harrewijn, T. L., Vergoossen, R. P. H. & Lantsoght, E. O. L. (2021). Reverse engineering of existing reinforced concrete slab bridges. *IABSE Congress: Resilient Technologies for Sustainable Infrastructure, Christchurch, New Zealand*. https://doi.org/10.2749/christchurch.2021.0140

Hasofer, A. M. & Lind, N. C. (1974). Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division, 100*(1), 111–121. https://doi.org/10.1061/JMCEA3.0001848

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Helton, J. C. & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety, 81*(1), 23–69. https://doi.org/10.1016/S0951-8320(03)00058-9

Hohenbichler, M., Gollwitzer, S., Kruse, W. & Rackwitz, R. (1987). New light on first- and second-order reliability methods. *Structural Safety, 4*, 267–284. https://doi.org/10.1016/0167-4730(87)90002-6

Holický, M. (2019). New European document on assessment of existing structures and building stock. *IOP Conference Series: Earth and Environmental Science, 290*(1), 012133. https://doi.org/10.1088/1755-1315/290/1/012133

Hornberger, G. M. & Spear, R. C. (1980). Eutrophication in peel inlet–I. The problem-defining behavior and a mathematical model for the phosphorus scenario. *Water Research, 14*(1), 29–42. https://doi.org/10.1016/0043-1354(80)90039-1

ISO. (2015). *General principles on reliability for structures*. Standard 2394, International Organisation for Standardisation. Geneva, Switzerland.

Jacinto, L., Neves, L. A. C. & Santos, L. O. (2015). Bayesian assessment of an existing bridge: A case study. *Structure and Infrastructure Engineering, 12*(1), 1–17. https://doi.org/10.1080/15732479.2014.995105

JCSS. (2001). *Probabilistic assessment of existing structures*. Joint Committee on Structural Safety, RILEM Publications S.A.R.L., Editor D. Diamantidis.

JCSS. (2015). *Probabilistic model code*. Joint Committee on Structural Safety. https://www.jcss-lc.org/jcss-probabilistic-model-code/

Jeffreys, H. (1961). *Theory of probability* (3rd edition). Oxford University Press.

Jones, E. M. C. & Iadicola, M. A. (2018). *A good practices guide for digital image correlation*. International Digital Image Correlation Society. https://doi.org/10.32720/idics/gpg.ed1

Jonkman, S. N., Steenbergen, R. D. J. M., Morales-Nápoles, O., Vrouwenvelder, A. C. W. M. & Vrijling, J. K. (2015). *Probabilistic design: Risk and reliability analysis in civil engineering*. Lecture notes, CIE4130, Delft University of Technology. Delft, The Netherlands.

Kapoor, M. (2021). *Optimal structural health information approaches for the efficient classification and management of structural systems*. PhD thesis, Department of Civil Engineering, Technical University of Denmark (DTU). Kongens Lyngby, Denmark. https://orbit.dtu.dk/en/publications/optimal-structural-health-information-approaches-for-the-efficien

Kass, R. E. & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91*(435), 1343–1370.

KIVI. (1938). *Voorschrift voor het ontwerpen en voor het vervaardigen en opstellen van stalen bruggen (Regulations for the design and for the manufacture and erection of steel bridges)*. Standard N 1008, Koninklijk Instituut van Ingenieurs. The Hague, The Netherlands.

KIVI. (1950). *Gewapend-beton-voorschriften (Reinforced concrete regulations)*. Standard N 1009, Koninklijk Instituut van Ingenieurs. The Hague, The Netherlands.

Kloek, T. & Van Dijk, H. K. (1978). Bayesian estimates of equation system parameters: An application of integration by monte carlo. *Econometrica*, *46*(1), 1–19. https://doi.org/10.2307/1913641

Koekkoek, R. T., Lantsoght, E. O. L., Yang, Y., De Boer, A. & Hordijk, D. A. (2016). Defining loading criteria for proof loading of existing reinforced concrete bridges. *fib Symposium 2016: Performance-based approaches for concrete structures, Cape Town, South Africa*.

Koekkoek, R. T., Lantsoght, E. O. L., Yang, Y. & Hordijk, D. A. (2016). *Analysis report for the assessment of viaduct De Beek by proof loading*. Stevin report 25.5-16-01, Delft University of Technology. Delft, The Netherlands.

Kong, P. & Rangan, B. V. (1997). Reinforced high strength concrete (HSC) beams in shear. *Australian Civil Engineering Transactions*, *39*(1), 43–50.

Krefeld, W. J. & Thurston, C. W. (1966). Studies of the shear and diagonal tension strength of simply supported reinforced concrete beams. *Journal of the American Concrete Institute*, *63*(4). https://doi.org/10.14359/7633

Lantsoght, E. O. L. (Ed.). (2019a). *Load testing of bridges: Current practice and diagnostic load testing*. CRC Press. https://doi.org/10.1201/9780429265426

Lantsoght, E. O. L. (Ed.). (2019b). *Load testing of bridges: Proof load testing and the future of load testing*. CRC Press. https://doi.org/10.1201/9780429265969

Lantsoght, E. O. L. (2023). Assessment of existing concrete bridges by load testing: Barriers to code implementation and proposed solutions. *Structure and Infrastructure Engineering*, *20*(7-8), 1002–1014. https://doi.org/10.1080/15732479.2023.2264825

Lantsoght, E. O. L., Van der Veen, C., De Boer, A. & Walraven, J. C. (2013). Recommendations for the shear assessment of reinforced concrete slab bridges from experiments. *Structural Engineering International*, *23*(4), 418–426. https://doi.org/10.2749/101686613X13627347100239

Lantsoght, E. O. L., Yang, Y., Van der Veen, C., Hordijk, D. A. & De Boer, A. (2019). Stop criteria for flexure for proof load testing of reinforced concrete structures. *Frontiers in Built Environment*, *5*. https://doi.org/10.3389/fbuil.2019.00047

Lantsoght, E. O. L., Van der Veen, C., De Boer, A. & Hordijk, D. A. (2017a). State-of-the-art on load testing of concrete bridges. *Engineering Structures*, *150*, 231–241. https://doi.org/10.1016/j.engstruct.2017.07.050

Lantsoght, E. O. L., Van der Veen, C., De Boer, A. & Hordijk, D. A. (2017b). Proof load testing of reinforced concrete slab bridges in the Netherlands. *Structural Concrete*, *18*(2), 597–606. https://doi.org/10.1002/suco.201600171

Lantsoght, E. O. L., Koekkoek, R. T., Hordijk, D. & De Boer, A. (2017c). Towards standardisation of proof load testing: Pilot test on viaduct Zijlweg. *Structure and Infrastructure Engineering*, *14*(3), 365–380. https://doi.org/10.1080/15732479.2017.1354032

Lantsoght, E. O. L., Koekkoek, R. T., Van der Veen, C., Hordijk, D. A. & De Boer, A. (2017d). Pilot proof-load test on viaduct De Beek: Case study. *Journal of Bridge Engineering*, *22*(12). https://doi.org/10.1061/(ASCE)BE.1943-5592.0001131

Laplace, P.-S. (1814). *Essai philosophique sur les probabilités (A philosophical essay on probabilities)*. Courcier, Paris, France.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. & Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, *20*(5), 491–505. https://doi.org/10.1002/net.3230200503

Leonhardt, F. & Walther, R. (1962). *Schubversuche an einfeldrigen Stahlbetonbalken mit und ohne Schubbewehrung (Shear tests on simply supported reinforced concrete beams with and without shear reinforcement)*. Heft 151, DAfStb. Berlin, Germany.

Lichtenstein, A. G. (1993). *Bridge rating through nondestructive load testing*. Technical report, NCHRP Project 12-28(13)A, National Cooperative Highway Research Program. Washington D.C., USA.

Lin, T. S. & Nowak, A. S. (1984). Proof loading and structural reliability. *Reliability Engineering, 8*, 85–100. https://doi.org/10.1016/0143-8174(84)90057-X

Lubell, A. S., Sherwood, E. G., Bentz, E. C. & Collins, M. P. (2004). Safe shear design of large wide beams. *Concrete International, 26*(1), 66–78.

MacKay, D. J. C. (1991). *Bayesian methods for adaptive models*. PhD thesis, California Institute of Technology. Pasadena, USA. https://doi.org/10.7907/H3A1-WM07

Madsen, H. O., Krenk, S. & Lind, N. C. (2006). *Methods of structural safety*. Dover Publications.

Marx, S., Grünberg, J. & Schacht, G. (2019). Methods to evaluate test results based on small sample sizes. *Civil Engineering Design, 1*(2), 74–84. https://doi.org/10.1002/cend.201900012

Melchers, R. E. & Beck, A. T. (2018). *Structural reliability analysis and prediction* (3rd edition). John Wiley & Sons.

Melhem, M. M., Caprani, C. C., Stewart, M. G. & Zhang, S. (2020). *Bridge assessment beyond the AS 5100 deterministic methodology*. Research report AP-R617-20, Austroads. Sydney, Australia. https://austroads.gov.au/publications/bridges/ap-r617-20

Metropolis, N. (1953). Equations of state calculations by fast computing machines. *Chemical Physics, 21*(6), 1087–1091. https://doi.org/10.1063/1.1699114

NEN. (2011). *Beoordeling van de constructieve veiligheid van een bestaand bouwwerk bij verbouw en afkeuren – Grondslagen (Structural safety assessment of existing structures for renovation and disapproval – Basis)*. Standard 8700, Nederlands Normalisatie-instituut. Delft, The Netherlands.

Nishijima, K. & Faber, M. H. (2007). Bayesian approach to proof loading of quasi-identical multi-components structural systems. *Civil Engineering and Environmental Systems, 24*(2), 111–121. https://doi.org/10.1080/10286600601159172

O'Brien, E. J., O'Connor, A. J. & Arrigan, J. E. (2012). Procedures for calibrating Eurocode traffic load model 1 for national conditions. *Proceedings of the 6th International Conference on Bridge Maintenance, Safety and Management (IABMAS 2012)*, 2597–2603.

Özden, K. (1967). *An experimental investigation on the shear strength of reinforced concrete beams: Tests performed at structural research laboratory, technical university of denmark*. Technical University of Istanbul. Istanbul, Turkey.

Papoulis, A. & Pillai, S. U. (2002). *Probability, random variables, and stochastic processes* (4th edition). McGraw-Hill.

Qian, S. S., Stow, C. A. & Borsuk, M. E. (2003). On Monte Carlo methods for Bayesian inference. *Ecological Modelling, 159*(2-3), 269–277. https://doi.org/10.1016/S0304-3800(02)00299-5

Rackwitz, R. & Schrupp, K. (1985). Quality control, proof testing and structural reliability. *Structural Safety, 2*, 239–244. https://doi.org/10.1016/0167-4730(85)90030-X

Regan, P. E. (1971). *Shear in reinforced concrete – An experimental study*. CIRIA-report, Imperial College of Science and Technology, Department of Civil Engineering. London, United Kingdom.

Reineck, K.-H., Bentz, E. C., Fitik, B., Kuchma, D. A. & Bayrak, O. (2013). ACI-DAfStb database of shear tests on slender reinforced concrete beams without stirrups. *ACI Structural Journal, 110*(5), 867–876. https://doi.org/10.14359/51685839

Reineck, K.-H., Bentz, E. C., Fitik, B., Kuchma, D. A. & Bayrak, O. (2014). ACI-DAfStb databases for shear tests on slender reinforced concrete beams with stirrups. *ACI Structural Journal*, *111*(5), 1147–1156. https://doi.org/10.14359/51686819

Roscoe, K., Diermanse, F. & Vrouwenvelder, T. (2015). System reliability with correlated components: Accuracy of the equivalent planes method. *Structural Safety*, *57*, 53–64. https://doi.org/10.1016/j.strusafe.2015.07.006

Russell, B. T. & Huang, W. K. (2021). Modeling short-ranged dependence in block extrema with application to polar temperature data. *Environmetrics*, *32*, e2661. https://doi.org/10.1002/env.2661

RWS. (2013). *Richtlijnen Beoordeling Kunstwerken (RBK) – Beoordeling van de constructieve veiligheid van een bestaand kunstwerk bij verbouw, gebruik en afkeur (Guidelines for assessing infrastructure – Assessment of the structural safety of an existing infrastructure for renovation, use and rejection)*. Report, Version 1.1, Rijkswaterstaat, Utrecht, The Netherlands. https://standaarden.rws.nl/link/standaard/1059

Sarkhosh, R., Walraven, J. & Den Uijl, J. (2015). Shear-critical reinforced concrete beams under sustained loading – Part I: Experiments. *Heron*, *60*(3), 181–206. https://heronjournal.nl/60-3/3.html

Savage, L. J. (1961). The foundations of statistics reconsidered. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*.

Schacht, G., Bolle, G. & Marx, S. (2019). *Load testing of concrete building constructions*. Book chapter, Load testing of bridges: Proof load testing and the future of load testing, Editor E. O. L. Lantsoght. https://doi.org/10.1201/9780429265969-4

Schmidt, J. W., Halding, P. S., Jensen, T. W. & Engelund, S. (2018). High magnitude loading of concrete bridges. *ACI Structural Journal*, *323*, 9.1–9.20. https://doi.org/10.14359/51702439

Schmidt, J. W., Thöns, S., Kapoor, M., Christensen, C. O., Engelund, S. & Sørensen, J. D. (2020). Challenges related to probabilistic decision analysis for bridge testing and reclassification. *Frontiers in Built Environment*, *6*. https://doi.org/10.3389/fbuil.2020.00014

Sherwood, E. G. (2008). *One-way shear behaviour of large, lightly-reinforced concrete beams and slabs*. PhD thesis, University of Toronto, Department of Civil Engineering. Toronto, Canada. http://hdl.handle.net/1807/119841

Sibuya, M. (1959). Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics*, *11*(2), 195–210.

Spaethe, G. (1994). Die Beeinflussung der Sicherheit eines Tragwerks durch Probebelastung (The influence of proof load testing on structural reliability). *Bauingenieur*, *69*, 459–468.

Steenbergen, R. D. J. M. & Vrouwenvelder, A. C. W. M. (2010). Safety philosophy for existing structures and partial factors for traffic loads on bridges. *Heron*, *55*(2), 123–139. https://heronjournal.nl/55-2/2.html

Stewart, M. G., Rosowsky, D. V. & Val, D. V. (2001). Reliability-based bridge assessment using risk-ranking decision analysis. *Structural Safety*, *23*(4), 397–405. https://doi.org/10.1016/S0167-4730(02)00010-3

Straub, D. & Papaioannou, I. (2015). Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, *141*(3). https://doi.org/10.1061/(ASCE)EM.1943-7889.0000839

Stroustrup, B. (2013). *The C++ programming language* (4th edition). Pearson Education.

Šýkora, M. & Holický, M. (2013). Assessment of uncertainties in mechanical models. *Applied Mechanics and Materials*, *378*, 13–18. https://doi.org/10.4028/www.scientific.net/AMM.378.13

Sýkora, M., Holický, M., Prieto, M. & Tanner, P. (2015). Uncertainties in resistance models for sound and corrosion-damaged rc structures according to en 1992-1-1. *Materials and Structures, 48*(10), 3415–3430. https://doi.org/10.1617/s11527-014-0409-1

Tvedt, L. (1983). *Two second order approximations to the failure probability*. Technical report RDIV/20-004-83, Det Norske Veritas (DNV). Høvik, Norway.

Val, D. V. & Stewart, M. G. (2019). *Determination of remaining service life of reinforced concrete bridge structures in corrosive environments after load testing*. Book chapter, Load testing of bridges: Proof load testing and the future of load testing, Editor E. O. L. Lantsoght. https://doi.org/10.1201/9780429265969-10

Vanmarcke, E. H. (1983). *Random fields: Analysis and synthesis*. MIT Press.

Vu, K. A. T. & Stewart, M. G. (2000). Structural reliability of concrete bridges including improved chloride-induced corrosion models. *Structural Safety, 22*(4), 313–333. https://doi.org/10.1016/S0167-4730(00)00018-7

Wang, N., Ellingwood, B. R. & Zureick, A. H. (2011). Bridge rating using system reliability assessment. II: Improvements to bridge rating practices. *Journal of Bridge Engineering, 16*(6), 863–871. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000171

Wasserman, L. (2004). *All of statistics: A concise course in statistical inference*. Springer. https://doi.org/10.1007/978-0-387-21736-9

Wilkinson, M. D., Dumontier, M., Aalbersberg, IJ. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W. & ... (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Yang, Y., Van der Ham, H. W. M. & Naaktgeboren, M. (2021). Shear capacity of RC slab structures with low reinforcement ratio – An experimental approach. *Proceedings to the 18th fib Symposium, Lisbon, Portugal*.

Yuefei, L., Dagang, L. & Xueping, F. (2014). Reliability updating and prediction of bridge structures based on proof loads and monitored data. *Construction and Building Materials, 66*, 795–804. https://doi.org/10.1016/J.CONBUILDMAT.2014.06.025

Zabell, S. L. (1992). R. A. Fisher and the fiducial argument. *Statistical Science, 7*(3), 369–387. https://doi.org/10.1214/ss/1177011233

Zarate Garnica, G. I., De Vries, R. & Lantsoght, E. O. L. (2022). *Analysis report of reinforced concrete slabs for stop criteria*. Stevin report 25.5-22-02, Delft University of Technology. Delft, The Netherlands.

Zarate Garnica, G. I. & Lantsoght, E. O. L. (2021). Stop criteria for proof load testing of reinforced concrete structures. *Proceedings of the 2021 session of the 13th fib International PhD Symposium in Civil Engineering*, 195–202. https://phdsymp2020.sciencesconf.org

Zarate Garnica, G. I., Lantsoght, E. O. L. & Yang, Y. (2022). Monitoring structural responses during load testing of reinforced concrete bridges: A review. *Structure and Infrastructure Engineering, 18*(10-11), 1558–1580. https://doi.org/10.1080/15732479.2022.2063906

Zarate Garnica, G. I., Lu, J., Hendriks, M. A. N. & Lantsoght, E. O. L. (2024a). Shear experiments on straight reinforced concrete slabs. *Proceedings of International Structural Engineering and Construction, 11*(1). https://doi.org/10.14455/ISEC.2024.11(1).STR-07

Zarate Garnica, G. I., Lantsoght, E. O. L., Yang, Y. & Hendriks, M. A. N. (2024b). Shear stop criteria for reinforced concrete slab strips. *Proceedings of the 13th International Conference on Bridge Maintenance, Safety and Management (IABMAS), Copenhagen, Denmark*. https://doi.org/10.1201/9781003483755-36

Zhang, F. (2022). *Acoustic emission-based indicators of shear failure of reinforced concrete structures without shear reinforcement*. PhD thesis, Delft University of Technology. Delft, The Netherlands. https://resolver.tudelft.nl/uuid:9220a0c2-f4c1-46e6-a0a9-0069e4662730

Zhang, W.-H., Lu, D.-G., Qin, J., Thöns, S. & Faber, M. H. (2021). Value of information analysis in civil and infrastructure engineering: A review. *Journal of Infrastructure Preservation and Resilience, 2*. https://doi.org/10.1186/s43065-021-00027-0

Zheng, X., Yi, T.-H., Yang, D.-H. & Li, H.-N. (2023). Multisection optimization-based target proof load determination method for bridge load testing. *Journal of Bridge Engineering, 28*(6), 04023025. https://doi.org/10.1061/JBENF2.BEENG-6073

Zhou, Y. E. & Guzda, M. R. (2020). Bridge load rating through proof load testing for shear at dapped ends of prestressed concrete girders. *Frontiers in Built Environment, 6*, 117. https://doi.org/10.3389/fbuil.2020.00117

Nomenclature

Abbreviations

BMC	Bayesian Monte Carlo
CDF	Cumulative distribution function
COV	Coefficient of variation
DIC	Digital image correlation
FORM	First order reliability method
LVDT	Linear variable differential transducer
MCMC	Markov-chain Monte Carlo
PDF	Probability density function
SORM	Second order reliability method
WIM	Weigh-in-motion

Symbols

$\mathbb{1}[\cdot]$	Indicator function: yields 1 if the argument is true and 0 otherwise
b	Width of a section or lane
C_{0Q}	Time-invariant part of the traffic load uncertainty
E	Load effect
E_{PL}	Load effect achieved during proof load
$g(\cdot)$	Limit state function
G	Permanent load effect
G_{DL}	Dead load effect, i.e. self-weight
G_{SDL}	Super-imposed dead load effect, e.g. asphalt layer, railings
I	Indicator, e.g. function of strain, crack width, etc.
m	(Sample) mean
M	Bending moment
$\mathcal{N}(\cdot)$	Normally distributed
L	Span length
$p(\cdot)$	Probability distribution (Bayesian)
$\mathrm{P}(\cdot)$	Probability of
P_{f}	Failure probability
$P_{\mathrm{f,sys}}$	System failure probability
Q	Variable load effect, e.g. traffic load
Q_{k}	Characteristic value of the traffic load
Q_{PL}	Proof load effect
R	Resistance (to load effect)

s	(Sample) standard deviation
$t_\nu(\cdot)$	Student's t-distribution with ν degrees of freedom
U	Standard normal random variable
V	Coefficient of variation, or shear force
$w_{\max,\mathrm{w}}$	Maximum weighted crack width
$\mathbf{x}$	Data vector (Bayesian)
X	Random variable, resistance ratio ($X = R/E$)
$\mathbf{X}$	Vector of random variables
$\bar{x}$	Sample mean
Z	Limit state function, regular traffic situation
Z_{PL}	Limit state function, proof load testing situation
β	Reliability index, scale parameter of the Gumbel distribution
γ	Euler-Mascheroni constant (≈ 0.5772)
γ_{PL}	Proof load factor ($Q_{\mathrm{PL}}/Q_{\mathrm{k}}$)
γ_{t}	Transfer factor
ε_{s}	Steel strain
θ_E	Model uncertainty of the load effect
$\theta_{E,\mathrm{PL}}$	Model uncertainty of the proof load effect
θ_R	Model uncertainty of the resistance
$\boldsymbol{\theta}$	Vector of model parameters (Bayesian)
μ	Location parameter of the normal and Gumbel distribution
ρ	Pearson correlation coefficient
$\varphi(\cdot)$	Standard normal probability density function
ν	Matérn kernel shape parameter
$\Phi(\cdot)$	Standard normal cumulative distribution function

Terminology

The following list of technical terms and their descriptions is included to support the reader's understanding of the dissertation; it is not intended to establish formal or normative definitions.

Annual reliability	A reliability framework in which the annual probability of failure is evaluated conditional on survival up to the year under consideration.
Autocorrelation	The statistical dependence between different realisations of the same variable across time or, more generally, across spatial intervals.
Bayes' theorem	A mathematical rule, named after Thomas Bayes, for reversing conditional probabilities, enabling the probability of a cause to be determined given an observed effect.

Characteristic load effect	A representative high load effect, derived either from a codified load model or from a specified upper fractile of its statistical distribution (e.g. 0.999).
Component	A structural element, such as a beam or slab, or more generally, an element considered within a system reliability framework.
Consequence class	A categorisation of the severity of consequences associated with structural failure or assigned importance, commonly defined in accordance with the Eurocodes.
Disapproval level	Minimum acceptable reliability level before intervention is required (Dutch NEN 8700 concept)
Design life	The intended period for which a structure is designed to fulfil its function, typically 50–100 years.
Design load	A load level derived by probabilistic consideration of uncertainties in both resistance and load effects, usually determined by semi-probabilistic procedures in standards.
Hierarchical model	A Bayesian model in which parameters are themselves treated as random variables, allowing separation of global (population-level) and local (component-level) effects.
Indicator	A quantity representing structural performance, measured or inferred from observations during testing, such as strain or crack width.
Likelihood	A function quantifying how likely, or compatible, observed data are with specific parameter values.
Limit state function	A mathematical expression defining the boundary between safe and failed structural states, typically formulated as the difference between resistance and load effect.
Load effect	The structural response resulting from applied loads, for example, bending moments, shear forces, or stresses.
Loading protocol	A specification of the procedure for conducting a load test, including the magnitude and number of load increments, their duration, and any unloading steps.
Low-informative	Describing a prior distribution that reflects limited but physically plausible knowledge, typically characterised by a large spread but still realistic.
Lower-bound approach	A method in which only the load effect achieved during proof load testing is used to infer a minimum resistance.
Nominal crack width	Results from the consolidation of several smaller cracks that are closely clustered, forming a single significant crack.

Posterior distribution	The updated probability distribution of an uncertain parameter obtained after combining the prior distribution with observed data.
Prior distribution	A probability distribution expressing beliefs about an uncertain parameter before considering new data.
Proof load factor	Factor by which a characteristic traffic load effect, codified or following from WIM-data, is multiplied to obtain the (target) proof load effect.
Proof load test	An in-situ test in which comparatively high loads are applied to demonstrate that a structure possesses sufficient capacity to withstand future loading.
Random field	A *stochastic process* addressing spatial variability and correlation of random quantities, for example, material properties, resistances, or load effects.
Reference period	The time interval associated with a time-dependent property or statistical description, such as an extreme value distribution or a reliability requirement.
Reliability requirement	The maximum acceptable probability of failure within a specified reference period.
Reliability updating	The process by which the reliability of a structure is reassessed in light of new information, such as monitoring data, or proof load test outcomes.
Resistance	The capacity of a structure to withstand load effects, such as bending moments or shear forces.
Series system	A structural system in which failure of any individual component results in failure of the entire system.
Stop criteria	Predefined limiting values of performance indicators, derived from measurements during testing, suggesting termination of the test.
Structural reliability	The probability that a structure performs without failure over a specified reference period, typically with respect to the ultimate limit state (ULS).
Target load	The prescribed load level to be reached during a proof load test to demonstrate adequate capacity.
Traffic load model	A standardised representation of traffic actions used to calculate load effects, typically comprising axle loads and distributed loads, for example Eurocode Load Model 1 (LM1).
Transfer factor	A multiplicative factor applied to the target proof load to compensate for uncertainty arising from untested components within a structural system.

A

IABMAS 2022 article

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M. & Fennis, S. A. A. M. (2022). Reliability assessment of existing reinforced concrete bridges and viaducts through proof load testing. *Proceedings of the 11 th International Conference on Bridge Maintenance, Safety and Management (IABMAS), Barcelona, Spain*, 467–475. https://doi.org/10.1201/9781003322641-54

Authors' postprint

A

Reliability assessment of existing reinforced concrete bridges and viaducts through proof load testing

R. de Vries
Delft University of Technology, Faculty of Civil Engineering and Geosciences
Netherlands Organization for Applied Scientific Research (TNO), Structural Reliability

E.O.L. Lantsoght
Delft University of Technology, Faculty of Civil Engineering and Geosciences
Universidad San Francisco de Quito, College of Sciences and Engineering

R.D.J.M. Steenbergen
Netherlands Organization for Applied Scientific Research (TNO), Structural Reliability
Ghent University, Faculty of Engineering and Architecture

S.A.A.M. Fennis
Rijkswaterstaat, Ministry of Infrastructure and Water Management

Mini-symposium: [MS05] Assessment of existing infrastructure assisted by field data

ABSTRACT: In the assessment of existing infrastructure performing only a desk study is often not sufficient to determine the structural reliability of a bridge or viaduct. For concrete structures gathering field data by performing a proof load test offers detailed information about the structural performance. However, the relation between the magnitude of the load and the structural reliability is not immediately clear. In the present study the challenges in determining the target load and the uncertainties that require attention are described. An approach is presented that addresses the time-dependent character of the structural reliability, the need for accurate stop-criteria, the knowledge level and spatial uncertainty. It is shown how both past traffic loads and a proof load test may contribute to the proven strength of a structure. The described methodology provides a starting point towards a flexible approach for proof load testing in which structure-specific information and requirements are considered.

Keywords: Deterioration, existing structures, load testing, proof loading, reliability, time-dependence

1 INTRODUCTION

Due to the constant aging of infrastructure, increased traffic load and traffic intensities, methods are explored by which the reliability of existing road bridges and viaducts can be assessed. In case limited information of the structure is available or its condition is of concern, load testing may be used to prove that the structure can still satisfactorily carry the live loads. A test conducted with this aim is referred to as a 'proof load test'. Historically, before complex structural analysis was commonplace, proof load testing was regularly performed prior to opening a bridge to the public. In a number of countries performing a proof load test before use is still required (Lantsoght et al. 2017).

The magnitude of the load to be applied, or target load, is of particular importance. If the, relatively large, target load is successfully carried by the structure then it has proven to be sufficiently structurally reliable for future use. The test can be performed using regular trucks, a special vehicle (Figure 1) or other methods such as a loading frame including ballast.

Often, for existing structures less stringent reliability requirements hold. Therefore, a design load used for new structures is not necessarily useful in the assessment of existing structures. In addition, the condition of the structure may be of particular concern due to the effect of deterioration or other time-dependent processes (Ellingwood 1996).

This article describes the challenges in determining the target load in relation to structural reliability and the associated uncertainties. In particular, the following aspects will be discussed: the evolution of the structural reliability with time, the reliability of stop-criteria, the level of knowledge about the structure and assessment at the system-level.

2 LITERATURE

2.1 *International standards*

Proof load testing is not a standardized assessment procedure in many countries. If national guidance is lacking, standards or guidelines from other countries can provide useful insight into accepted practices.

Figure 1: Pilot proof load test being performed on the Vechtbrug in the Netherlands (October 2016).

In the USA, the Manual for Bridge Evaluation (MBE) (AASHTO 2018) is used as a guideline for proof load testing. The target proof load is expressed in terms of the regular load model and is magnified by a proof load factor (X_p). Its default value (1.4) was derived in a basic probabilistic analysis (Lichtenstein 1993) that did not address the challenges described in this article. Another relevant American standard is the ACI 437.2M (ACI 2013) which describes the requirements for load testing of existing concrete buildings including loading protocols and acceptance criteria.

Recently the German committee for reinforced concrete has published a new version of its guideline for proof load tests on concrete structures (DAfStb 2020). The guideline is mainly intended for buildings, but speaks in more general terms such as structures and components. The magnitude of the proof load is expressed in a format that resembles the load effect in Equation (6.10) of EN 1990 (CEN 2019). An interesting aspect of the guideline is the consideration of multiple similar components. When a limited number of similar components are tested, the remaining uncertainty associated with their slight differences is accounted for by increasing the magnitude of the test load (Marx 2019).

2.2 *Research*

2.2.1 *General*

Proof load testing is still an active field of research and continues to gain attention due to the growing need for versatile assessment methods for existing structures (Lantsoght 2019).

It is desirable that the assessment of infrastructure is not overly conservative because that may lead to the replacement or upgrading of bridges that are actually satisfactory. Proof load magnitudes can vary depending on the load rating, dead/live load ratios, degradation, bridge age, reference period and prior service loads (Faber, Val & Stewart 2000).

In Casas & Gómez (2013) proof load factors are presented that were developed as part of the large scale ARCHES (Assessment and Rehabilitation of Central European Highway Structures) project. The factors are applied to the characteristic (or nominal) live-load obtained from a traffic load model (e.g. Eurocode LM1). A distinction is made between the case where bridge documentation is available and the case where it is not. Depending on the target reliability and estimated strength, different factors apply. Weigh-in-motion (WIM) data from five European countries, including the Netherlands, is used to describe the traffic load. Their study presents a step forward from current code-based approaches by making use of recent traffic load data and the option to include bridge documentation.

For existing structures flexibility is needed regarding the reference period, therefore the time-dependent nature of the structural reliability is of particular interest. An early description of the time-dependence in relation to proof load testing is found in Spaethe (1994). It is shown that during the proof load test the reliability of the structure is low, due to the relatively large load that is applied, but afterwards the reliability is increased – if the test was successful. In the more recent work by Schacht, Bolle & Marx (2019) and Frangopol et al. (2019) the decrease of reliability with time in case of deterioration is also recognized.

2.2.2 *System reliability*

One of the main aspects in which the assessment by load testing is different from the design of new structures is the influence of system action on the performance of an existing structure. Therefore, system reliability is of particular interest to proof load testing.

A system may be thought to be comprised of multiple components; the performance of all components needs to be combined according to a certain scheme to result in the system reliability. In this scheme the components may act in parallel or in series. In addition, the combined performance of a group of elements may interact with one component, or another group. In the context of system failure a diagram of the interaction is called a fault tree (Fussel 1975).

Various methods may be used to calculate the failure probability of a system. The Monte Carlo Simulation (MCS) is a straightforward method that is always applicable, but it is computationally expensive (Metropolis & Ulam 1949). For better computational efficiency the equivalent planes method (Roscoe, Diermanse & Vrouwenvelder 2015) is used in this article. The method is based on the equivalent component method (Gollwitzer & Rackwitz 1983) and the first-order system reliability method described by Hohenbichler & Rackwitz (1983). The reliability of the individual components may be determined using the first order reliability method (FORM) (Hasofer & Lind 1974) or any other method that also provides the influence coefficients of the random variables.

A

The equivalent planes method works on the basis of two components. The (linearized) limit state functions Z_i of two components may be written using the reliability index (β_i) of the component and the influence coefficients (α_{ij}) of all random variables present in the system:

$$Z_1 = \beta_1 + \alpha_{11}U_{11} + \alpha_{12}U_{12} + \cdots + \alpha_{1n}U_{1n} \quad (1a)$$

$$Z_2 = \beta_2 + \alpha_{21}U_{21} + \alpha_{21}U_{21} + \cdots + \alpha_{2n}U_{2n} \quad (1b)$$

In this equation U_{ij} are standard normally distributed random variables that are statistically independent (i.e. uncorrelated) within the component. However, auto-correlation $\rho_j = \rho(U_{1j}, U_{2j})$ may exist.

In case of more components, the combination process needs to be repeated several times until just one component remains. Each time two components are combined to give a new component that replaces the two original components. The most accurate results are obtained when the components with the highest correlation between the limit state functions are combined first in every step (Gong & Zhou 2017).

2.2.3 *Reliability updating*

Proof load testing is starting to be considered in the light of maintenance and durability. In particular, the so-called 'updating' of structural reliability as performed on the basis of Bayesian theory provides the opportunity to incorporate various sources of information. The theory can provide a mathematical basis for the updated distributions of the reliability (Yuefei, Dagang & Xueping 2014).

The more generally applicable Bayesian decision theory is also used in the context of proof load testing. It can provide decision support and the identification of information to aid in modelling and monitoring of structures (Schmidt et al. 2020). In Bayesian decision theory, today often mentioned in the context of value of information (Zhang et al. 2021), the state of information about an object results in three possible types of analysis: prior analysis, posterior analysis and pre-posterior analysis (see Figure 2). Each stage in the analysis has its own set of possibilities (E, X, A, Θ), dependent on earlier choices or outcomes. All possible paths lead to certain consequences or costs (C), which may also include the risk of losing human life.

In a prior analysis, the decision alternatives are directly associated with possible outcomes and the associated consequences. In a posterior analysis, additional information is added to update the probabilistic model. The added value of additional data may be quantified, even when it has not been collected yet – i.e. the *value of information* is studied. This gives rise to the pre-posterior analysis, where the extra costs of acquiring additional data is evaluated against possible gains. Proof load testing itself may also be viewed as an additional source of information about the structure and a pre-posterior analysis may be employed to determine if it is beneficial (Nishijima & Faber 2007).

3 CHALLENGES AND SUGGESTED APPROACH

Because of the long history of proof load testing, methods applied in practice possess a strong deterministic character or are based on basic probabilistic calculations that do not consider the aspects described in the following sections.

3.1 *Time-dependence*

In case a variable load acts on a structure, e.g. traffic load, the structural reliability is time-dependent. Additional time-dependent processes such as load trends

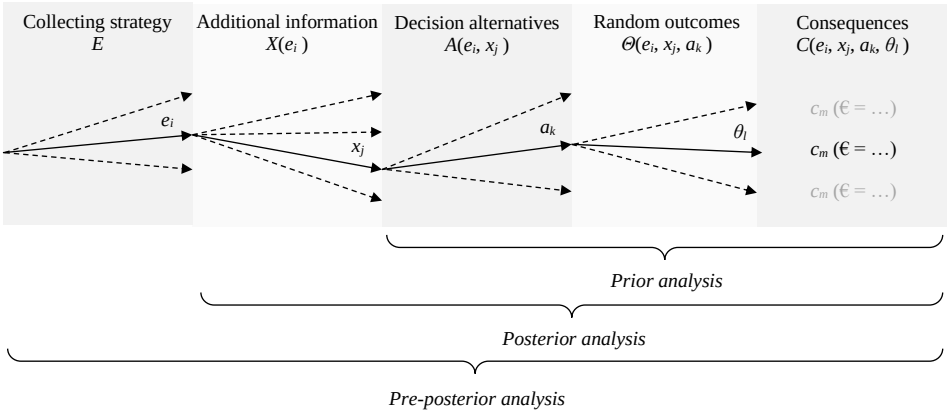


Figure 2: Analysis type depending on the state of information.

and material deterioration add to this effect. Also, during a proof load test the structural reliability is markedly lower than in normal operation. In relation to the desired reliability level, a fixed design reference period (e.g. 50 years) lacks the flexibility to accommodate the remaining functional life span of an existing structure. Moreover, the reliability level for new structures is larger than strictly required for human safety because of economic benefit. A structure that lasts longer may be profited from for a longer period of time (Steenbergen & Vrouwenvelder 2010).

To judge if the structural reliability is satisfactory after a successful proof load test, an assessment on the basis of annual reliability is suggested, but any small time period suffices. In this way the framework is also flexible regarding the remaining life span of an existing structure. Considering the time-dependence is also beneficial in relation to the proven strength by past traffic loads. In a sense, every truck passing a bridge may be viewed as a sort of proof load test, contributing to the proven strength of the bridge. Standard texts on reliability theory describe the proven strength from past loading via the 'bathtub curve' of the failure probability (Jonkman et al. 2015).

3.2 *Reliability of stop-criteria*

Although various stop-criteria have been developed (Zarate Garnica & Lantsoght 2021), little attention is paid to the link between structural reliability during the load test and the magnitude of the load. Reaching the target load, as calculated before the execution of the load test, may not always be feasible due to signs of distress appearing when the load is gradually applied. If the reliability of the stop-criteria is low, the proof load test may be aborted before the structure is actually near is maximum capacity.

Collapse or partial collapse of a structure during the proof load test is undesired. Considering the reliability during testing can mitigate the risk of collapse of the structure and provide decision support in the selection of sensors and measurement plans. In addition, if enough evidence of proof load tests reaching the target load successfully with some signs of distress is incorporated in the analysis, the risk at damage may be quantified instead of avoided (risk aversion).

3.3 *Knowledge level*

The knowledge level (available information in drawings, material tests, etc.) varies between structures, especially because of their age. Therefore, a flexible method is needed that can utilize various types of information. To assess the bending moment capacity the cross-sectional area of the reinforcement is of main interest. However, to assess the shear capacity the concrete quality is more valuable. Various data sources and their influence on the state of information are collected in Figure 3. In addition, a balance is sought between how much information is collected and analyzed before performing the proof load test and regarding the proof load test itself as the primary source of information about the structural performance (Medha et al. 2019).

In the suggested approach the state of information plays a key role in the reliability analysis preceding a proof load test. The basic information in a prior analysis is complemented by additional information in a

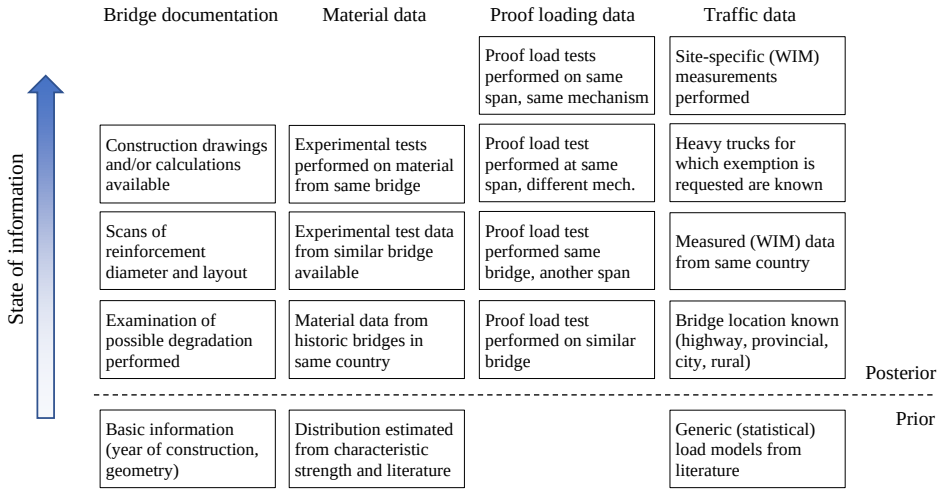


Figure 3: State of information considering various information sources.

A

posterior analysis (see Figure 2 and Figure 3). By introducing additional uncertainty into the probabilistic model it may subsequently be updated (reduced) through the application of Bayes' theorem:

$$\mathrm{P}(H \mid D) = \frac{\mathrm{P}(D \mid H)\,\mathrm{P}(H)}{\mathrm{P}(D)} \tag{2}$$

where H is the hypothesis (e.g. assume that parameter x of the model takes on value y, etc.) and D is the additional data that has become available. The data is a possible outcome of the model observed in real life and may take on many forms (e.g. a numeric value, a flag indicating failure, etc.). Because of this, various kinds of data (or evidence) can be combined.

Uncertainty in the resistance may be split into an objective (natural) and subjective part (model), often written as a multiplication ($\theta_R R$). In this formulation the model uncertainty (θ_R) is the parameter that is updated in a Bayesian approach (Lin & Nowak 1984). Very large uncertainties may be introduced purposely as 'objective' low informative priors (Ditlevsen & Vrouwenvelder 1994). If available, other broad prior distributions following from basic information such as the bridge span and traffic type are also suitable.

3.4 *System-level assessment*

A bridge or viaduct consists of physical components such as the bridge deck, supports, etc. However, also cross-sections and connections may be regarded as components. In addition, the same component may take part in different failure mechanisms. In a design all components are typically verified to achieve the desired structural reliability at element-level. It may subsequently be assumed that the reliability of the structure, i.e. the system, is at least equal to the reliability at the element level. Assessing the reliability of just one component, cross-section, or failure mechanism by proof load testing does not provide the reliability of the structure. Usually practical limitations apply with regard to the number of tests because of the hindrance caused to traffic and the costs of testing.

In a system-level assessment the performance of multiple components and spatial variability is incorporated. Also here Bayesian analysis can be utilized to update the reliability of the system (joint PDF) with incomplete and uncertain information about a limited number of parameters (Schneider 2020). An example of a simplified bridge with two spans is provided in Figure 4. In this case only load and spatial variation in the longitudinal direction is considered (and not over the width of the bridge). The structural schematization with a distributed load indicates three common design checks: bending moment at midspan (blue), support moment (green) and shear force near the support (orange). The corresponding cross-sections are indicated in the lower part of the figure. Because of spatially varying material properties and execution details other cross-sections may be critical. In Figure 4 these cross-sections have been drawn with same color, but transparently.

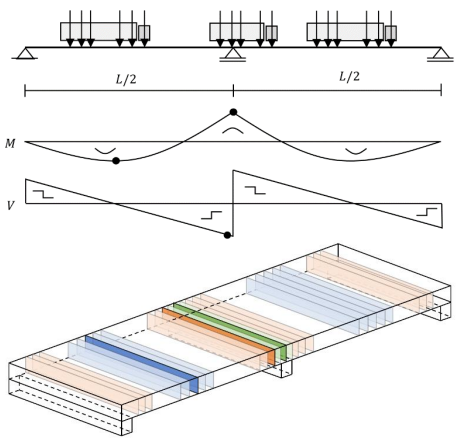


Figure 4: Visualization of the cross-sections to be assessed in a system-level assessment.

4 EXAMPLE TIME-DEPENDENT ANALYSIS

4.1 *Description*

To illustrate the effect of proof load testing in a time-dependent reliability analysis an example calculation is presented (De Vries, Lantsoght & Steenbergen 2021). Under consideration is a simply supported slab bridge located in one of the highways in the Netherlands. The structure was built in 1960 and designed according to the prevailing standards of that time (KIVI 1938, 1950). The traffic load used in its dated design is inappropriate when compared to today's high traffic intensity. But, the design values of material properties (e.g. steel and concrete strength) were quite conservative. As a result, old bridges and viaducts can still possess adequate structural strength to resist today's higher loads.

In case the original bridge documentation such as drawings and calculations are still available, they may be used to infer the (prior) probabilistic description of the resistance of the structure. In this case the bridge documentation is not available. Therefore its design was 'reverse engineered' by using historic standards (Harrewijn, Vergoossen & Lantsoght 2021).

In this example only the bending moment at midspan will be considered. In reality, the shear capacity of the slab near the supports and the capacity of other bridge components will require assessment as well.

4.2 *Conditional annual reliability*

The annual reliability is calculated under the condition that no failure occurs in any of the years before the year under consideration. Using the following events:

A failure in the year *i*;
B failure in the years 1 to *i* – 1;
B′ no failure in the years 1 to *i* – 1 (complement).

the conditional annual probability of failure can be written as:

$$\mathrm{P}(A \mid B') = \frac{\mathrm{P}(A \cap B')}{\mathrm{P}(B')} = \frac{\mathrm{P}(A \cup B) - \mathrm{P}(B)}{1 - \mathrm{P}(B)} \tag{3}$$

The probability $\mathrm{P}(A \cup B)$ may be read as the cumulative failure probability up to and including the year *i*, whereas P(*B*) is the cumulative failure probability up to, but not including, the year *i*.

To calculate the conditional annual reliability using the system reliability method, first the reliability index and influence coefficients of each year need to be calculated, e.g. using FORM. The individual years are the system components in this calculation. Next, the cumulative probability of failure can be calculated using the equivalent planes method (OR-combination). Then, the conditional probability of failure in year *i* is:

$$P_{\mathrm{f,cond},i} = \frac{P_{\mathrm{f},i} - P_{\mathrm{f},i-1}}{1 - P_{\mathrm{f},i-1}} \tag{4}$$

where $P_{\mathrm{f},i}$ is the cumulative failure probability up to and including the year *i*. In the first year no conditionality holds and thus $P_{\mathrm{f,cond},1} = P_{\mathrm{f},1}$.

Table 1: Random variables used in the limit state function.

Var.	Description	Dist.*	Mean	COV
θ_R	Model uncertainty of the resistance	LN	1	0.05
R	Bending moment resistance at midspan	LN	4100 kNm	0.05
θ_E	Model uncertainty of the load effect	LN	1	0.11
G_{DL}	Load effect of the dead load	N	721 kNm	0.05
G_{SDL}	Load effect of the superimposed dead load	N	101 kNm	0.1
C_{0Q}	Time-independent uncertainty of the variable load, including bias for dynamic load effect	LN	1.1	0.1
Q	Load effect of the annual traffic load	G	1150 kNm	0.025
t_{R0}	Initiation time to deterioration	LN	20 yr	0.1
Δ_{cR}	Degradation per year	LN	0.0025	0.1
c_{Q0}	Starting value of the trend	LN	0.78	0.1
Δ_{cQ}	Increase of traffic load per year	LN	0.004	0.1
Q_{PL}	Load effect of the proof load, including uncertainty	N	(varies)	0.05

*Distribution LN is lognormal, N is normal and G is Gumbel.

4.3 *Probabilistic model*

The limit state function for the probabilistic calculation is formulated in line with fib (2016):

$$Z = \theta_R c_R R - \theta_E (G_{\mathrm{DL}} + G_{\mathrm{SDL}} + c_Q C_{0Q} Q) \tag{5}$$

where the properties of the random variables are provided in Table 1. Each variable is characterized by a distribution, the mean value and coefficient of variation (COV).

To include the deterioration of the resistance and a trend in the traffic load, the limit state function makes use of the following time-dependent coefficients:

$$c_R(t) = \begin{cases} 1 & t \le t_{R0} \\ 1 - \Delta c_R (t - t_{R0}) & t > t_{R0} \end{cases} \tag{6a}$$

$$c_Q(t) = c_{Q0} + \Delta c_Q t \tag{6b}$$

where the parameters are random variables, listed as well in Table 1. This degradation model includes a time to initiation (t_{R0}), followed by a linear reduction of strength (Enright & Frangopol 1998). Corrosion leading to a reduction of the effective steel area in a cross-section was modelled by a quadratic function in Vu & Stewart (2000). In case of deterioration, a large degree of uncertainty exists with respect to the current capacity of the bridge. In this example only a limited amount of uncertainty is considered for simplicity. It thus represents the rather uncommon scenario where the deterioration process is well-known.

The system reliability method (Section 2.2.2) is applied to combine the failure probability in time. All random variables are fully auto-correlated except the annual traffic load which is assumed to be independent (i.e. a new realization each year).

WIM data from 2015 was analyzed to determine the load effect on the bridge, expressed as the largest bending moment at midspan within a certain period of time. Only the traffic in the right-most lane, where the trucks drive, has been analyzed.

4.4 *Results*

Using the presented probabilistic description, a time-dependent reliability analysis can be performed. The result is displayed in Figure 5. The base case displays the reliability without traffic trend and degradation. In this case the annual reliability increases gradually due to proven strength of past traffic loads. The traffic trend and degradation are incorporated subsequently to display their detrimental effect on the evolution of the annual reliability. A higher reliability is attained in the first years when including the traffic trend because the adopted linear trend expresses a reduction before the year 2015 and an increase afterwards.

Note that in this example the parameters of the degradation and traffic load trend have been tuned to

A

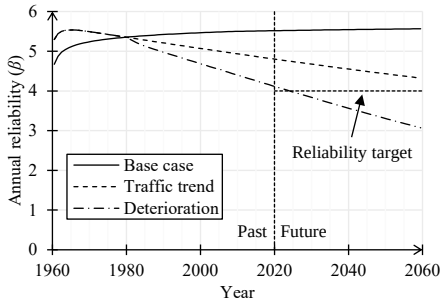


Figure 5: Development of the conditional annual reliability with time, incorporating a traffic load trend and deterioration.

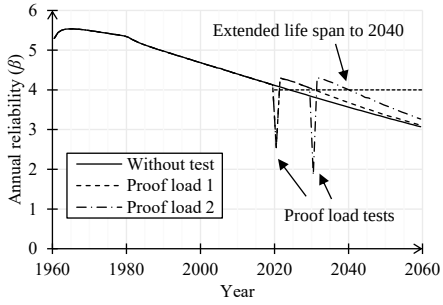


Figure 6: Effect of proof load testing on the annual reliability.

yield a reliability index that drops below the acceptable annual reliability $\beta = 4$ for CC3 (Steenbergen & Vrouwenvelder 2010) around 2020. In a real-life situation the parameters will need to be determined by studying the effect of all possible degradation mechanisms and the actual trend in traffic loads.

Proof load testing is adopted to ensure the bridge meets the required structural reliability. When a proof load test is performed an additional term is included for the proof load effect in the limit state function:

$$Z = \theta_R c_R R - \theta_E [G_{\mathrm{DL}} + G_{\mathrm{SDL}} + \max(c_Q C_{0Q} Q, Q_{\mathrm{PL}})] \quad (7)$$

The first proof load test is performed in the year 2020 and has a target (mean) value of 1800 kNm. Then, in the year 2030 the second proof load test is performed with a higher target load effect of 2000 kNm (Figure 6). In the year the test is performed, the annual reliability is markedly lower, but as a reward the reliability in the following years is higher. The target loads have been determined such that the annual reliability remains above the target in the next 10 years. Alternatively, the higher target load could have been applied directly in 2020, also leading to sufficient reliability until 2040. But, then the probability of failure in the first test in 2020 would be larger.

5 DISCUSSION

In this article, a hypothetical slab bridge was analyzed to show the development of the reliability with time. With regard to the knowledge level, in this example a scenario was depicted where the structural properties of the bridge, the traffic trend and deterioration process are known to large degree. Normally, this will not be the case. Especially the rate by which deterioration occurs will be difficult to establish. Suitable treatment of these uncertainties is critical.

The mathematical form of the limit state function is not definitive, other formulations are possible as well. In the presented formulation the model uncertainty of the load effect (θ_E) acts on both the traffic load effect and the load effect produced during the proof load test. In practice, the methods to calculate both effects will likely be similar (e.g. finite element analysis), but it does not guarantee full correlation.

In addition, it was investigated to what extent the historic traffic load influences the reliability at the moment of proof load testing. If the traffic load before the year 2015 (date of measurements) is ignored, comparable outcomes are produced. In this way insensitivity to the (difficulty to estimate) historic traffic load was established.

A future framework for proof load testing should be flexible, in such way that it could also consider the method by which a proof load test is performed. For example, when driving over a bridge less spatial uncertainty remains than when a single position is loaded.

6 CONCLUSIONS

With the suggested approach to the reliability assessment of existing reinforced structures through proof load testing a new framework can be developed that addresses existing challenges. An assessment based on annual reliability highlights the evolution of the structural reliability before, during and after the proof load test. A clear need for stop-criteria and an evaluation of their reliability, emerges from the relatively large probability of failure during the proof load test.

By adopting a flexible method, various types of information can be combined to assess the structural reliability through proof load testing. How much benefit is obtained when considering various kinds of information should be quantified in future research.

By judging the reliability on the system-level, uncertainty with regard to multiple failure mechanisms and spatial variability can be addressed. In this way, reservations regarding the assessment of shear capacity through proof load testing may be lifted.

ACKNOWLEDGEMENT

The authors wish to express their gratitude and sincere appreciation to the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat) for financing the research work. In addition, the fruitful discussions with M. Naaktgeboren of Rijkswaterstaat and M.A.N. Hendriks of Delft University of Technology have been of great help.

REFERENCES

AASHTO. (2018). The manual for bridge evaluation. *Standard, 3rd Edition. Washington, D.C., USA*.

ACI. (2013). Code requirements for load testing of existing concrete structures. *Standard, ACI 437.2M, Farmington Hills, Michigan, USA*.

Casas, J.R. & J.D. Gómez. (2013). Load rating of highway bridges by proof-loading. *KSCE Journal of Civil Engineering* 17 (3):556-567.

CEN. (2019). Eurocode 0: Basis of structural design. *Standard, EN 1990+A1+A1/C2:2019*.

DAfStb. (2020). DAfStb-Richtlinie: Belastungsversuche an Betonbauwerken. *Guideline, July*.

De Vries, R., E.O.L. Lantsoght & R.D.J.M. Steenbergen. (2021). Case study proof loading in an annual reliability framework. *Delft University of Technology, Stevin Report 25.5-21-02*.

Ditlevsen, O. & A. Vrouwenvelder. (1994). "Objective" low informative priors for Bayesian inference from totally censored Gaussian data. *Structural Safety* 16:175-188.

Ellingwood, B.R. (1996). Reliability-based condition assessment and LRFD for existing structures. *Structural Safety* 18 (2):67-80.

Enright, M.P. & D.M. Frangopol. (1998). Probabilistic analysis of resistance degradation of reinforced concrete bridge beams under corrosion. *Engineering Structures* 20 (11):960-971.

Faber, M.H., D.V. Val & M.G. Stewart. (2000). Proof load testing for bridge assessment and upgrading. *Engineering Structures* 22:1677-1689.

fib. (2016). Partial factor methods for existing concrete structures. *Fédération internationale du béton, Bulletin 80, Recommendation, Task Group 3.1*.

Frangopol, D.M., D.Y. Yang, E.O.L. Lantsoght & R.D.J.M. Steenbergen. (2019). Reliability-based analysis and lifecycle management of load tests. *Book chapter, Load testing of bridges: Proof load testing and the future of load testing, Editor E.O.L. Lantsoght*.

Fussel, J.B. (1975). A Review of fault tree analysis with emphasis on limitations. *IFAC Proceedings Volumes* 8 (1):552-557.

Gollwitzer, S. & R. Rackwitz. (1983). Equivalent components in first-order system reliability. *Reliability Engineering* 5:99-115.

Gong, C. & W. Zhou. (2017). Improvement of equivalent component approach for reliability analyses of series systems. *Structural Safety* 68:65-72.

Harrewijn, T.L., R.P.H. Vergoossen & E.O.L. Lantsoght. (2021). Reverse engineering of existing reinforced concrete slab bridges. *IABSE Congress, Christchurch 2021: Resilient technologies for sustainable infrastructure*.

Hasofer, A.M. & N.C. Lind. (1974). Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division* 100 (1):111-121.

Hohenbichler, M. & R. Rackwitz. (1983). First-order concepts in system reliability. *Structural Safety* 1:177-188.

Jonkman, S.N., R.D.J.M. Steenbergen, O. Morales-Nápoles, A.C.W.M. Vrouwenvelder & J.K. Vrijling. (2015). Probabilistic design: Risk and reliability analysis in civil engineering. *Lecture notes CIE4130, Delft University of Technology*.

KIVI. (1938). Voorschrift voor het ontwerpen en voor het vervaardigen en opstellen van stalen bruggen. *Koninklijk Instituut van Ingenieurs, Standard, N 1008*.

KIVI. (1950). Gewapend-beton-voorschriften. *Koninklijk Instituut van Ingenieurs, standard, N 1009*.

Lantsoght, E.O.L. (2019). *Load testing of bridges: Proof load testing and the future of load testing*. Edited by D.M. Frangopol, *Structures and Infrastructures, Vol. 13*. Boca Raton, FL: CRC Press.

Lantsoght, E.O.L., C. van der Veen, A. de Boer & D.A. Hordijk. (2017). State-of-the-art on load testing of concrete bridges. *Engineering Structures* 150:231-241.

Lichtenstein, A.G. (1993). Bridge rating through nondestructive load testing. *Technical Report, NCHRP Project 12-28(13)A*.

Lin, T.S. & A.S. Nowak. (1984). Proof loading and structural reliability. *Reliability Engineering* 8:85-100.

Marx, S. (2019). Neufassung der DAfStb-Richtlinie: Belastungsversuche an Betonbauwerken. *Presentation, 7. Jahrestagung, 60. Forschungskolloquium des DAfStb*.

Medha, K., J.W. Schmidt, J.D. Sørensen & S. Thöns. (2019). A decision theoretic approach towards planning of proof load tests. *13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP)*.

Metropolis, N. & S. Ulam. (1949). The Monte Carlo method. *Journal of the American Statistical Association* 44 (247):335-341.

Nishijima, K. & M.H. Faber. (2007). Bayesian approach to proof loading of quasi-identical multi-components structural systems. *Civil Engineering and Environmental Systems* 24 (2):111-121.

Roscoe, K., F. Diermanse & T. Vrouwenvelder. (2015). System reliability with correlated components: Accuracy of the equivalent planes method. *Structural Safety* 57:53-64.

Schacht, G., G. Bolle & S. Marx. (2019). Load testing of concrete building constructions. *Book chapter, Load testing of bridges: Proof load testing and the future of load testing, Editor E.O.L. Lantsoght*.

Schmidt, J.W., S. Thöns, M. Kapoor, C.O. Christensen, S. Engelund & J.D. Sørensen. (2020). Challenges related to probabilistic decision analysis for bridge testing and reclassification. *Frontiers in Built Environment* 6.

Schneider, R. (2020). Time-variant reliability of deteriorating structural systems conditional on inspection and monitoring data. *PhD thesis, Technical University of Munich, Bundesanstalt für Materialforschung und -prüfung (BAM)*.

Spaethe, G. (1994). Die Beeinflussung der Sicherheit eines Tragwerks durch Probebelastung. *Bauingenieur* 69:459-468.

Steenbergen, R.D.J.M. & A.C.W.M. Vrouwenvelder. (2010). Safety philosophy for existing structures and partial factors for traffic loads on bridges. *Heron* 55 (2):123-139.

Vu, K.A.T. & M.G. Stewart. (2000). Structural reliability of concrete bridges including improved chloride-induced corrosion models. *Structural Safety* 22:313-333.

Yuefei, L., L. Dagang & F. Xueping. (2014). Reliability updating and prediction of bridge structures based on proof loads and monitored data. *Construction and Building Materials* 66:795-804.

Zarate Garnica, G.I. & E.O.L. Lantsoght. (2021). Stop criteria for proof load testing of reinforced concrete structures. *Proceedings of the 2021 session of the 13th fib International PhD Symposium in Civil Engineering*:195-202.

Zhang, W-H., D-G. Lu, J. Qin, S. Thöns & M.H. Faber. (2021). Value of information analysis in civil and infrastructure engineering: A review. *Journal of Infrastructure Preservation and Resilience* 2.

B

Transportation Research Record article

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M. & Naaktgeboren, M. (2023a). Proof load testing method by the American association of state highway and transportation officials and suggestions for improvement. *Transportation Research Record*, *2677*(11), 245–257. https://doi.org/10.1177/03611981231165026

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

B

Research Article

Proof Load Testing Method by the American Association of State Highway and Transportation Officials and Suggestions for Improvement

Transportation Research Record
2023, Vol. 2677(11) 245–257

DOI: 10.1177/03611981231165026
journals.sagepub.com/home/trr
SAGE

Rein de Vries[1,2], Eva O. L. Lantsoght[1,3], Raphaël D. J. M. Steenbergen[2,4], and Marius Naaktgeboren[5]

Abstract

Because of the aging of infrastructure, methods are explored by which the reliability of existing bridges and viaducts can be assessed. In cases in which limited information of the structure is available or its condition is of concern, proof load testing may be used to demonstrate sufficient live load carrying capacity. Proof load tests in the U.S.A. are typically performed using the Manual for Bridge Evaluation (MBE) published by the American Association of State Highway and Transportation Officials (AASHTO). The proof load is expressed by the regular live load model magnified by the target proof load factor. The level of reliability obtained using the target proof load factor is not explicitly stated in the MBE, but is of particular interest. In this article, relevant background documents are investigated to uncover the underlying calculations, assumptions, and input data. Current challenges in proof load testing are described in which the considerations of time dependence, stop criteria, available information, and system-level assessment are highlighted. Subsequently, improvements to the MBE proof load testing background are suggested. An example calculation using traffic data from the Netherlands shows that the HL93 load model and Eurocode LM1 provide a reasonably constant proof load factor with span length for bending and shear. However, the HS20 load model does not scale well with increasing span length. It is found that the magnitude of the target load as specified through the proof load factor is directly related to the desired level of reliability. Although the MBE proof load testing method is practical, several challenges remain.

Keywords

infrastructure, structures, testing and evaluation of transportation structures, bridge assessment, load rating, load testing

Normally, a structure that has just been completed fulfills functional and safety requirements as specified by the prevailing design standards. However, after years of use, the environment or societal demands may have changed (for instance, larger traffic intensity or more stringent safety requirements). In addition, the structure may have suffered from degradation. To evaluate if the structure fulfils the requirements, an assessment needs to be carried out (*1*). Proof load testing is one of the methods available for assessment, competing with desk studies that often make use of finite element models.

Recent advances concern the usage of load test information to update finite element models and structural reliability estimations (*2*), and are collected in the Transportation Research Board (TRB) circular *Primer on Bridge Load Testing* (*3*). Proof load testing as a means to assess the structural reliability found its way into the literature in the 1980s (*4–6*). The probabilistic treatment

[1]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[2]Structural Reliability, Netherlands Organization for Applied Scientific Research (TNO), Delft, The Netherlands
[3]College of Sciences and Engineering, Universidad San Francisco de Quito, Quito, Ecuador
[4]Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium
[5]Ministry of Infrastructure and Water Management, Rijkswaterstaat, Utrecht, The Netherlands

Corresponding Author:
Rein de Vries, rein.devries@tudelft.nl

Figure 1. The German BELFA load testing vehicle on the Vlijmen-Oost viaduct in the Netherlands. Reprinted with permission from (*1*).

of proof load testing can result in appropriate target loads depending on the load rating, dead/live load ratios, degradation, bridge age, reference period, and prior service loads (*7*). In Europe, proof load factors were developed as part of the large-scale ARCHES (Assessment and Rehabilitation of Central European Highway Structures) project (*8*). Recently, efficient strategies for bridge reclassification based on probabilistic decision analysis have gained attention (*9–11*).

In a proof load test, a relatively large load is applied to a bridge or viaduct to demonstrate sufficient load-carrying capacity. If the structure is able to withstand the large load without showing signs of distress, the test is a success. The load can be applied by using one or more heavy trucks (as is common in the U.S.A.), a loading frame with ballast, or a specialized load testing vehicle (Figure 1). The magnitude of the load to be applied in the proof load test is commonly referred to as the *target load*. If the target load is chosen to be very large (e.g., multiple times a heavy truck weight), the probability of failure during the test is also large. However, if the bridge can withstand the large load without showing signs of distress, then it has proven to have a high reliability. If a relatively small target load is selected (e.g., a normal truck weight), the probability of failure during the test is small, but it also does not prove that the bridge has high reliability. Therefore, the target load is directly associated with the structural reliability (and safety) of the bridge or viaduct being tested. If the target load could not be reached during the proof load test, because signs of distress were detected, then load posting (load restrictions) may be applied, or the bridge needs to be renovated/replaced. Such decisions depend on the load level reached during the test and the nature of the observed distress.

In the case in which a proof load test is performed in the U.S.A., the Manual for Bridge Evaluation (MBE) (*12*) is used as a guideline. The MBE describes many aspects concerning the inspection and evaluation of existing highway bridges. In this context, proof loading is mentioned as one of the methods in which a *load rating* may be accomplished. Load rating is the determination of the live load carrying capacity of a structure. The MBE provides guidance on how to use diagnostic and proof load tests for the purpose of load rating. This article focusses on the latter, and in particular the reliability background described in a report dating from 1993 by Lichtenstein (*13*), which was also included in the 1998 Manual for Bridge Rating Through Load Testing (*14*).

Objective

The objective of this article is to describe the proof load testing method by the American Association of State Highway and Transportation Officials (AASHTO) as provided in the MBE, communicate the background to the target proof load factor (X_p), provide suggestions for improvement, and highlight current challenges in proof load testing—all within the context of structural reliability.

The level of reliability that the MBE proof load testing method provides is of interest in relation to possible application in other countries where different reliability requirements and traffic conditions apply. To aid the knowledge transfer between countries, a significant part of this article is devoted to explaining the MBE proof load testing method and its probabilistic background. The novelty of this research is contained in the improvements suggested to the MBE reliability background and in the identification of the remaining challenges in proof load testing.

Reliability Assessment of Existing Structures

Reliability expresses the probability of success (no failure) within a certain time period, that is, the reference period. In the design of new structures, the reference period is the design life. In the context of annual reliability the reference period is 1 year. Reliability is commonly related to the failure probability via $P_f = \Phi(-\beta)$, where β is the reliability index and $\Phi(\cdot)$ indicates the standard normal cumulative distribution function. In the reliability assessment of structures, failure indicates the exceedance of a limit state. Commonly, such a failure comprises loss of capacity or another significant change that leads to an unsafe situation (*15*). Reliability assessment of existing structures may be done via various methods, including proof load testing (*7*).

The reliability requirements, or targets, for existing structures are commonly lower than the requirements for new structures. For new structures, on top of safety and sociopolitical considerations, additional economy or quality requirements typically govern. In the U.S.A., a reliability index of $\beta = 3.5$ with a reference period of 75 years is common for new structures, whereas a reliability index of $\beta = 2.3$–2.5 with a reference period of 2–5 years is common for existing structures (*16*, *17*). In the U.S.A., a reliability requirement is connected to a failure mode, load combination, and type of uncertainty (*18*). Alternatively, a differentiation can be made based on the consequence class (CC) for the structure, as adopted in the Eurocode standards (*19*). CC2 is typically associated with bridges in the secondary road network, whilst CC3 applies to structures in the primary (highway) network. The assessment criteria for existing structures in the Netherlands have been determined by a risk-based approach. The minimum allowable annual reliability indices, associated with human safety, are 3.4 for CC2 and 4.0 for CC3 (*20*).

Reliability Background of the Manual for Bridge Evaluation Proof Load Testing Method

For a full description of the proof load testing method, the reader is referred to Section 8.8.3 of the MBE (*12*). This article is mainly concerned with the calculation of the target load and its relation to structural reliability. Therefore, only the relevant parts of the method are described here.

Rating Factor

The MBE uses a so-called rating factor to indicate to which degree a structure is able to carry live loads. In principle every bridge is thought to have a resistance (or capacity) R, dead load (or self-weight) D, and a live load (or variable load) L. A structure is considered safe if the resistance is equal to, or larger than, the self-weight plus the live load. The rating factor (RF) is derived as follows:

$$\begin{aligned} R &\geqslant D + L \\ R - D &\geqslant L \\ \frac{R-D}{L} &\geqslant \frac{L}{L} \\ RF = \frac{R-D}{L} &\geqslant 1 \end{aligned} \tag{1}$$

In Equation 6A.4.2.1–1 of the MBE additional factors are included, various types of dead load are differentiated, and the dynamic contribution to the live load (dynamic load allowance) is included explicitly.

Table 1. Adjustment to Target Proof Load. Reproduced with permission from (*12*).

Consideration	Adjustment
One-lane load controls	+ 15%
Nonredundant structure	+ 10%
Fracture-critical details present	+ 10%
Bridges in poor condition	+ 10%
In-depth inspection performed	−5%
Ratable, existing $RF \geqslant 1$	−5%
$ADTT \leqslant 1000$	−10%
$ADTT \leqslant 100$	−15%

Target Proof Load

Bridge-specific circumstances may be included via the adjusted proof load factor (X_{pA}). The proof load factor is increased or decreased by an associated percentage (see Table 1, which is Table 8.8.3.3.1–1 of the MBE). In cases in which multiple considerations apply, the adjustment percentages are summed.

The value of the adjusted proof load factor is calculated via $X_{pA} = X_p(1 + \Sigma\,\% / 100)$. If a one-lane load controls the load effect (i.e., no redistribution of forces between lanes exists), then an increase of 15% is required. The motivation for this adjustment is the 0.85 factor found in comparing two-lane traffic to one-lane traffic loads (*13*). The target proof load (L_T) is expressed with respect to the load model and is magnified by an (adjusted) proof load factor, leading to the following expression:

$$L_T = X_{pA} L_R (1 + IM) \tag{2}$$

where L_R is the comparable unfactored live load because of the rating vehicle for the lanes loaded and IM is the dynamic load allowance (or impact). The background to the target proof load factor (X_p) may be found in the 1998 Manual for Bridge Rating Through Load Testing (*14*), which references and attaches a report by Lichtenstein (*13*). In Chapter 3 of that technical report, the default value $X_p = 1.4$ is derived from a probabilistic analysis. The remainder of this section describes the analysis that was performed in the Lichtenstein report (*13*) and reproduces the results.

Considered Case

In Lichtenstein (*13*), a simply supported bridge with span of $l = 60$ ft (18.3 m) is considered as a base case. It has two driving lanes and its capacity is unknown. The AASHTO HS20 load effect is calculated by gradually moving the HS20 truck along the beam whilst keeping track of the maximum moment occurring at each location. The HS20 truck is defined by a front axle loaded with 8 kip (35.6 kN), followed at 14 ft (4.27 m) by an axle

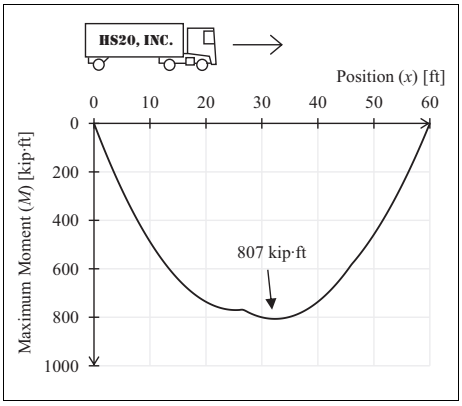


Figure 2. Moment envelope of a 60 ft (18.3 m) simply supported beam subjected to the HS20 load. Conversion factors: 1 ft = 0.305 m, 1 kip·ft = 1.35 kNm.

loaded with 32 kip (142.3 kN) and then by another axle of 32 kip. The distance between the heavily loaded axles is 14–30 ft (4.27–9.14 m), whichever produces the largest load effect—in this case it is 14 ft. The calculated moment envelope is plotted in Figure 2, which indicates a maximum of 807 kip·ft (1094 kNm) around the midspan.

Limit State Function

The limit state function adopted in the Lichtenstein report (*13*) for the probabilistic calculation is as follows:

$$Z = R - (D + L + I) \tag{3}$$

where R is the resistance of the structure, D is the dead load (permanent load), L is the live load, and I is the impact load (dynamic load effect).

In the case in which $Z < 0$, structural failure occurs. The resistance and dead load are regarded as deterministic values in the calculation. It is assumed that their values are known after the proof load test. The value of the dead load effect is taken to be equal to the AASHTO HS20 live load effect, that is, $D = L_A = 807$ kip·ft per lane. The effect of different dead load contributions in relation to the live load is studied by Lichtenstein (*13*) as well. The live load (L) and impact load (I) are modeled as random variables.

Traffic Load

The mean value of the 75-year maximum traffic load is equal to 1.79 times the HS20 load effect (L_A) as determined by extrapolation of a traffic survey (*21*). For a reference period of two years, it is 1.65 times the HS20 load effect. The latter reference period is deemed to be appropriate in the context of proof load testing because it is based on an inspection interval of two years.

If two lanes are considered, a reduction factor of 0.85 applies because of expected redistribution of the load between lanes. The live load has a coefficient of variation of 0.18. This value is thought to cover both the uncertainty in the occurrence of heavy truck loads and the analysis. If only the truck load uncertainty is considered, a reduced value of 0.14 applies. The dynamic load allowance is different than that defined by the coefficient of AASHTO. Its mean value is estimated to be about 0.1 of the live load, with a coefficient of variation of 0.8. The mean value of the live load for a return period of 2 years is thus calculated as follows:

$$m_L = 0.85 \cdot 1.65 \cdot 807 = 1132 \text{ kip} \cdot \text{ft} \ (1535 \text{ kNm})$$

Only the uncertainty in the occurrence of heavy truck loads is deemed to apply, and therefore the coefficient of variation $V_L = 0.14$ is used. The mean value of the impact load is calculated as follows:

$$m_I = 0.1 \cdot 1132 = 113.2 \text{ kip} \cdot \text{ft} \ (153.5 \text{ kNm})$$

with coefficient of variation $V_I = 0.8$.

Resistance

If the proof load test was successful, the resistance of the structure is at least equal to the sum of the target proof load (L_T) and the dead load. In the probabilistic analysis the mean resistance of the structure is calculated as follows:

$$R = 1.12(L_T + D) = 1.12\left[X_p L_A (1 + C_{I,A}) + D\right] \tag{4}$$

where the factor 1.12 accounts for higher mean strengths in respect to the nominal (or "design") strengths used in regular code calculations. The value of the AASHTO impact coefficient for 60 ft (18.2 m) is $C_{I,A} = 50/(60 + 125) = 0.27$. The resistance for this case is calculated as follows:

$$R = 1.12\left[X_p \cdot 807(1 + 0.27) + 807\right] = 1148 X_p + 904 \text{ kip} \cdot \text{ft}$$

All information required to perform the probabilistic analysis has now been described. Table 2 provides an overview of the parameters used in the limit state function.

Table 2. Overview of Variables in the Limit State Function

Var.	Description	Dist.[a]	Mean (*m*) [kip·ft]	CoV (*V*) [-]
R	Resistance	na	$1148X_p + 904$	na
D	Dead load effect	na	807	na
L	Live load effect	N	1132	0.14
I	Impact load effect	N	113.2	0.8

[a]Distribution type. N = normal; LN = lognormal; G = Gumbel.
CoV = coefficient of variation, na = not applicable.

Table 3. Calculated Reliability with Increasing Proof Load Factor

	Reliability index for 2 years (β) [-]	
Proof load factor (X_p) [-]	Lichtenstein	Recalculated
1.2	1.26	1.26
1.3	1.89	1.89
1.4	2.57	2.52
1.5	3.15	3.15
1.6	3.78	3.77

Results

Since all distributions are assumed to be normal, the calculation method in Lichtenstein (*13*) simply consists of dividing the mean value of the limit state function Z by its standard deviation. Making use of the first-order reliability method (FORM) (*22*) is also possible. In this case, the FORM converges to the exact solution in one iteration. The reliability index (β) that results from the probabilistic calculation for various values of X_p is provided in Table 3. The values in the column *Recalculated* have been obtained by the author and are in correspondence with the original numbers.

The value $X_p = 1.4$ was selected because a reliability index of $\beta = 2.3$ was found to be in line with the operating level according to the AASHTO Load and Resistance Factor Design (LRFD) studies (*16*). The reliability index is lower than that used for the inventory and design level. The lower value is justified, since it reflects past rating practices at the operating level. In relation to varying span lengths, two additional calculations are performed by Lichtenstein (*13*). It is concluded that the factor $X_p = 1.4$ provides adequate safety for other spans as well.

Suggested Improvements to the Manual for Bridge Evaluation Proof Load Testing Background

In this section various improvements are suggested to be incorporated into the background of the MBE proof load testing method. It does, however, not result in a ready-to-be-adopted new format. Instead, the most important facets are highlighted and improvements are suggested.

Probabilistic Model

In the Lichtenstein report (*13*), the dead load is treated as a deterministic value equal to the live load. The value of the dead load (effect) is not exactly known, as it is generally not possible to measure its value. However, because the structure can carry the dead load, it does not matter if it is lower or higher than expected. When including the dead load as a random variable, it can also be eliminated from the limit state function, as shown in Equation 5.

An additional factor of 1.12 is used in Equation 4 to convert from nominal to mean strength. Such a factor is appropriate when R is a design or nominal strength. However, here R is a random variable. After a successful proof load test, it is known that the resistance must be equal to or larger than the load effect following from the self-weight and the target load ($R \geq D + L_T$). Assigning the resistance with a value that is 12% higher than that obtained from the test is speculative.

With the suggested alterations, the limit state function may be rewritten such that only the live load and the dynamic amplification remain as random variables. In essence, the probability of failure of the structural part or cross-section is directly reformulated into the probability that a future live load effect (including dynamic amplification) exceeds the load effect produced during the proof load test:

$$\begin{aligned} Z &= R - (D + L + I) \\ &= (D + L_T) - (D + L + I) \\ &= L_T - (L + I) \end{aligned}$$

Here, L_T is the target proof load effect (deterministic value), L is the traffic live load effect (random variable), and I is the dynamic contribution (random variable).

The dynamic load effect (impact) should be included in the target load (L_T) as part of the load model via the regular design procedure. Since comparable extreme values for the traffic load are considered, the design

Table 4. Overview of Variables Included in the Suggested Limit State Function

Var.	Description	Dist.[a]	Mean (m) [kip·ft]	CoV (V) [-]
θ_{LT}	Model uncertainty load effect produced in the test	LN	1	0.1
L_T	Load effect caused by proof loading vehicle or frame	na	(varies)	na
θ_L	Model uncertainty load effect produced by the traffic load	LN	1	0.11
C_{0L}	Time-invariant part of the traffic load variability	LN	1	0.1
L	Annual maximum of the traffic load effect	G	(varies)	(varies)

[a]Distribution type. N = normal; LN = lognormal; G = Gumbel.
CoV = coefficient of variation, na = not applicable.

procedure to account for the dynamic loads is suitable here as well. Therefore, the impact (I) may be removed from the limit state function. In this way, the probabilistic analysis can be performed using recorded traffic loads without, or with minimal, dynamic contribution (e.g., weight-in-motion [WIM] data).

Missing in the limit state function of Equation 5 are model uncertainties. Our understanding of the translation from applied loads, in a test or from actual traffic, to the load effect is limited. The degree of uncertainty depends on the level of sophistication incorporated in the mechanical model. Additional uncertainty stems from the statistical modeling of the load effect—that is, the assumed distribution functions. The variability of the traffic load may be split into time-invariant (C_{0L}) and time-variant parts (L) (*23*). By including model uncertainties, splitting the live load variability, and removing the dynamic contribution, the limit state function becomes as follows:

$$Z = \theta_{LT} L_T - \theta_L C_{0L} L \tag{6}$$

An overview and description of the parameters in the suggested limit state function is provided in Table 4. The statistical properties of the random variables are based on general recommendations for probabilistic modeling (*23*, *24*). The coefficient of variation of the model uncertainty concerning the load effect produced in the proof load test (θ_{LT}) is based on the value of the model uncertainty related to the traffic load. Because the conditions are more controlled during a test, a lower value may seem more appropriate. However, when viewed as a resistance parameter, it should also cover the uncertainty associated with selecting the most critical locations to test. This issue is alleviated when using a moving vehicle to perform the test, but also here some uncertainty remains related to the transverse location and axle configuration.

The mean value of the model uncertainty relating to the traffic load effect (θ_L) can be altered to introduce a certain bias. In this way, a traffic load model derived on the basis of roads with great intensity may be corrected with a factor lower than 1 to reflect the general case. On the other hand, the trend of continuously increasing traffic loads would increase the mean value. The latter depends largely on the time period that is considered. The trend will also depend on the span length (or influence length) because of longer trucks and smaller intervehicle distances, for example, truck platooning (*25*). This affects large spans more than short spans.

Traffic Load

In the Lichtenstein report (*13*), the statistical description of the live load and impact (dynamic) load is based on the study by Nowak from 1993 (*21*). It is recommended to use more recent data, preferably obtained from the measurement of axle loads at multiple locations and for a longer period of time (e.g., one year or more). WIM data is well-suited to obtain an accurate statistical representation of the traffic load effect.

In the Netherlands, WIM recording stations are positioned at several traffic-intense highway locations. Using WIM data from 2015 in the Netherlands, traffic simulations have been performed to obtain the maximum bending moment at the midspan and the maximum shear force near the supports of a simply supported span. Over a period of one year, various block maxima may be determined: hourly, daily, weekly, and so on. Considering the difference in traffic on weekdays and at the weekend, a week represents an appropriate cycle, leading to 52 data points. Subsequently, the Gumbel extreme value distribution is fitted to the data using the maximum likelihood estimation (MLE) method. A threshold value is chosen (probability of exceedance $S = 0.25$) to capture, on the log-scale, the linearly descending right-hand tail of the distribution. Figure 3 shows the fitted distribution to the data points of the maximum bending moment of Dutch highway A27L lane 1—the rightmost lane mostly occupied by trucks.

The traffic load model uncertainty incorporated in the probabilistic model should reflect the quality of the data and its modeling. Ideally, data collected over a long

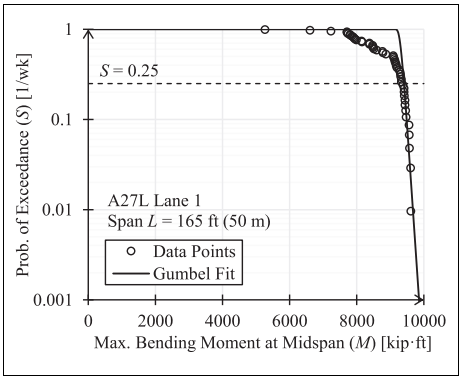


Figure 3. Gumbel fit of the load effect data points for the maximum bending moment at the midspan of a simply supported span. Conversion factors: 1 ft = 0.305 m, 1 kip·ft = 1.35 kNm.

period of time is used for distribution fitting. However, traffic data from many years ago may not be representative for the traffic today. Over longer periods of time, trends in the data may cause a distorted view if all data points are processed as if they would have originated from the same stationary process.

Because the weekly maxima are sufficiently uncorrelated, the Gumbel distribution may be converted to annual maxima by shifting the location parameter (μ) via $\mu_a = \mu_w + \beta_G \ln(52)$, where β_G is the scale parameter and 52 is the number of weeks in a year. The mean and coefficient of variation of the distribution are obtained as $m = \mu + \beta_G\gamma$ and $V = s/m = (\beta_G \pi/\sqrt{6})/m$, where $\gamma \approx 0.5772$ is the Euler–Mascheroni constant (*26*). Distributions have been fitted for various WIM datasets and span lengths, as displayed in Figure 4. The analyzed roads show a comparable trend in the mean and coefficient of variation with span length. In a reliability analysis, the average of the four different roads is used (plus model uncertainty θ_L; see Table 4).

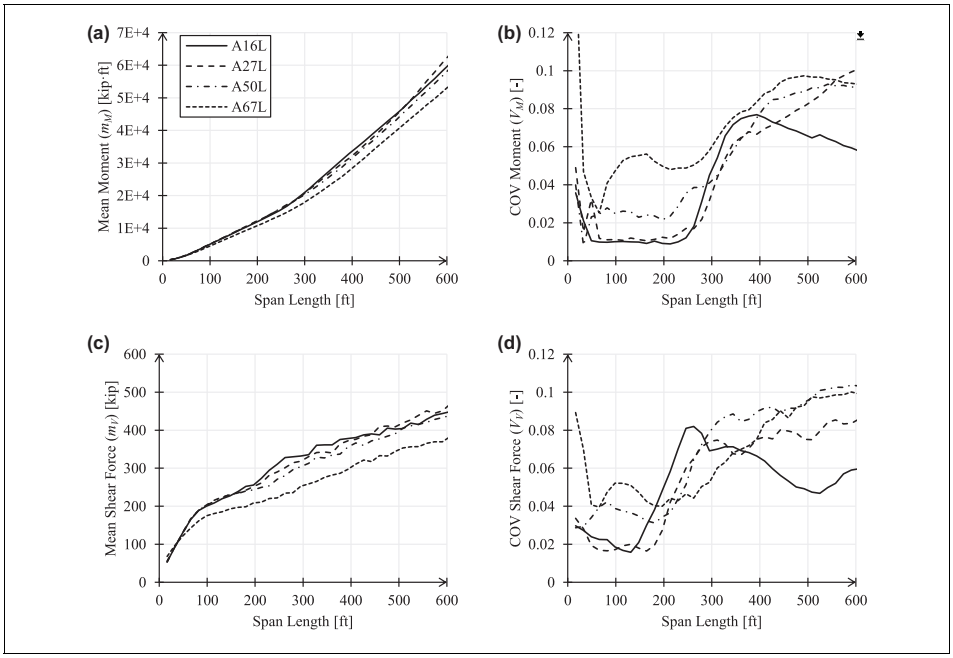


Figure 4. Parameters of the fitted Gumbel distribution for the annual maximum load effect: (*a*) mean of the bending moment, (*b*) coefficient of variation (COV) of the bending moment, (*c*) mean of shear force, and (*d*) coefficient of variation of shear force. Conversion factors: 1 ft = 0.305 m, 1 kip·ft = 1.35 kNm, 1 kip = 4.45 kN.

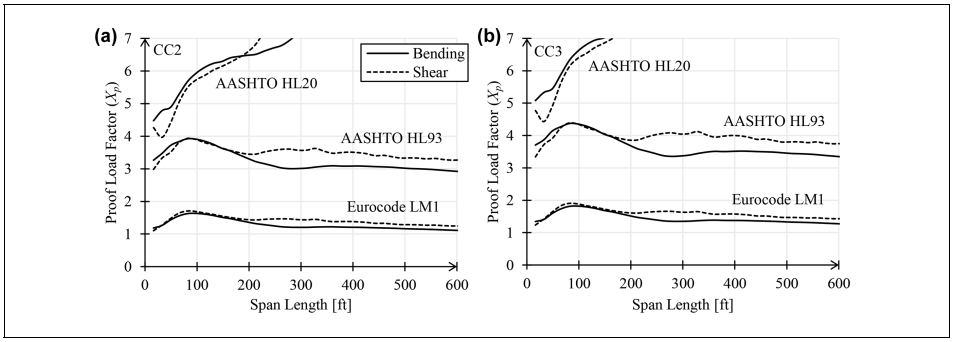


Figure 5. Relation between the span length and target proof load factor considering unfactored load models in the bending and shear: (*a*) consequence class (CC) 2 with annual reliability requirement β = 3.4 and (*b*) consequence class 3 with annual reliability requirement β = 4.0. Conversion factor: 1 ft = 0.305 m. AASHTO = American Association of State Highway and Transportation Officials.

Influence of Span Length

The configuration of a bridge that is subjected to a proof load test is often different than the simply supported span for which the load effect was calculated and the reliability analysis was performed. To overcome this limitation, the target proof load is related to a load model via the proof load factor; see Equation 2. In the Lichtenstein report (*13*) the HS20 load model is used, but today the HL93 load model (*16*) describes the traffic better. In addition to an HS20 truck or a (military) load tandem, the latter also includes a lane load (distributed load) that represents the other traffic present on the bridge. The HL93 load model is comparable to the Eurocode LM1 specification but has significantly lower loads. In Nowak et al. (*27*) it is found that the Eurocode LM1 load effects are about a factor two higher than those calculated using AASHTO HL93, owing to the higher unfactored loads in the traffic model. After applying (partial) factors, the design load effect varies from country to country.

To study the relation between the span length and the target proof load factor, an example calculation is made with the improved probabilistic model and traffic data from the Netherlands. Per the span length, two probabilistic analyses are performed: one considering the bending moment at the midspan and one considering the shear force near the supports. CC2 and CC3 of EN 1990 (*19*) are considered with target annual reliability indices of 3.4 and 4.0, respectively. The distributed load of the Eurocode LM1 and the AASHTO HL93 load models are applied over a lane width of 3 m. The proof load factors following from the reliability analyses are displayed in Figure 5.

It is observed that the target proof load factor is considerably larger when using the AASHTO HS20 and HL93 load models in comparison to Eurocode LM1. This follows from the relatively high unfactored load effect following from LM1. A quick comparison of the axle loads signifies the difference: 67.4 kip (300 kN) for LM1 versus 32 kip (145 kN) for HS20 and HL93, respectively. The traffic load from the Netherlands is relatively high compared to other countries (*28*) and meshes with the high loads of Eurocode LM1, leading to moderate values of the proof load factor (X_p). Because of the large discrepancy between unfactored load models, it is recommended to cautiously evaluate traffic models and statistical descriptions for application within the U.S.A.

Another observation is the continuously increasing proof load factor with span length when the HS20 load model is used. This is because the load model only includes a single truck, whereas in reality many vehicles may be present on the bridge. The issue is overcome by the HL93 load model, which also includes a distributed lane load. For both the Eurocode LM1 and the HL93 load models, an almost constant factor is obtained over various spans. Only around 100 ft (30 m) is a relatively large factor is required. This may be explained by the occurrence of long and heavy vehicles (an oversize load for which usually an exemption must be requested) that are not accurately represented by the load model.

Remaining Challenges in Proof Load Testing

In the more general context of proof load testing, several current challenges are highlighted in relation to structural reliability. Because the practice of proof load testing has been established in the past, where assessments had a predominantly deterministic character, various probabilistic aspects deserve extra attention. The aspects highlighted

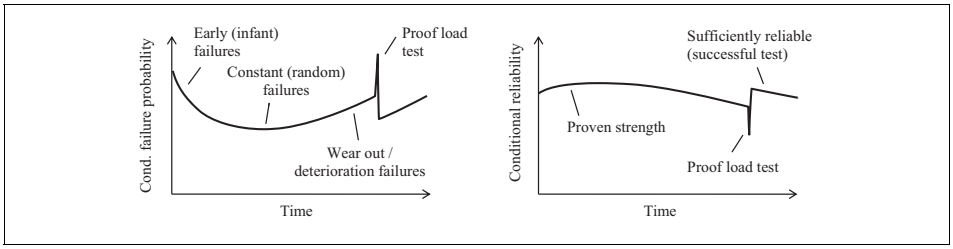


Figure 6. Evolution of conditional failure probability and (annual) reliability with time.

below could provide further advancement of the reliability background (and suggestions for improvement) presented in the previous sections.

Time Dependence

The structural reliability of a bridge or viaduct that is subject to time-variant loads (such as traffic loads), or other time-dependent processes, is not constant in time (*29*). It is intuitive to understand that the reliability of a bridge or viaduct that suffers from corrosion, or any other form of degradation, will slowly decrease with time. Less intuitive is the increase of reliability with time as the bridge continuously sustains a high traffic load, and thus displays *proven strength* with time. Both of these effects result in a variable conditional failure probability, often in the form of a *bathtub curve* (*30*). Conditional indicates the condition in which no failure has occurred in the past. By considering a sufficiently small period of time, the evolution of structural reliability may be followed. To this end, the annual reliability is considered in time-dependent probabilistic calculations (*31*). When a proof load test is performed, the failure probability is high because of the large load that is applied. Afterwards, in the case of a successful test, the structural reliability is higher (*32*). The bathtub curve may be extended to include a proof load test, as displayed in Figure 6.

When a proof load test is successful, a sufficient degree of structural reliability is demonstrated at that moment. However, because of the time-dependent effects (such as deterioration and traffic load trends) the reliability may not be sufficient in the future. Substantiated statements about the development of future reliability require probabilistic analyses in which time-dependent effects are explicitly considered (*33*). An appropriate increase of the target load (e.g., via the factor X_{pA}) may be quantified, depending on the bridge-specific circumstances. In this way, the expected time-dependent effects and their uncertainty can be compensated by a higher target load.

Stop Criteria

During a proof load test the load is gradually increased until the target load is reached. In this process, the structure may show signs of distress before the full target load is applied. Therefore, criteria for stopping the proof test are required. The use of sensors during the test provides extra information about the structural response. The stop criteria are typically related to the structural response (not directly the measurements). The sensor readings are interpreted with respect to stop criteria. If a stop criterion is exceeded in the proof load test, irreversible damage may occur. Depending on the damage and failure mechanism, different stop criteria apply (*34*, *35*).

Many bridges are constructed using reinforced concrete. Generic stop criteria are difficult to apply in practice for reinforced concrete structures where there may be existing cracks caused by material degradation (*3*). In addition, a distinction between linear and non-linear behavior is not useful as a stop criterion because even moderate loads can cause the exceedance of the concrete tensile strength. In some cases small cracks are acceptable, while in other cases they are not.

In the definition and application of stop criteria, little attention is paid to the link with structural reliability. By considering the reliability during the test, the risk of collapse can be mitigated and additional guidance provided for the selection and placement of sensors (*36*).

Information About the Structure and Its Context

The knowledge level (available information in drawings, calculations, material tests, etc.) may vary significantly between structures (*8*). For older structures it is common that documentation is missing and only basic information (year of construction, geometry, etc.) remains. Valuable information may also relate to the context of the structure. For example, a bridge may be part of the highway network or part of a larger group of infrastructure all designed to the same specifications.

Many information sources can be included to estimate the capacity of a structure. In addition, extra material tests may be performed to obtain a better estimate of the important parameters in the structural model. Ultimately, a balance must be sought between acquiring information about the structural performance and with respect to the proof load test as the main source of information (*9*). Through the application of Bayes' theorem, various sources of information (evidence) can be included to improve the estimate of the resistance. For this purpose, the uncertainty of the resistance is split into the objective (natural) and subjective (model) parts. In this way, the model uncertainty may be systematically updated in the Bayesian approach (*5*).

System-level Assessment

A bridge or viaduct consists of several components, such as the deck, girders, supports, and so forth. In addition, cross-sections and connections can also be regarded as components. In the calculation of the structural reliability of the entire structure, all components matter. In regular design approaches, the reliability is verified at the structural component or element level. Under the assumption of sufficient parallel performance (redundancy), correlation between the components or overstrength, the system reliability is approximately equal to the reliability at element level. However, this single component methodology may actually lead to both under- and overconservative designs (*37*). Explicitly considering the system behavior in the probabilistic assessment leads to more accurate estimates of structural reliability.

Checking only a single cross-section (e.g., the bending moment capacity at the midspan) in a proof load test is not always sufficient to verify the reliability of the bridge as a whole. In this regard, the MBE states that "loads must also be moved to different positions to properly check all load path components." In practice this is accomplished by using various heavy vehicles and driving paths.

Because of practical limitations or economic reasons, not all components may be verified explicitly in a proof load test. Bayesian analysis can be used to update the reliability of the system with information about a limited number of components (*38*). In this way, uncertainty may be compensated through application of a higher target load.

Discussion

Suggested Improvements

The suggested probabilistic model includes model uncertainties for both the actual live load (θ_L) and the live load produced in a proof load test (θ_{LT}). Their statistical description has been estimated and requires further refinement. In particular, the uncertainty associated with the proof load test will need to cover different aspects depending on the application: how the load is applied, how many positions and lanes are tested, whether the bending or shear are critical, and so forth. In addition, it is likely that the model uncertainties (θ_{LT} and θ_L) are correlated because the same mathematical principles/models are used to calculate both load effects. Because there are several remaining challenges, the probabilistic model and the results presented in this article should be viewed as indicative.

The traffic load analysis was performed using highway measurements obtained in the Netherlands; therefore, the resulting distributions have limited applicability. By using the method followed in this article, applicable distributions can be derived for different countries. For completeness, also other configurations besides the single span case need to be considered (*8*).

In the MBE an adjustment to the target live load of + 15% is suggested in cases in which a one-lane load controls the load effect. This is a measure to counteract the more favorable two-lane traffic load description used by Lichtenstein (*13*). An important assumption in the two-lane situation is that the bridge is able to redistribute the traffic load between its lanes. This is not always the case. In today's computer-aided design process, all lanes and their corresponding loads are included in the model. The load effect for additional lanes follows from the load model, just as in the design calculation. (The rationale is the same for excluding the dynamic effect in the probabilistic calculation; it is assumed that the established calculation rules account for these particularities correctly.) The use of a multiple presence factor (MPF) calibrated on the basis of WIM data is recommended (*39*). With the one-lane situation as the default, a probabilistic assessment of the first lane (as performed in this article) will correspond to MPF = 1.0.

Remaining Challenges

Proof load testing is a valuable tool to demonstrate sufficient load-carrying capacity. However, the derivation of load factors and rules to carry out a test that results in the desired reliability remains challenging.

The time dependence of the structural reliability can be incorporated into a probabilistic analysis directly to deliver the point in time where the annual reliability is not sufficient any longer. An example of such a calculation is provided by De Vries et al. (*33*, *40*). If such analyses are performed for several typical cases, the required increase of the target load can be determined.

A future framework for proof load testing should be flexible with respect to which information is utilized. In some cases, bridge documentation, material data, traffic

data, or even proof load testing data on similar bridges may be available. It would not be economical to ignore evidence of good structural performance. Additional rules may be established on the basis of Bayesian inference to utilize knowledge about the structure and its context (traffic loads, environment, geographical location, etc.) (*33*).

By thinking about the bridge as a system of components, the question is how many components are tested in a proof load test. This may also depend on the type of bridge or failure mechanism being considered. In a (successful) proof load test, one can only observe that the system (i.e., the entire structure) carries the load. However, the load may not follow the expected load path and/or redistribution of forces can take place. For this reason, the proof load test result does not necessarily tell us something about the performance of a component.

Conclusion

The magnitude of the target load is directly related to the desired level of reliability. In the MBE (*12*), the target load is obtained through application of the live load model multiplied by a factor for proof load testing that can also include bridge-specific adjustments (X_{pA}). In this way, the target proof load can be easily calculated for any bridge or viaduct under consideration. The background report by Lichtenstein (*13*) was studied to uncover the underlying probabilistic model. The calculations resulting in the basic value of $X_p = 1.4$ as used in the MBE have been reproduced with success.

Although a method based on the probabilistic analysis of the live load alone (such as the MBE method) is practical, several challenges remain: the influence of time-dependent effects, reliability of stop criteria, usage of information about the structure, and the importance of system-level assessment. Only a part of the challenges can be overcome by adjustments to the target load or factor. Verifying the reliability of an existing bridge or viaduct through proof load testing is markedly different from the design process.

The main idea behind the MBE method (i.e., the resistance is at least equal to the self-weight and the applied live load in the test) remains valuable and therefore suggestions for improvement have been provided. In summary, the improvements entail including model uncertainties in the probabilistic model, updating the traffic load description, and adopting the appropriate live load model. Dutch roads with high traffic intensity display a comparable trend in the statistical description of the load effect with span length. With a probabilistic analysis it was shown that live load models HL93 and Eurocode LM1 can provide reasonably constant proof of load factors over a large range of span lengths, for both bending and shear.

Notation

C_{0L}	Time-invariant part of the traffic load variability
D	Dead load effect (or self-weight)
I	Dynamic load allowance (or impact load)
IM	Dynamic load allowance (or impact load)
L	Live load effect (or variable load)
L_R	Unfactored live load effect because of the rating vehicle for the lanes loaded
L_T	Target load effect for the proof load test
m_I	Mean value of the impact load
m_L	Mean value of the live load
P_f	Probability of failure
R	Resistance (or load-carrying capacity of the structure)
RF	Rating factor
V_I	Coefficient of variation of the impact load
V_L	Coefficient of variation of the live load
X_p	Proof load factor
X_{pA}	Adjusted proof load factor
Z	Limit state function
β	Reliability index
β_G	Scale parameter of the Gumbel distribution
γ	Euler–Mascheroni constant
θ_L	Model uncertainty load effect produced by the traffic load
θ_{LT}	Model uncertainty load effect produced in the test
μ	Location parameter of the Gumbel distribution
Φ	Standard normal cumulative distribution function

Acknowledgments

The authors wish to express their gratitude and sincere appreciation for the fruitful discussions with Sonja A.A.M. Fennis of Rijkswaterstaat and Max A.N. Hendriks from Delft University of Technology, which have been of great help.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: E. Lantsoght, R. Steenbergen, R. de Vries; data collection: R. de Vries; analysis and interpretation of results: R. de Vries, E. Lantsoght, R. Steenbergen, M. Naaktgeboren; draft manuscript preparation: R. de Vries, E. Lantsoght, R. Steenbergen, M. Naaktgeboren. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat).

ORCID iD

Marius Naaktgeboren https://orcid.org/0000-0003-2412-8650

References

1. Lantsoght, E. O. L., C. van der Veen, A. de Boer, and D. A. Hordijk. State-of-the-Art on Load Testing of Concrete Bridges. *Engineering Structures*, Vol. 150, 2017, pp. 231–241.
2. Alampalli, S., D. M. Frangopol, J. Grimson, M. W. Halling, D. E. Kosnik, E. O. L. Lantsoght, D. Y. Yang, and E. Zhou. Bridge Load Testing: State-of-the-Practice. *Journal of Bridge Engineering*, Vol. 26, No. 3, 2021, p. 03120002.
3. Alampalli, S., D. M. Frangopol, J. Grimson, D. Kosnik, M. Halling, E. O. L. Lantsoght, J. S. Weidner, D. Y. Yang, and Y. E. Zhou. *Primer on Bridge Load Testing, TRB Circular E-C257*. Transportation Research Board, Washington, D.C., 2019.
4. Grigoriu, M., and W. B. Hall. Probabilistic Models for Proof Load Testing. *Journal of Structural Engineering*, Vol. 110, No. 2, 1984, pp. 260–274.
5. Lin, T. S., and A. S. Nowak. Proof Loading and Structural Reliability. *Reliability Engineering*, Vol. 8, 1984, pp. 85–100.
6. Rackwitz, R., and K. Schrupp. Quality Control, Proof Testing and Structural Reliability. *Structural Safety*, Vol. 2, 1985, pp. 239–244.
7. Faber, M. H., D. V. Val, and M. G. Stewart. Proof Load Testing for Bridge Assessment and Upgrading. *Engineering Structures*, Vol. 22, 2000, pp. 1677–1689.
8. Casas, J. R., and J. D. Gómez. Load Rating of Highway Bridges by Proof-Loading. *KSCE Journal of Civil Engineering*, Vol. 17, No. 3, 2013, pp. 556–567.
9. Medha, K., J. W. Schmidt, J. D. Sørensen, and S. Thöns. A Decision Theoretic Approach Towards Planning of Proof Load Tests. *13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP)*, Seoul, South Korea, 2019.
10. Schmidt, J. W., S. Thöns, M. Kapoor, C. O. Christensen, S. Engelund, and J. D. Sørensen. Challenges Related to Probabilistic Decision Analysis for Bridge Testing and Reclassification. *Frontiers in Built Environment*, Vol. 6, 2020, p. 14.
11. Zhang, W. H., D. G. Lu, J. Qin, S. Thöns, and M. H. Faber. Value of Information Analysis in Civil and Infrastructure Engineering: A Review. *Journal of Infrastructure Preservation and Resilience*, Vol. 2, 2021, p. 16.
12. American Association of State Highway Transportation Officials. *The Manual for Bridge Evaluation, Standard*, 3rd ed. AASHTO, Washington, D.C., 2018.
13. Lichtenstein, A. G. *Bridge Rating Through Nondestructive Load Testing*. Technical Report, NCHRP Project 12-28(13)A, TRB, National Research Council, Washington, D.C., 1993.
14. NCHRP. *Manual for Bridge Rating Through Load Testing, Research Results Digest, No. 234*. Transportation Research Board (TRB), Washington, D.C., 1998.
15. Madsen, H. O., S. Krenk, and N. C. Lind. *Methods of Structural Safety*. Prentice Hall, Englewood Cliffs, New Jersey, 1986, p. 403.
16. American Association of State Highway Transportation Officials. *LRFD Bridge Design Specifications, Standard*, 9th ed. AASHTO, Washington, D.C., 2020.
17. NCHRP. *Load Capacity Evaluation of Existing Bridges*. Report 301, Project 12-28(1), TRB, National Research Council, Washington, D.C., 1987.
18. Bhattacharya, B., D. Li, M. Chajes, and J. Hastings. Reliability-Based Load and Resistance Factor Rating Using in-Service Data. *Journal of Bridge Engineering*, Vol. 10, No. 5, 2005, pp. 530–543.
19. CEN. *Eurocode 0: Basis of Structural Design*. Standard, EN 1990 + A1 + A1/C2:2019, European Committee for Standardization, Brussels, Belgium, 2019.
20. Steenbergen, R. D. J. M., and A. C.W. M. Vrouwenvelder. Safety Philosophy for Existing Structures and Partial Factors for Traffic Loads on Bridges. *Heron*, Vol. 55, No. 2, 2010, pp. 123–139.
21. Nowak, A. S. Live Load Model for Highway Bridges. *Structural Safety*, Vol. 13, No. 1–2, 1993, pp. 53–66.
22. Hasofer, A. M., and N. C. Lind. Exact and Invariant Second-Moment Code Format. *Journal of the Engineering Mechanics Division*, Vol. 100, No. 1, 1974, pp. 111–121.
23. fib. *Partial Factor Methods for Existing Concrete Structures*. Bulletin 80, Recommendation, Task Group 3.1, Fédération internationale du béton, 2016.
24. JCSS. *Probabilistic Model Code*. 2015. https://www.jcss-lc.org/jcss-probabilistic-model-code/
25. Thulaseedharan, N., and M. T. Yarnold. Prioritization of Texas Prestressed Concrete Bridges for Future Truck Platoon Loading. *Bridge Structures*, Vol. 16, No. 4, 2020, pp. 155–167.
26. Gumbel, E. J. *Statistical Theory of Extreme Values and Some Practical Applications*. Applied Mathematics Series, National Bureau of Standards, Washington, D.C., Vol. 33, 1954.
27. Nowak, A. S., M. Lutomirska, and F. I. Sheikh Ibrahim. The Development of Live Load for Long Span Bridges. *Bridge Structures*, Vol. 6, No. 1, 2010, pp. 73–79.
28. O'Brien, E. J., A. J. O'Connor, and J. E. Arrigan. Procedures for Calibrating Eurocode Traffic Load Model 1 for National Conditions. *Proc., 6th International Conference on Bridge Maintenance, Safety and Management (IABMAS 2012)*, Stresa, Italy, 2012, pp. 2597–2603.
29. Enright, M. P., and D. M. Frangopol. Probabilistic Analysis of Resistance Degradation of Reinforced Concrete Bridge Beams Under Corrosion. *Engineering Structures*, Vol. 20, No. 11, 1998, pp. 960–971.
30. Smith, D. J.Understanding Terms and Jargon. In *Reliability, Maintainability and Risk*, 7th ed., (D. J. Smith, Butterworth-Heinemann, eds.), Oxford, UK, 2005, pp. 11–23.

31. Melhem, M. M., C. C. Caprani, M. G. Stewart, and S. Zhang. *Bridge Assessment Beyond the AS 5100 Deterministic Methodology*. Research Report AP-R617-20, Austroads, 2020.
32. Spaethe, G. Die Beeinflussung Der Sicherheit Eines Tragwerks Durch Probebelastung. *Bauingenieur*, Vol. 69, 1994, pp. 459–468.
33. De Vries, R., E. O. L. Lantsoght, R. D. J. M. Steenbergen, and S. A. A. M. Fennis. Reliability Assessment of Existing Reinforced Concrete Bridges and Viaducts Through Proof Load Testing. *Proc., IABMAS Conference*, Barcelona, Spain, 2022.
34. DAfStb. *DAfStb-Richtlinie: Belastungsversuche an Betonbauwerken*. Guideline, Deutscher Ausschuss für Stahlbeton, Berlin, Germany, 2020.
35. Zarate Garnica, G. I., and E. O. L. Lantsoght. Stop Criteria for Proof Load Testing of Reinforced Concrete Structures. *Proc., 2021 Session of the 13th fib International PhD Symposium in Civil Engineering*, Paris, France, 2021, pp. 195–202.
36. Zarate Garnica, G. I., E. O. L. Lantsoght, and Y. Yang. Monitoring Structural Responses During Load Testing of Reinforced Concrete Bridges: A Review. *Structure and Infrastructure Engineering*, Vol. 18, No. 10–11, 2022, pp. 1558–1580.
37. Hendawi, S., and D. M. Frangopol. System Reliability and Redundancy in Structural Design and Evaluation. *Structural Safety*, Vol. 16, 1994, pp. 47–71.
38. Schneider, R. *Time-Variant Reliability of Deteriorating Structural Systems Conditional on Inspection and Monitoring Data*. PhD thesis, Technical University of Munich, Bundesanstalt für Materialforschung und -prüfung (BAM), Munich, 2020.
39. Fu, G., L. Liu, and M. Bowman. Multiple Presence Factor for Truck Load on Highway Bridges. *Journal of Bridge Engineering*, Vol. 18, No. 3, 2013, pp. 240–249.
40. De Vries, R., E. O. L. Lantsoght, and R. D. J. M. Steenbergen. *Case Study Proof Loading in an Annual Reliability Framework*. Stevin Report 25.5-21-02, Delft University of Technology, Delft, The Netherlands, 2021.

C

Structure and Infrastructure Engineering article

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M. & Fennis, S. A. A. M. (2023b). Time-dependent reliability assessment of existing concrete bridges with varying knowledge levels by proof load testing. *Structure and Infrastructure Engineering: IABMAS 2022 Special Issue*, *20*(7-8), 1053–1067. https://doi.org/10.1080/15732479.2023.2280712

STRUCTURE AND INFRASTRUCTURE ENGINEERING
2024, VOL. 20, NOS. 7–8, 1053–1067
https://doi.org/10.1080/15732479.2023.2280712

C

Time-dependent reliability assessment of existing concrete bridges with varying knowledge levels by proof load testing

Rein de Vries[a,b], Eva O. L. Lantsoght[a,c], Raphaël D. J. M. Steenbergen[b,d] and Sonja A. A. M. Fennis[e]

[a]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands; [b]Reliable Structures, Netherlands Organisation for Applied Scientific Research (TNO), Delft, The Netherlands; [c]College of Sciences and Engineering, Universidad San Francisco de Quito, Quito, Ecuador; [d]Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium; [e]Rijkswaterstaat, Ministry of Infrastructure and Water Management, Utrecht, The Netherlands

ABSTRACT

In the evaluation of existing bridges and viaducts, relying solely on a desk study is often inadequate for determining their structural reliability. Performing a proof load test provides valuable field data that offers detailed information about the structural integrity. However, the relation between the magnitude of the load and the structural reliability is not immediately clear. This study addresses the challenges associated with determining the target load and highlights the uncertainties that play a key role. A case study is presented that shows the time-dependent character of the structural reliability and the influence of an informative and a weakly informative prior distribution in a Bayesian context. It is shown how both past traffic loads and a proof load test may contribute to the proven strength of a structure. The described method provides a starting point towards a flexible approach for proof load testing in which structure-specific knowledge levels and requirements are considered.

ARTICLE HISTORY
Received 15 March 2023
Revised 8 August 2023
Accepted 12 September 2023

KEYWORDS
Existing structures; incomplete knowledge; load testing; proof loading; reliability updating; time-dependence; traffic load

1. Introduction

Due to the constant aging of infrastructure, increased traffic load and traffic intensities, methods are explored by which the reliability of existing road bridges and viaducts can be assessed. In case limited information of the structure is available or its condition is of concern, load testing may be used to obtain additional information about the structure. Historically, before complex structural analysis was commonplace, load testing was regularly performed prior to opening a bridge to the public. In a number of countries performing a load test before use is still required (Lantsoght, van der Veen, de Boer, & Hordijk, 2017b).

Three types of load testing may be distinguished: a diagnostic test, a proof load test and a collapse test. A diagnostic load test is performed at moderate load levels to gain improved understanding of the distribution of forces, stiffness of materials or structural components, fixity of connections, composite action, etc. The measurements are typically used to adjust the structural (finite element) model and/or its input parameters. During a proof load test the level of load is typically much higher. The intent of this test is to prove that a bridge or viaduct can satisfactorily carry the traffic live loads. The method of load application varies from heavy vehicles to loading frames with ballast blocks. A load test in which the load is continuously increased until failure occurs is referred to as a collapse test. This type of test is used to determine the capacity of the bridge and study the mechanisms leading to failure (Lantsoght, 2019a).

Examples of proof load tests in the USA, Denmark and the Netherlands are provided in Zarate Garnica, Lantsoght, and Yang (2022). In the Netherlands, the tests were carried out as a precursor to proof load tests demonstrating sufficient structural reliability. In these pilot tests, the loads were continually increased until collapse was established. One of the pilot tests was performed on the prestressed concrete T-beam bridge Vechtbrug (Ensink et al., 2018). In the test, use was made of a hydraulic jack installed in the loading frame to apply a concentrated load (Figure 1).

In proof load testing the magnitude of the load to be applied, or target load, is of particular importance. If the, relatively large, target load is successfully carried by the structure then it has proven to be sufficiently structurally reliable for future use. In contrast to desk studies and numerical verification methods, different uncertainties play a role during proof load testing. In addition, the condition of the structure may be of particular concern due to the effect of deterioration or other time-dependent processes (Ellingwood, 1996).

In this article, which is an extension of De Vries et al. (2022), it is examined how the structural reliability of reinforced concrete bridges and viaducts can be established by proof load testing. From the literature study, the challenges

CONTACT Rein de Vries rein.devries@tudelft.nl Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands.

Figure 1. Collapse test being performed in October, 2016 on the Vechtbrug in The Netherlands (De Vries, Lantsoght, Steenbergen, & Fennis, 2022).

in determining the target proof load and the associated uncertainties are highlighted. An approach to address the challenges is suggested and illustrated by a case study.

2. Literature review

2.1. International standards

Proof load testing is not a standardised assessment procedure in many countries. If national guidance is lacking, standards or guidelines from other countries can provide useful insight into accepted practices. In the USA, the Manual for Bridge Evaluation (MBE) (AASHTO, 2018) is used as a guideline for diagnostic and proof load testing. The target proof load is expressed in terms of the regular load model and is magnified by a proof load factor (X_p). Its default value (1.4) was derived in a basic probabilistic analysis (Lichtenstein, 1993) that did not address the challenges described in this article. Suggested improvements to the probabilistic background are provided in De Vries, Lantsoght, Steenbergen, and Naaktgeboren (2023). Another relevant American standard is the ACI 437.2 M (ACI, 2013a) which describes the requirements for proof load testing of existing concrete buildings including loading protocols and acceptance criteria.

Recently the German committee for reinforced concrete published a new version of its guideline for proof load tests on concrete structures (DAfStb, 2020). The guideline is intended for buildings, but refers in more general terms such to structures and components. The magnitude of the proof load is expressed in a format that resembles the load effect in Equation (6.10) of EN 1990:2019 (CEN, 2019). An interesting aspect of the guideline is the consideration of multiple similar components. It is recognised that two or more components of a structure may not be exactly the same but may be very similar. Similarity is to be expected when a component occurs multiple times and the same design applies (e.g. the floors in a building). The additional uncertainty introduced by not testing every component is compensated by increasing the test load slightly compared to the case where only each component is tested. The increase depends on the total number of components (elements), the number of elements tested (sample size) and the coefficient of variation (COV) associated with the material that governs failure (Marx, 2019).

2.2. State-of-the-art

2.2.1. Proof load testing

Proof load testing is still an active field of research and continues to gain attention due to the growing need for versatile assessment methods for existing structures (Lantsoght et al., 2017b). It is desirable that the assessment of infrastructure is not overly conservative because that may lead to the replacement or upgrading of bridges that are actually satisfactory. Proof load magnitudes can vary depending on the load rating, dead/live load ratios, degradation, bridge age, reference period and prior service loads (Faber, Val, & Stewart, 2000). Recent advances in measurement techniques and the treatment of proof load testing within a reliability-based decision-making context are described in Lantsoght (2019b). In Casas and Gómez (2013) proof load factors are presented that were developed as part of the large scale ARCHES (Assessment and Rehabilitation of Central European Highway Structures) project. The study presents a sophistication with respect to current code-based approaches by making use of recent traffic load data and differentiating the case where bridge documentation is available and the case where it is not.

Because the desired remaining life of existing structures is often less than the normative design life (e.g. 50–100 years), flexibility in choosing an appropriate reference period is needed (Vrouwenvelder & Scholten, 2010). The reference period holds significance in the context of structural reliability as it considers the time-dependent nature of

reliability. Using a time-dependent reliability analysis, it is possible to directly determine if the structural reliability is sufficient for the desired remaining lifespan. An early description of the time-dependence in relation to proof load testing is found in Spaethe (1994). During the proof load test the reliability of the structure is low, due to the relatively large load that is applied, but afterwards reliability increases – in case of a successful test. In more recent works by Schacht, Bolle, and Marx (2019) and Frangopol, Yang, Lantsoght, and Steenbergen (2019), the decrease of reliability with time in case of deterioration is also recognised. In the recent developments on proof load testing the link with structural reliability is recognised. However, the aspects in which the proof load testing situation is markedly different from the design situation are not sufficiently recognised and addressed. The incomplete understanding may lead to over-conservative assessment methods or potentially unsafe situations.

2.2.2. System reliability

The assessment of structures by proof load testing is markedly different from the conventional design procedure for new structures. In proof load testing, only the entire system's performance can be observed, whereas in the design process, verifications are typically conducted at the component level rather than the system level. Therefore, system reliability is of particular interest to proof load testing. A system may be thought to be comprised of multiple components. In this scheme, the components may act in parallel or in series. In addition, the combined performance of a group of elements may interact with one component, or another group. In the context of system failure a diagram of the interaction is called a fault tree (Fussell, 1975).

Various methods may be used to calculate the failure probability of a system. The Monte Carlo Simulation (MCS) is a straightforward method that is always applicable, but it is computationally expensive (Metropolis & Ulam, 1949). For better computational efficiency, the equivalent planes method (Roscoe, Diermanse, & Vrouwenvelder, 2015) is used in this article. The method is based on the equivalent component method (Gollwitzer & Rackwitz, 1983) and the first-order system reliability method described by Hohenbichler and Rackwitz (1982). The reliability of the individual components may be determined using the first-order reliability method (FORM) (Hasofer & Lind, 1974), the second-order reliability method (SORM) (Breitung, 1984) or any other method that also provides the influence coefficients of the random variables.

2.2.3. Reliability updating

Proof load testing as a means to assess the performance of a structure in relation to its structural reliability was recognised in the 1980s, with pioneering work by Grigoriu and Hall (1984), Lin and Nowak (1984), and Rackwitz and Schrupp (1985). Proof load testing is starting to be considered in the light of maintenance and durability. In particular, the so-called 'updating' of structural reliability as performed on the basis of Bayesian theory provides the opportunity to incorporate various sources of information. The theory can provide a mathematical basis for the updated distributions of the reliability (Yuefei, Dagang, & Xueping, 2014).

The more generally applicable Bayesian decision theory is also used in the context of proof load testing. It can provide decision support and the identification of information to aid in modelling and monitoring of structures (Schmidt et al., 2020). In Bayesian decision theory, today often mentioned in the context of value of information, the state of information about a structure at a given point in time results in three possible types of analysis: prior analysis, posterior analysis and pre-posterior analysis (Zhang, Lu, Qin, Thöns, & Faber, 2021). Each stage in the analysis has its own set of possibilities (E, X, A, Θ), dependent on earlier choices or outcomes (Figure 2).

The collecting strategy (E) involves selecting informative observations or experiments (X) to enhance the accuracy of the posterior analysis. Decision alternatives (A) represent available choices prior to obtaining new information, while random outcomes (Θ) depict uncertain events associated with each chosen alternative. All possible paths lead to certain consequences or costs (C), which may also include the risk of losing human life. Proof load testing may be viewed as a source of information and a pre-posterior analysis can be used to determine its value (Nishijima & Faber, 2007). A decision analytic approach was developed for reclassifying bridges using proof load testing information and pre-posterior decision analysis (Kapoor, Christensen, Schmidt, Sørensen, & Thöns, 2023).

3. Challenges and suggested approach

The literature at the interface between proof load testing and structural reliability has been briefly described. In relation to a full probabilistic treatment of proof load testing, it is believed that a number of aspects deserve further attention. These aspects, and their combined usage, give rise to the challenges and the suggested approach described in the following subsections.

3.1. Time-dependence

The time-dependent nature of structural reliability may be addressed by using an annual reliability safety format. In this way, flexibility with regard to the remaining functional life span is obtained (De Vries et al., 2022). Considering the time dependence is also beneficial in relation to the proven strength by past traffic loads. In a sense, every truck passing a bridge may be viewed as a test, contributing to the service-proven strength of the bridge (Wang, Ellingwood, & Zureick, 2011). Standard texts on reliability theory describe the concept of proven strength and degradation (or wear out) via the 'bathtub curve' of the failure probability (Smith, 2005).

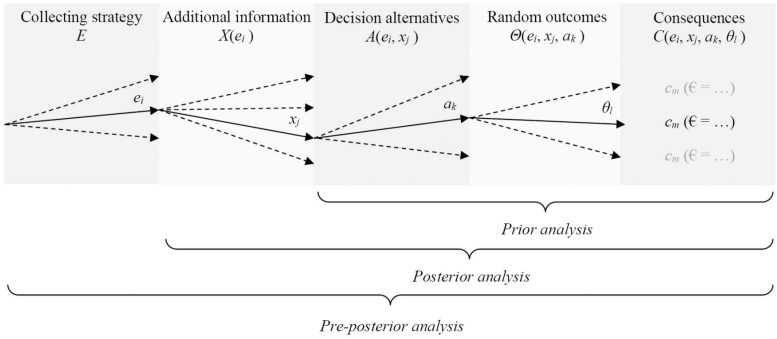


Figure 2. Analysis type depending on the state of information (De Vries et al., 2022).

3.2. Stop criteria and structural reliability

During the proof load test signs of distress may appear when the load is gradually increased towards the target load. To test whether distress occurs, stop criteria may be evaluated. Stop criteria typically address unwanted structural behaviour, but act on a measurable property (indicator). For example, excessive strains indicate that the reinforcement is yielding, then a stop criteria on the strains is formulated. In the German guideline for proof load testing (DAfStb, 2020) various criteria are provided. Also in the Czech Republic, Slovakia, Spain, Switzerland, Poland and Hungary guidelines with stop criteria are available (Lantsoght, Yang, Van der Veen, Hordijk, & De Boer, 2019). The effectiveness of various stop criteria for proof load tests is studied in Zarate Garnica and Lantsoght (2021). In the United States of America acceptance criteria apply and may be found in ACI 437.2 M-13 (ACI, 2013b). Acceptance criteria are used to evaluate the state of the structure or component after the test.

The link between structural reliability, after and during the load test, and the formulation of stop criteria including safety margins needs to be explored. If the safety margins of stop criteria are too stringent, the proof load test may be aborted long before the structure is actually near its maximum capacity. Theoretically, the capacity of a structure may be extrapolated from the observed behaviour (e.g. location and width of cracks, deflection, etc.) at low levels of loading. However, this approach has not been applied in practice or thoroughly studied yet.

3.3. Knowledge level

A flexible method is needed that can utilise various types of information. Various data sources and their influence on the state of information are collected in Figure 3. A balance should be sought between how much information is collected and analysed prior a proof load test and regarding the proof load test itself as the primary source of information (Kapoor, Schmidt, Sørensen, & Thöns, 2019). It is suggested to follow a Bayesian approach in which the state of information plays a key role in structural reliability predictions (De Vries et al., 2022). In this approach, large uncertainties may be introduced purposely as 'objective' low informative priors (Ditlevsen & Vrouwenvelder, 1994). If available, other broad prior distributions following from basic information (bridge span, traffic type, etc.) may be included.

3.4. System-level assessment

In a system-level assessment, the performance of multiple components and spatial variability is incorporated. In addition to the physical components of a bridge, its cross-sections may also be regarded as components. By modelling the bridge as a system including correlations the reliability analysis can address the associated uncertainties directly (De Vries et al., 2022). Also here Bayesian analysis can be utilised to update the system reliability (joint PDF) with incomplete and uncertain information about a limited number of parameters (Schneider, 2020).

An example of a simplified bridge with two spans is provided in Figure 4. In this case only load and spatial variation in the longitudinal direction is considered (and not over the bridge width). The structural schematisation with a distributed load indicates three common design checks: bending moment at midspan (blue), support moment (green) and shear force near the support (orange). The corresponding cross-sections are indicated in the lower part of the figure. Because of spatially varying material properties and execution details other cross-sections may be critical. In Figure 4, these cross-sections have been drawn with the same colour, but transparently.

4. Methods

To address the challenges and implement the suggested approach discussed in Section 3, a number of probabilistic methods are required. The methods described in the following subsections will be utilised in the case study (Section 5).

C

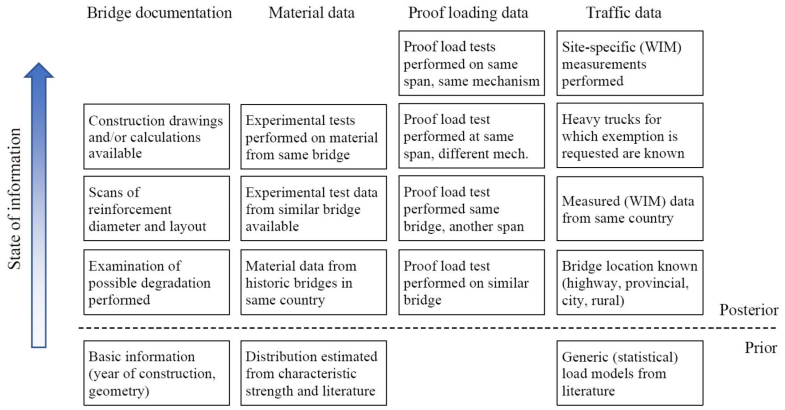


Figure 3. State of information considering various information sources (De Vries et al., 2022).

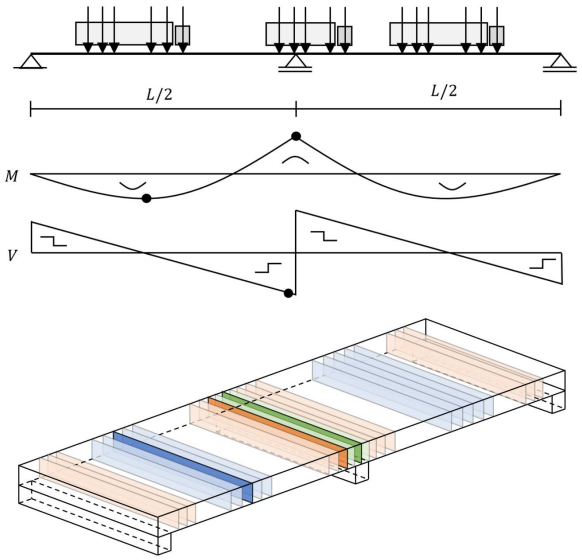


Figure 4. Visualisation of the cross-sections to be assessed in a system-level assessment (De Vries et al., 2022).

4.1. System reliability

Calculating the system reliability using the equivalent planes method works on the basis of two components. The (linearised) limit state functions Z_i of two components may be written using the reliability index (β_i) of the component and the influence coefficients (α_{ij}) of all random variables present in the system (Roscoe et al., 2015):

$$Z_1 = \beta_1 + \alpha_{11}U_{11} + \alpha_{12}U_{12} + \cdots + \alpha_{1n}U_{1n} \quad (1a)$$

$$Z_2 = \beta_2 + \alpha_{21}U_{21} + \alpha_{21}U_{21} + \cdots + \alpha_{2n}U_{2n} \quad (1b)$$

In this equation, U_{ij} are standard normally distributed random variables that are statistically independent (i.e. uncorrelated) within the component. However, auto-correlation $\rho_j = \rho(U_{1j}, U_{2j})$ may exist. In case one or more random variables are used in both components, correlation

between the two limit state functions exists and it is calculated via:

$$\rho = \rho(Z_1, Z_2) = \sum_{j=1}^{n} \alpha_{1j}\alpha_{2j}\rho_j \tag{2}$$

Using the component correlation coefficient ρ, the limit state functions of the two components may be expressed using just two independent standard normally distributed random variables (U_1 and U_2):

$$Z_1 = \beta_1 + U_1 \tag{3a}$$

$$Z_2 = \beta_2 + \rho U_1 + \sqrt{1-\rho^2}U_2 \tag{3b}$$

It should be noted that standard normal random variables U_1 and U_2 are unrelated to the random variables U_{ij} in Equation (1a) and (1b). The same holds for correlation coefficient ρ used in Equation (3) and the correlation coefficients denoted as ρ_j in Equation (2).

In case of a parallel system, failure occurs if components 1 *and* 2 fail. In a series system, failure occurs if component 1 *or* 2 fails. In the latter case the system failure probability (of only these two components) $P_{f,or}$ is calculated via:

$$P_{f,or} = P_{f,1} + P_{f,2} - P_{f,and} \tag{4}$$

where $P_{f,1} = P(Z_1 < 0)$, $P_{f,2} = P(Z_2 < 0)$ and $P_{f,and} = P(Z_1 < 0 \cap Z_2 < 0)$. The last probability may be rewritten to a conditional probability such that it may be easily computed using a standard reliability method such as FORM (Hasofer & Lind, 1974).

In case of more components, the combination process needs to be repeated several times until just one component remains. Each time two components are combined to give a new component that replaces the two original components. The most accurate results are obtained when the components with the highest correlation between the limit state functions are combined first in every step (Gong & Zhou, 2017).

4.2. Time-dependent reliability analysis

4.2.1. Limit state functions

The limit state function used in a time-dependent analysis is the same as used for a regular probabilistic analysis. The main difference is in the reference period used for the variable loads. In a regular probabilistic analysis, the reference period of the load effect will be large, commonly 50–100 years, but if the time-dependence is explicitly studied it will be small. As discussed in Section 3.1, the reference period is chosen as one year, resulting in annual reliability values. Because of the time-dependence auto-correlation of the random variable becomes important. Normally only the variable loads will be uncorrelated in time.

The limit state function for the probabilistic analysis of structural failure may be formulated in terms of resistance (R) and the load effect fib (2016). The load effect is split into the contributions from the dead load (G_{DL}), the superimposed dead load (G_{SDL}) and the variable load (Q). Both the resistance and load effect are associated with model uncertainty (θ_R and θ_E). Specific to the variable load is the time-invariant part of the variability (C_{0Q}). In addition, two random variables are added that account for the deterioration of the resistance (c_R) and trend in traffic load (c_Q):

$$Z = \theta_R c_R R - \theta_E(G_{DL} + G_{SDL} + c_Q C_{0Q} Q) \tag{5}$$

The limit state function is subsequently adjusted to incorporate the proof load test event. When a proof load test is performed an additional term (Q_{PL}) is included for the proof load effect in the limit state function:

$$Z = \theta_R c_R R - \theta_E\left[G_{DL} + G_{SDL} + \max(c_Q C_{0Q} Q, Q_{PL})\right] \tag{6}$$

where the max-function is used to ensure that the regular traffic load is also considered for the year in which a proof load test is conducted. If a very low target load is used for proof load testing, it will have no effect. The adoption of the same model uncertainty for the load effect of traffic action and proof load testing is discussed in Section 6.

4.2.2. Conditional annual reliability

The annual reliability is calculated under the condition that no failure occurs in any of the years before the year under consideration. Using the following events:

A failure in the year i;
B failure in the years 1 to i – 1;
B' no failure in the years 1 to i – 1 (complement).

the conditional annual probability of failure can be written as:

$$P(A|B') = \frac{P(A \cap B')}{P(B')} = \frac{P(A \cup B) - P(B)}{1 - P(B)} \tag{7}$$

The probability $P(A \cup B)$ may be read as the cumulative failure probability up to and including the year i, whereas $P(B)$ is the cumulative failure probability up to, but not including, the year i. In the case of a proof load test, the failure probability (in the year) after the test will be significantly less since $P(B)$ includes the relatively large failure probability associated with proof load testing.

To calculate the conditional annual reliability using the system reliability method, first the reliability index and influence coefficients of each year need to be calculated, e.g. using FORM. The individual years are the system components in this calculation. Next, the cumulative probability of failure can be calculated using the equivalent planes method (OR-combination). Then, the conditional probability of failure in year i is:

$$P_{f,cond,i} = \frac{P_{f,i} - P_{f,i-1}}{1 - P_{f,i-1}} \tag{8}$$

where $P_{f,i}$ is the cumulative failure probability up to and including the year i. In the first year no conditionality holds and thus $P_{f,cond,1} = P_{f,1}$.

The conditional probability calculation via Equations (7) or (8), in combination with the limit state functions in Section 4.2.1, effectively performs the update of structural reliability. In this updating process all random variables that are correlated in time will be updated each year. They

C

include, the resistance, permanent loads, model uncertainties, the deterioration of the resistance and the trend in traffic load – but not the traffic load effect (Q). When moving to a context in which subjective knowledge about the resistance plays a significant role, the updating process is often referred to as Bayesian updating.

C

4.3. Bayesian reliability updating

4.3.1. Updating the resistance distribution

At the heart of the probabilistic treatment of proof load testing is the expectation that after a successful test, the load effect produced during a proof load test E_{PL} can be considered as a lower bound for the resistance (R) Lin and Nowak (1984):

$$R \geq E_{\mathrm{PL}} \tag{9}$$

Truncating the left tail of the random variable for the resistance (R) leads to the posterior distribution for the resistance and may be expressed as:

$$f_R{}^*(r) = \begin{cases} \dfrac{f_R(r)}{1 - F_R(E_{\mathrm{PL}})} & \text{for } r \geq E_{\mathrm{PL}} \\ 0 & \text{for } r < E_{\mathrm{PL}} \end{cases} \tag{10}$$

where $f_R(\cdot)$ is the probability density function and $F_R(\cdot)$ is the cumulative density function of the prior distribution for R. It may also be obtained via the application of Bayes' theorem together with a likelihood function providing the value 0 when $r < E_{\mathrm{PL}}$ and 1 otherwise. As noted by Ditlevsen and Madsen (1996), the proof loading must be made at rather high levels in order to achieve a high reliability. If E_{PL} is a deterministic value, the probabilistic calculation may be performed directly using the updated distribution of resistance R, i.e. Equation (10). However, commonly the load effect achieved within a proof load test is not precisely known and is better described using random variables (see Section 4.3.2). It should be noted that Equation (10) is not needed when the update is performed using the conditional probability calculation in Section 4.2.2.

As an assessment method for existing structures, proof load testing is typically applied when large doubts exist about the resistance of the structure. Even if drawings and original calculations are available, there may be such significant evidence of deterioration that they become irrelevant. In this context, often a weakly or low-informative prior distribution is desired (Ditlevsen & Vrouwenvelder, 1994). A reliability analysis will typically result in an unacceptably low reliability index when using a prior distribution for R with such large uncertainty. But, after a successful proof load test the truncation, Equation (10), will lead to an increase of the reliability index. The prior distribution does not necessarily need to be a (log)normal distribution. In Kapoor, Sørensen, Ghosh, and Thöns (2021), a uniform distribution was chosen to reflect the lack of knowledge about the resistance (Figure 5).

4.3.2. Limit state function

When transitioning from a context where the resistance is based on reliable information (informative) to a context with significant uncertainty surrounding this parameter (weakly informative), it is necessary to adjust the limit state function to reflect the situation. As in the time-dependent analysis, the limit function for a specific failure mechanism is considered (e.g. bending or shear). The corresponding limit state function is expressed as:

$$Z = R - \theta_E(G_{\mathrm{DL}} + G_{\mathrm{SDL}} + C_{0Q}Q) \tag{11}$$

where the definition of the random variables is the same as in Section 4.2. Here the time-dependent coefficients c_R and c_Q have been excluded for simplicity. In this context, no mechanical model will be used to calculate the resistance from a set of basic parameters (such as geometry and material properties). Therefore, θ_R is not explicitly included in the limit state function, the remaining random variable R may be regarded as the resistance including any probabilistic uncertainty. As it is already known that the structure can carry the permanent loads (G_{DL} and G_{SDL}), they may be eliminated from the limit state equation:

$$\begin{aligned} Z &= (\theta_E G_{\mathrm{DL}} + \theta_E G_{\mathrm{SDL}} + \hat{R}) - \theta_E(G_{\mathrm{DL}} + G_{\mathrm{SDL}} + C_{0Q}Q) \\ &= \hat{R} - \theta_E C_{0Q} Q \end{aligned} \tag{12}$$

where $\hat{R}$ represents the remainder of the capacity available to resist the variable traffic load. Note that this principle is similar to the rating factor used in the MBE (AASHTO, 2018).

At the moment of proof load testing, the load effect following from the traffic is replaced by the load effect achieved during the proof load test (assuming no traffic is allowed onto the bridge during the test):

$$Z_{\mathrm{PL}} = \hat{R} - \theta_{E,\mathrm{PL}} Q_{\mathrm{PL}} \tag{13}$$

where $\theta_{E,\mathrm{PL}}$ is the model uncertainty of the load effect specific to the proof load testing situation. This model uncertainty may be smaller than the one applied in the traffic load situation because it only needs to account for a load or vehicle placed at a known location. Normally the same structural schematisation or finite element model will be used to calculate the load effect in both the regular traffic

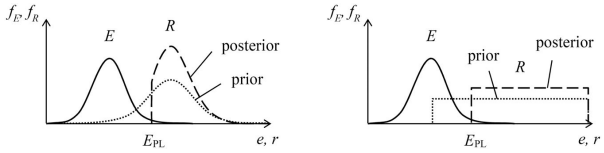


Figure 5. Schematic representation of the Bayesian update of the reliability distribution following a successful proof load test.

load and proof load testing situation. This implies a reasonably strong correlation between θ_E and $\theta_{E,PL}$.

4.3.3. Lower bound

Although customary, it is not necessary to treat the Bayesian (or probabilistic model) uncertainty of the resistance R in the same way as the other random variables. Instead of calculating the predictive posterior (integrating the posterior distributions), the failure probability may also be calculated using other methods such as point estimates, nested reliability analysis and confidence bounds (Der Kiureghian, 2022). To reduce complexity, specific attention can be paid to the first 'slice' of the posterior distribution – essentially providing the most conservative reliability estimate.

Conceptually, making this slice increasingly smaller results in the resistance being equal to the proof load effect (Figure 6). The limit state function in Equation (13) may then be written as (including model uncertainties):

$$Z = \theta_{E,\,PL} Q_{PL} - \theta_E C_{0Q} Q \tag{14}$$

This limit state function may also be directly obtained by assuming that the resistance is at least equal to the load effect caused by the permanent loads and the proof load ($R \geq G + Q_{PL}$) as shown in De Vries et al. (2023). This assumption lies at the basis of the probabilistic background for the proof load testing method in the MBE (AASHTO, 2018) as described by Lichtenstein (1993).

5. Case study

5.1. Description

A case study was performed to explore the implications of the challenges described in Section 3 using the methods described in Section 4. The hypothetical structure under consideration is a concrete slab bridge with a relatively short span of $L = 10$. This type of bridge is very common in the Netherlands and also in many other countries (Christensen et al., 2022). Most of the bridges that are in service today were built during the 1960s and 1970s. Due to the continuously increasing traffic intensity and loads, bridges situated in the highway network are of primary interest in relation to their (load-carrying) capacity. If assessed via today's Eurocode standards these bridges and viaducts often do not fulfil capacity requirements (Lantsoght, Van der Veen, De Boer, & Hordijk, 2017a).

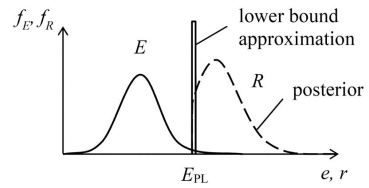


Figure 6. Schematic representation of the lower bound approximation when compared to a posterior distribution of the resistance.

5.2. Time-dependent analysis

5.2.1. Assumed design

For this example, it is assumed that the structure was built in 1960 and designed according to the prevailing standards of that time (KIVI, 1938, 1950). The traffic load used in its dated design is inappropriate when compared to today's high traffic intensity. But, the design values of material properties (e.g. steel and concrete strength) were quite conservative. As a result, old bridges and viaducts can still possess adequate structural strength to resist today's higher loads. In this case study, only the bending moment at midspan will be considered. In reality, the shear capacity of the slab near the supports and the capacity of other bridge components will require assessment as well.

In case the original bridge documentation such as drawings and calculations are still available, they may be used to infer the (prior) probabilistic description of the resistance of the structure. In this case, the bridge documentation is not available. Therefore, its design was 'reverse engineered' by using historic standards (Harrewijn, Vergoossen, & Lantsoght, 2021). For simplicity, only the slab bridge's rightmost lane is considered, primarily used by trucks. This conservative approach does not take into account the distribution of forces that typically occurs across multiple lanes. A top view and cross-section with the inferred bottom reinforcement layout from the historic standards is provided in Figure 7 (De Vries, Lantsoght, & Steenbergen, 2021).

The bending moment resistance (R) is calculated from the balance of normal forces in the cross-section when the reinforcement yields (Figure 8). Equating the force in the concrete (F_c) with the force in the reinforcing steel (F_s) and solving for the location of the neutral axis (x) gives:

$$F_c = F_s \iff \alpha_{cc} f_c b \lambda x = A_s f_y$$
$$x = \frac{A_s f_y}{\alpha_{cc} f_c b \lambda} \tag{15}$$

Using the effective depth $d = h - a$ the moment arm is calculated as $z = d - \lambda x/2$. Note that the compressive stress block reduction factor (λ) disappears in the expression moment arm when neutral axis location x, Equation (15), is inserted. Finally, the moment resistance is obtained as:

$$R = F_c z = F_s z = A_s f_y \left(d - \frac{A_s f_y}{2 a_{cc} f_c b} \right) \tag{16}$$

where A_s is the cross-sectional area of the reinforcement over cross-section width $b = 3$ m, f_y is the yield strength of the reinforcement, f_c is the concrete compressive strength. The concrete compressive strength is reduced with factor $\alpha_{cc} = 0.85$ to account for long-term effects and possible unfavourable effects from the way the load is applied.

5.2.2. Probabilistic model

The limit state equations in Section 4.2 are used in the following to perform the time-dependent reliability analysis. The random variables are described in Table 1, where each variable is characterised by a distribution, the mean value, the COV and the auto-correlation coefficient. The auto-

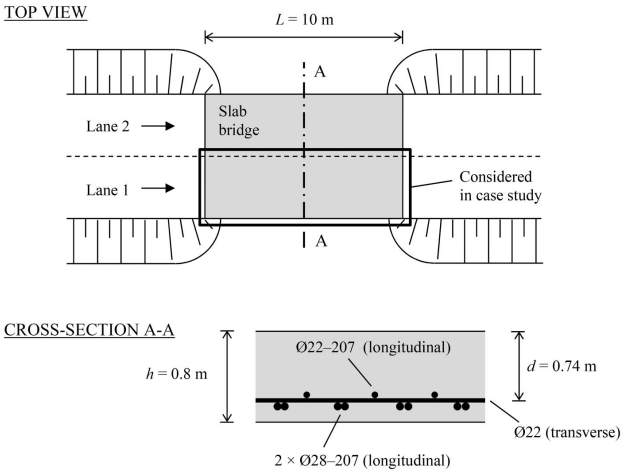


Figure 7. Layout of the reverse-engineered bottom reinforcement of the slab. Rebar size and spacing in mm.

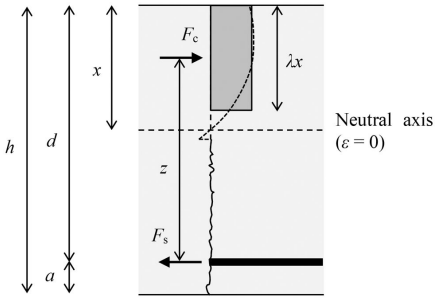


Figure 8. Calculation of moment resistance capacity.

correlation coefficient describes the correlation of the random variable in time. The chosen distribution types and parameter values are based on fib (2016) and JCSS (2015).

The now obsolete concrete type K250 (referring to a compressive strength of 250 kg/cm^2) was commonly used in the Netherlands in the 1960s. The COV of 0.15, a value commonly used for concrete, is increased to 0.20 to describe the uncertainty associated with historic concrete. Back then, smooth (i.e. not ribbed) steel rebars were extensively applied. The characteristic value (5th percentile) of the QR24 reinforcing steel yield strength corresponds to 2400 kg/cm^2 ≈ 235 N/mm^2 when a lognormal distribution is assumed with a COV of 0.05 (RWS, 2013). A mean value of 1.1 is used for the time-independent uncertainty of the traffic load (C_{0Q}) to include dynamic effects.

WIM data from 2015 was analysed to determine the load effect (Q), expressed as the largest bending moment at midspan within a certain period of time. Only the traffic in the right-most lane, where the trucks drive, has been analysed. Vehicles with a length smaller than 7 m, such as cars and small vans, have been excluded from the dataset. The combined effect of all vehicle axle loads located on the bridge is calculated via the linear superposition. Effectively a 'train' of axles is moved gradually over the bridge to record the largest moment at midspan within a certain period of time. Over a period of one year the weekly maxima have been collected for various highways in the Netherlands (A16, A27, A50 and A67). Depending on the span length, highway location and load effect considered (i.e. bending or shear), different values for the mean and COV of Q are obtained (De Vries et al., 2023). The values used for this case study represent an average of the considered highway locations.

Following fib (2016), the area of the reinforcing steel (A_s) is not included as a random variable because its small variability. Note that the resistance model uncertainty (θ_R) is intended to account for any remaining uncertainty that may exist between modelling and reality. If corrosion of the reinforcement plays a significant role, the decrease of the reinforcement area should not be neglected and may also be measured to update reliability predictions (Jacinto, Neves, & Santos, 2016). Instead of modifying parameter A_s, a separate and more general deterioration parameter c_R is included here. Use is made of the following relations for the time-dependent coefficients to include the deterioration of the resistance and a trend in the traffic load:

$$c_R(t) = \begin{cases} 1 & t \le t_{R0} \\ 1 - \Delta c_R(t - t_{R0}) & t > t_{R0} \end{cases} \quad (17a)$$

$$c_Q(t) = c_{Q0} + \Delta c_Q t \quad (17b)$$

where the parameters are random variables, listed as well in Table 1. The mean value of the parameters was chosen in such a way that the annual reliability is insufficient around the year 2020.

Table 1. Random variables used in the limit state function.

Var.	Description	Distribution	Mean	COV	Auto-corr.
θ_R	Model uncertainty of the resistance	Lognormal	1	0.05	1
f_c	Concrete compressive strength (K250)	Lognormal	21.1 MPa	0.20	1
f_y	Reinforcement steel yield stress (QR24)	Lognormal	261 MPa	0.05	1
h	Height of the slab	Normal	0.8 m	0.02	1
a	Distance of reinforcement to surface	Gamma	0.057 m	0.17	1
θ_E	Model uncertainty of the load effect	Lognormal	1	0.11	1
G_{DL}	Load effect of the dead load	Normal	721 kNm	0.05	1
G_{SDL}	Load effect of the superimposed dead load	Normal	101 kNm	0.1	1
C_{0Q}	Time-independent uncertainty of the variable load, including bias for dynamic load effect	Lognormal	1.1	0.1	1
Q	Load effect of the traffic load, annual maximum	Gumbel	1150 kNm	0.025	0
t_{R0}	Initiation time to deterioration	Lognormal	20 years	0.1	1
Δ_{cR}	Degradation per year	Lognormal	0.0025	0.1	1
c_{Q0}	Starting value of the trend	Lognormal	0.78	0.1	1
Δ_{cQ}	Increase of traffic load per year	Lognormal	0.004	0.1	1
Q_{PL}	Load effect of the proof load	Normal	1800 and 2000 kNm	0.01	1

This article aims to provide methods for updating the structural reliability after a proof load test, not to provide an accurate description of resistance degradation. The chosen degradation model includes a time to initiation (t_{R0}), followed by a linear reduction of strength (Enright & Frangopol, 1998). Corrosion leading to a reduction of the effective steel area in a cross-section was modelled by a quadratic function in Vu and Stewart (2000). In case of deterioration, a large degree of uncertainty exists with respect to the current capacity of the bridge. In this example only a limited amount of uncertainty is considered for simplicity. It thus represents the rather uncommon scenario where the deterioration process is well-known.

5.2.3. Results

Using the presented probabilistic description, a time-dependent reliability analysis can be performed. The independent reliability analysis for each year (a component) was performed using the improved SORM approximation by Hohenbichler, Gollwitzer, Kruse, and Rackwitz (1987). This approximation provides a good balance between computational effort and accuracy in the presented case study. For the combination of components the FORM (Hasofer & Lind, 1974) is used. Comparisons with MCS indicated an acceptable difference of about a tenth in the annual reliability index. The result of the calculations is displayed in Figure 9.

The base case displays the reliability without traffic trend and degradation. In this case, the annual reliability increases gradually due to proven strength of past traffic loads. The traffic trend and degradation are incorporated subsequently to display their detrimental effect on the evolution of the annual reliability. A higher reliability is attained in the first years when including the traffic trend because the adopted linear trend expresses a reduction before 2015 and an increase afterwards.

Note that in this example the parameters of the degradation and traffic load trend have been tuned to yield a reliability index that drops below the acceptable annual reliability $\beta = 4$ for CC3 (Steenbergen & Vrouwenvelder, 2010) around 2020. In a real-life situation, the parameters will need to be determined by studying the effect of all possible degradation mechanisms and the actual trend in traffic loads.

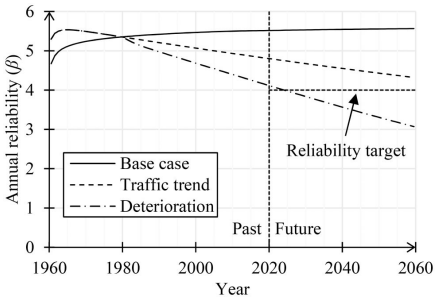


Figure 9. Development of the conditional annual reliability with time, incorporating a traffic load trend and deterioration (De Vries et al., 2022).

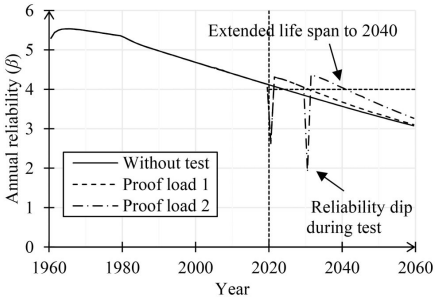


Figure 10. Effect of proof load testing on the annual reliability (De Vries et al., 2022).

Proof load testing is adopted to ensure the bridge meets the required structural reliability. When a proof load test is performed, an additional term is included for the proof load effect in the limit state function, see Equation (6). The first proof load test is performed in the year 2020 and has a

target (mean) value of 1800 kNm. Then, in the year 2030, the second proof load test is performed with a higher target load effect of 2000 kNm (Figure 10). In the year the test is performed, the annual reliability is markedly lower, but as a reward, the reliability in the following years is higher. The low reliability indices during proof load testing are below the acceptable target reliability. Hence, further safety measures are necessary during the load test, such as closing off the area underneath the viaduct or bridge. In addition, instrumenting the structure and evaluating the stop criteria after each load cycle can avert damage and collapse. Incorporating stop criteria results in an increased reliability index during the test. But if the target load is not attained, the post-test reliability is inadequate.

The target loads have been determined in an iterative fashion such that the annual reliability remains above the target in the next 10 years. Alternatively, the higher target load could have been applied directly in 2020, also leading to sufficient reliability until 2040. But, then the probability of failure in the first test in 2020 would be larger.

5.3. Bayesian reliability updating

5.3.1. Weakly informative prior

Instead of using a mechanical model, as adopted in the previous time-dependent analysis (Section 5.2), here a weakly informative prior distribution is used. This situation describes the other end of the spectrum; a situation in which very little is known about the resistance. The limit state function of Equation (12) is used to update $\hat{R}$, i.e. the remainder of the capacity available to resist the variable traffic load. To enable comparisons, the mean value of the load effect to be attained during the proof load is chosen as $m_{Q,\mathrm{PL}} = 1800$ kNm. This value corresponds to the initial proof load test conducted in the time-dependent analysis of Section 5.2.

The influence of the prior distribution is studied to determine if the distributions are indeed weakly informative. Three different types of prior distributions are employed to compare the reliability calculation outcomes. The parameters of the prior distributions will be chosen such that only little extra, sensible but subjective, information is included (Ditlevsen & Vrouwenvelder, 1994). Information about the traffic load will always be available because its statistical modelling is required to perform the reliability calculation once the posterior of $\hat{R}$ is obtained. (The accuracy of the traffic load model may be represented in the time-invariant coefficient C_{0Q}.) Therefore, the mean value of the annual traffic load is used as additional information to determine the prior distribution parameters.

For the normal prior distribution the mean value is equal to the mean value of the annual traffic load effect (1150 kNm, see Table 1). In addition, a normal prior is considered with the mean value equal to 1.5 times the traffic load effect (1.5 ▫ 1150 kNm = 1725 kNm). The factor 1.5 indicates that one expects a positive outcome – merited by the consideration of performing a proof load test in the first place. To reflect the large uncertainty, the value of the COV is chosen as 0.5. For the uniform prior distribution, the lower bound is not important. For the upper bound, the safety margin associated with new structures may be utilised. Referring to Table 1, the average capacity to resist live loads is $4100 - 721 - 101 = 3278$ kNm, which is about a factor 3 higher than the average annual traffic load effect. Therefore, the upper bound parameter value is chosen as 3 ▫ 1150 = 3450 kNm. A triangular distribution, with the same bounds as the uniform distribution, is also included to (conservatively) express a stronger belief in lower resistance values (Figure 11).

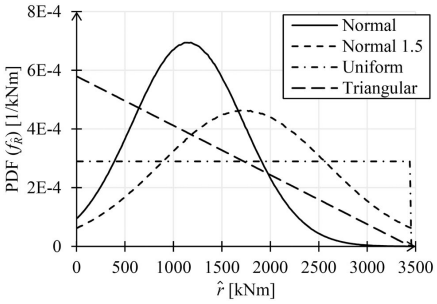


Figure 11. Prior distributions considered for the resistance.

5.3.2. Results

The Bayesian update is performed in a MCS by removing the samples that do not survive the proof load test. The number of samples in the simulation was $5 \cdot 10^8$. This update procedure also accounts for the uncertainty with regard to the proof load effect – in contrast to Equation (10). For simplicity, it is assumed that the same model uncertainty holds for the regular traffic load and the proof load testing situation – in line with the time-dependent example (Section 5.2). By assuming $\theta_E = \theta_{E,\mathrm{PL}}$ the single remaining model uncertainty may be eliminated from the equations (i.e. its contribution is now incorporated in the resistance $\hat{R}$). For each of the prior distribution types (normal, uniform and triangular) the reliability analysis is performed. The posterior distributions obtained for $\hat{R}$ are plotted in Figure 12. The result of the calculations is expressed as the annual reliability index for the first year after a successful proof load test (Table 2).

There is slight difference between the outcomes, indicating some sensitivity to the chosen prior distribution. The smallest reliability index is obtained using a normal prior distribution with the mean value equal to the mean value of the traffic load effect – which was to expected with this conservative prior. Additional calculations were performed in which the COV of the load Q has been increased and the shape of the right tail was varied between light (Weibull) and heavy (Fréchet). The same relative differences between the reliability indices were found, indicating no apparent sensitivity to such alterations.

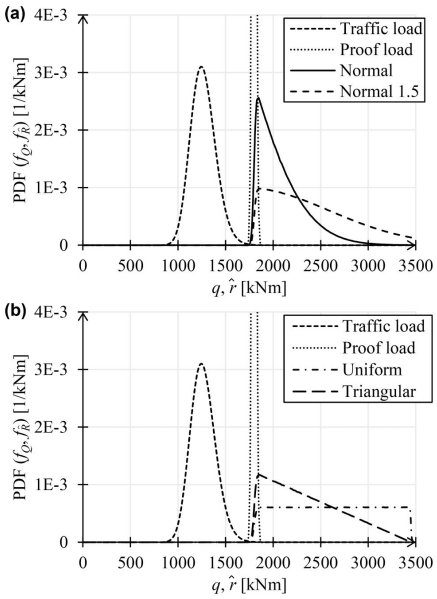


Figure 12. Traffic, proof load and resistance posterior distributions for (a) the normal priors and (b) the uniform and triangular priors.

Table 2. Results of reliability analyses with varying prior distributions.

Prior distribution	Parameter values	Annual reliability index (β) [–]
Normal	$\mu = 1150$ kNm, $V = 0.5$	3.95
Normal (factor 1.5)	$\mu = 1725$ kNm, $V = 0.5$	4.17
Uniform	$a = 0$, $b = 3450$ kNm	4.29
Triangular	$a = 0$, $b = 3450$ kNm, $c = 0$	4.14
Lower bound	-	3.43

The result of the lower bound reliability calculation using the limit state function of Equation (14) has also been included in Table 2. The somewhat lower reliability index is explained by the conservative lower bound assumption. In this light, the values obtained using a prior distribution should be viewed as best estimates of the reliability index. Compared to the time-dependent analysis (Section 5.2), similar reliability indices are found, with the exception of the lower bound calculation. However, the two outcomes are not directly comparable since resistance degradation and the trend in traffic loads is not included.

5.4. Time-dependent Bayesian analysis

The time-dependent and Bayesian analysis considered thus far may also be combined to incorporate (and visualise) time-dependent effects such as resistance deterioration and a trend in traffic loads. To express a lack of knowledge about the resistance the weak or informative normal prior distribution is used, with the mean value equal to 1.5 times the traffic load (Section 5.3.1). The time-dependent coefficients c_R and c_Q for this case study, Equations (17a) and (17b), are inserted in the limit state functions for Bayesian analysis, Equations (12) and (13), to result in:

$$Z = c_R\hat{R} - \max(c_Q C_{0Q} Q, Q_{PL}) \tag{18}$$

where again use is made of the remainder of the resistance available to resist variable loads ($\hat{R}$) and the model uncertainty assumption $\theta_E = \theta_{E,PL}$ (Section 5.3). For completeness, an overview of the random variables used in the time-dependent Bayesian analysis is provided in Table 3. The description of random variables provided for Table 1 provided in Section 5.2.2 also applies here.

The result of the time-dependent Bayesian analysis with two proof load tests is provided in Figure 13. The previous (informative) time-dependent analysis result (Section 5.2) is included as well for comparison. The conditional annual reliability is markedly lower for the weakly informative case when compared to the informative case – especially before a proof load test is performed. This is because the informative case has high reliability from the start, following from the assumed design.

The annual reliability in the year following the first proof load test is also lower than found in the simplified Bayesian analysis, namely $\beta = 3.67$ versus $\beta = 4.17$ (Section 5.3). However, in the simplified analysis the resistance deterioration and the trend in traffic load were not considered. Figure 13 also shows that an annual reliability index of about 2.5 just before the year 2020 is obtained considering proven strength by traffic loads alone – a value far from the reliability target $\beta = 4.0$. Also here the same remarks as provided in Section 5.2.3 regarding the unacceptably low reliability during a proof load test hold. To compensate for the lack of knowledge in the weakly informative case, the value of the target load in the proof load tests may be increased. To reach the desired reliability level in the period from 2020 to 2040 the required target loads are 2050 and 2150 kNm (For the informative case 1800 and 2000 kNm were required).

6. Discussion

In the literature review, international standards and the state-of-the-art on proof load testing have been briefly described. The body of literature related to proof load testing is much larger. In this article only the most relevant studies with regard to structural reliability have been highlighted and referred to. Although the various subjects identified as challenges are not new, their combined usage in the context of proof load testing deserves extra attention.

To illustrate the challenges, a case study of a hypothetical slab bridge was presented. With regard to the knowledge level, in the time-dependent example a scenario was depicted where the structural properties of the bridge, the traffic trend and the deterioration process are known to a large degree. Normally, this will not be the case. Especially the rate by which deterioration occurs will be difficult to establish. Suitable treatment of these uncertainties is critical. A clear need for stop criteria and their relation to structural

Table 3. Random variables used in the time-dependent Bayesian analysis.

Var.	Description	Distribution	Mean	COV	Auto-corr.
R	Resistance to variable loads	Normal	1725 kNm	0.5	1
C_{0Q}	Time-independent uncertainty of the variable load, including bias for dynamic load effect	Lognormal	1.1	0.1	1
Q	Load effect of the traffic load, annual maximum	Gumbel	1150 kNm	0.025	0
t_{R0}	Initiation time to deterioration	Lognormal	20 years	0.1	1
Δ_{cR}	Degradation per year	Lognormal	0.0025	0.1	1
c_{Q0}	Starting value of the trend	Lognormal	0.78	0.1	1
Δ_{cQ}	Increase of traffic load per year	Lognormal	0.004	0.1	1
Q_{PL}	Load effect of the proof load	Normal	1800 kNm and 2000 kNm	0.01	1

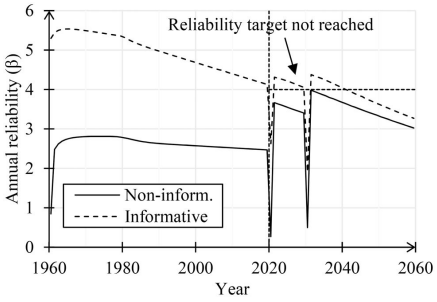


Figure 13. Conditional annual reliability of the weakly informative and the informative time-dependent analysis.

reliability emerges from the relatively large probability of failure during the proof load test.

The mathematical form of the limit state function for the proof load testing situation is not definitive; other formulations are possible as well. In the presented formulation, the model uncertainty of the load effect (θ_E) acts on both the traffic load effect and the load effect produced during the proof load test. In practice, the methods to calculate both effects will likely be similar (e.g. finite element analysis), but it does not guarantee full correlation. In the same way, the choice of the weakly informative prior distribution of the resistance, as used in Bayesian reliability updating, is not absolute. The case study demonstrates that the calculated reliability only differs slightly when the distribution type changes, and the same holds for its parameters. However, slight discrepancies will persist with the weakly informative prior approach.

A future framework for proof load testing should be flexible in such a way that it addresses the needs and reflects the knowledge level. As shown in the case study, the increased knowledge level can influence the level of proof loading needed to reach a certain target reliability. The framework could also consider the method by which a proof load test is performed. For example, less spatial uncertainty remains when driving over a bridge than when a single position is loaded. In the next steps towards a flexible framework for proof load testing, stop criteria in relation to capacity predictions and spatial uncertainty will be addressed.

7. Conclusions

With the suggested approach to the reliability assessment of existing reinforced structures through proof load testing a new framework can be developed that addresses existing challenges. An assessment based on annual reliability highlights the evolution of the structural reliability before, during and after the proof load test. The probabilistic methods that aid in addressing the challenges via time-dependent analysis and Bayesian updating have been described.

By adopting a flexible method, different types of information can be combined to assess the structural reliability through proof load testing. Future research should quantify how much benefit is obtained when considering various kinds of information. The case study of reinforced concrete slab bridge reveals that a significant difference exists in reliability predictions between the use of an informative and weakly informative prior distribution for the resistance. As a result, the required target loads in a proof load test are higher for the latter. Furthermore, uncertainty concerning multiple failure mechanisms and spatial variability can be addressed by judging the reliability on the system level rather than on the component level. This way, reservations regarding the assessment of shear capacity through proof load testing could be lifted.

The Bayesian reliability example considered the case in which very little is known about the resistance. Comparison of time-dependent analysis results between an informative and weakly informative prior distribution reveals that the effort devoted to obtaining an informative prior distribution leads to higher reliability estimates. This result of course depends on whether the original design was adequate and if the resistance deterioration was incorporated properly.

In addition, it was investigated to what extent the historic traffic load influences the reliability at the moment of proof load testing. If the traffic load before the year 2015 (date of measurements) is ignored, comparable outcomes are produced. This result established insensitivity towards historic traffic load (from 1960 to 2015), which is difficult to model. In the weakly informative Bayesian analysis it is found that the reliability before the proof load test is quite low, although the historic traffic load is included. It is concluded that the service-proven strength of past traffic load alone is not enough to reach the desired reliability level for a highway in the Netherlands.

List of symbols

a	Distance of reinforcement to surface
A_s	Cross-sectional area of the reinforcement
b	Width of the cross-section
$c_R(t)$	Deterioration of the resistance
$c_Q(t)$	Trend in traffic load
c_{Q0}	Starting value of the trend

C_{0Q}	Time-independent variability of traffic load
d	Effective depth of reinforcement
D	Data (Bayesian context)
e	Realisation of the load effect
E	Load effect
E_{PL}	Target proof load effect
f_c	Concrete compressive strength
f_y	Yield strength of the reinforcement
f_E	Probability density function of the load effect
f_R	Probability density function of the resistance
F_c	Force in the concrete
F_s	Force in the reinforcing steel
F_R	Cumulative probability function of the resistance
G_{DL}	Dead load
G_{SDL}	Super-imposed dead load
h	Height of the slab
H	Hypothesis (Bayesian context)
L	Span length
P(·)	Probability of (operator)
P_f	Probability of failure
Q	Traffic load
Q_{PL}	Target proof load
r	Realisation of the resistance
R	Resistance
$\tilde{R}$	Remainder of resistance available to resist variable loads
t	Time (year)
t_{R0}	Initiation time to deterioration
U	Standard normal random variable
x	Location of the neutral axis
Z	Limit state function
α	Influence coefficient
α_{cc}	Concrete compressive strength reduction factor
β	Reliability index
Δ_{cQ}	Starting value of the trend
λ	Factor for height of the compression block
ρ	Pearson correlation coefficient
θ_E	Model uncertainty of the load effect
$\theta_{E,PL}$	Model uncertainty of the target proof load effect
θ_R	Model uncertainty of the resistance

Acknowledgements

The authors wish to express their gratitude and sincere appreciation to the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat) for financing the research work. In addition, the fruitful discussions with M. Naaktgeboren of Rijkswaterstaat and M.A.N. Hendriks of Delft University of Technology have been of great help.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat).

References

AASHTO. (2018). *The manual for bridge evaluation.* Standard (3rd ed.). Washington, DC: AASHTO.

ACI. (2013a). *Code requirements for load testing of existing concrete structures. Standard, ACI 437.2M.* Farmington Hills, MI: American Concrete Institute.

ACI. (2013b). *Code requirenments for load testing of existing concrete structures (ACI 437.2M-13).* Farmington Hills, MI: American Concrete Institute.

Breitung, K. (1984). Asymptotic approximations for multinormal integrals. *Journal of Engineering Mechanics*, *110*(3), 357–366. doi:10.1061/(ASCE)0733-9399(1984)110:3(357)

Casas, J. R., & Gómez, J. D. (2013). Load rating of highway bridges by proof-loading. *KSCE Journal of Civil Engineering*, *17*(3), 556–567. doi:10.1007/s12205-013-0007-8

CEN. (2019). Eurocode 0: Basis of structural design. *Standard, EN 1990 + A1 + A1/C2:2019.* Brussels, Belgium: European Committee for Standardization.

Christensen, C. O., Zhang, F., Zarate Garnica, G. I., Lantsoght, E. O. L., Goltermann, P., & Smith, J. W. (2022). Identification of stop criteria for large-scale laboratory slab tests using digital image correlation and acoustic emission. *Infrastructures*, *7*(3), 36. doi:10.3390/infrastructures7030036

DAfStb. (2020). *DAfStb Guideline: Load tests on concrete structures (DAfStb-Richtlinie: Belastungsversuche an Betonbauwerken).* Berlin, Germany: Deutscher Ausschus für Stahlbeton.

De Vries, R., Lantsoght, E. O. L., & Steenbergen, R. D. J. M. (2021). *Case study proof loading in an annual reliability framework. Stevin Report 25.5-21-02.* Delft, The Netherlands: Delft University of Technology.

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M., & Fennis, S. A. A. M. (2022). Reliability assessment of existing reinforced concrete bridges and viaducts through proof load testing. *Proceedings of the 11th International Conference on Bridge Maintenance, Safety and Management (IABMAS)*, Barcelona, Spain.

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M., & Naaktgeboren, M. (2023). Proof load testing method by the American association of state highway and transportation officials and suggestions for improvement. *Transportation Research Record*, *2677*(2), 036119812311650. doi:10.1177/03611981231165026

Der Kiureghian, A. (2022). *Structural and system reliability.* Cambridge: Cambridge University Press.

Ditlevsen, O., & Madsen, H. O. (1996). *Structural reliability methods.* Hoboken, NJ: John Wiley & Sons.

Ditlevsen, O., & Vrouwenvelder, A. (1994). Objective" low informative priors for Bayesian inference from totally censored Gaussian data. *Structural Safety*, *16*(3), 175–188. doi:10.1016/0167-4730(94)00020-Q

Ellingwood, B. R. (1996). Reliability-based condition assessment and LRFD for existing structures. *Structural Safety*, *18*(2–3), 67–80. doi:10.1016/0167-4730(96)00006-9

Enright, M. P., & Frangopol, D. M. (1998). Probabilistic analysis of resistance degradation of reinforced concrete bridge beams under corrosion. *Engineering Structures*, *20*(11), 960–971. doi:10.1016/S0141-0296(97)00190-9

Ensink, S., Van der Veen, C., Hordijk, D. A., Lantsoght, E. O. L., Van der Ham, H. W. M., & De Boer, A. (2018). Full-size field test of prestressed concrete T-beam bridge. *Proceedings of the Structural Faults & Repair and European Bridge Conference*, Edinburgh.

Faber, M. H., Val, D. V., & Stewart, M. G. (2000). Proof load testing for bridge assessment and upgrading. *Engineering Structures*, *22*(12), 1677–1689. doi:10.1016/S0141-0296(99)00111-X

fib. (2016). *Partial factor methods for existing concrete structures.* Lausanne, Switzerland: fib.

Frangopol, D. M., Yang, D. Y., Lantsoght, E. O. L., & Steenbergen, R. D. J. M. (2019). Reliability-based analysis and lifecycle management of load tests. In E.O.L. Lantsoght (Ed.), *Load testing of bridges: Proof load testing and the future of load testing* (pp. 263–292). Boca Raton, FL: CRC Press.

Fussel, J. B. (1975). A Review of fault tree analysis with emphasis on limitations. *IFAC Proceedings Volumes*, *8*(1), 552–557.

Gollwitzer, S., & Rackwitz, R. (1983). Equivalent components in first-order system reliability. *Reliability Engineering*, *5*(2), 99–115. doi:10.1016/0143-8174(83)90024-0

Gong, C., & Zhou, W. (2017). Improvement of equivalent component approach for reliability analyses of series systems. *Structural Safety*, *68*, 65–72. doi:10.1016/j.strusafe.2017.06.001

Grigoriu, M., & Hall, W. B. (1984). Probabilistic models for proof load testing. *Journal of Structural Engineering*, *110*(2), 260–274. doi:10.1061/(ASCE)0733-9445(1984)110:2(260)

C

C

Harrewijn, T. L., Vergoossen, R. P. H., & Lantsoght, E. O. L. (2021). Reverse engineering of existing reinforced concrete slab bridges. *IABSE Congress, Christchurch 2021: Resilient Technologies for Sustainable Infrastructure*. Christchurch, New Zealand.

Hasofer, A. M., & Lind, N. C. (1974). Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division*, *100*(1), 111–121. doi:10.1061/JMCEA3.0001848

Hohenbichler, M., Gollwitzer, S., Kruse, W., & Rackwitz, R. (1987). New light on first- and second-order reliability methods. *Structural Safety*, *4*(4), 267–284. doi:10.1016/0167-4730(87)90002-6

Hohenbichler, M., & Rackwitz, R. (1982). First-order concepts in system reliability. *Structural Safety*, *1*(3), 177–188. doi:10.1016/0167-4730(82)90024-8

Jacinto, L., Neves, L. A. C., & Santos, L. O. (2016). Bayesian assessment of an existing bridge: A case study. *Structure and Infrastructure Engineering*, *12*(1), 61–77. doi:10.1080/15732479.2014.995105

JCSS. (2015). *Probabilistic model code*. Retrieved from https://www.jcss-lc.org/jcss-probabilistic-model-code/.

Kapoor, M., Christensen, C. O., Schmidt, J. W., Sørensen, J. D., & Thöns, S. (2023). Decision analytic approach for the reclassification of concrete bridges by using elastic limit information from proof loading. *Reliability Engineering & System Safety*, *232*, 109049. doi:10.1016/j.ress.2022.109049

Kapoor, M., Schmidt, J. W., Sørensen, J. D., & Thöns, S. (2019). A decision theoretic approach towards planning of proof load tests. *13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP)*. Seoul, South Korea.

Kapoor, M., Sørensen, J. D., Ghosh, S., & Thöns, S. (2021). Decision theoretic approach for identification of optimal proof load with sparse resistance information. *Proceedings of the 10th International Conference on Bridge Maintenance, Safety and Management (IABMAS 2020)*, Sapporo, Japan.

KIVI. (1938). Regulations for the design and for the manufacture and erection of steel bridges (Voorschrift voor het ontwerpen en voor het vervaardigen en opstellen van stalen bruggen). *Koninklijk Instituut Van Ingenieurs, Standard, N 1008*. The Netherlands: Dutch Standard.

KIVI. (1950). Reinforced concrete regulations (Gewapend-beton-voorschriften). *Koninklijk Instituut van Ingenieurs, Standard, N 1009*. The Netherlands: Dutch Standard.

Lantsoght, E. O. L. (2019a). *Load testing of bridges: Current practice and diagnostic load testing*. London: CRC Press/Balkema - Taylor & Francis Group.

Lantsoght, E. O. L. (2019b). *Load testing of bridges: Proof load testing and the future of load testing*. London: CRC Press/Balkema - Taylor & Francis Group.

Lantsoght, E. O. L., Van der Veen, C., De Boer, A., & Hordijk, D. A. (2017a). Proof load testing of reinforced concrete slab bridges in the Netherlands. *Proceedings of the 96th TRB Annual Meeting*, Washington, DC.

Lantsoght, E. O. L., van der Veen, C., de Boer, A., & Hordijk, D. A. (2017b). State-of-the-art on load testing of concrete bridges. *Engineering Structures*, *150*, 231–241. doi:10.1016/j.engstruct.2017.07.050

Lantsoght, E. O. L., Yang, Y., Van der Veen, C., Hordijk, D. A., & De Boer, A. (2019). Stop criteria for flexure for proof load testing of reinforced concrete structures. *Frontiers in Built Environment*, *5*, 47. doi:10.3389/fbuil.2019.00047

Lichtenstein, A. G. (1993). *Bridge rating through nondestructive load testing*. Technical Report, NCHRP Project 12-28(13)A. Washington, DC: TRB, National Research Council.

Lin, T. S., & Nowak, A. S. (1984). Proof loading and structural reliability. *Reliability Engineering*, *8*(2), 85–100. doi:10.1016/0143-8174(84)90057-X

Marx, S. (2019). Neufassung der DAfStb-Richtlinie: Belastungsversuche an betonbauwerken. *Presentation, 7. Jahrestagung, 60. Forschungskolloquium des DAfStb*.

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, *44*(247), 335–341. doi:10.1080/01621459.1949.10483310

Nishijima, K., & Faber, M. H. (2007). Bayesian approach to proof loading of quasi-identical multi-components structural systems. *Civil Engineering and Environmental Systems*, *24*(2), 111–121. doi:10.1080/10286600601159172

Rackwitz, R., & Schrupp, K. (1985). Quality control, proof testing and structural reliability. *Structural Safety*, *2*(3), 239–244. doi:10.1016/0167-4730(85)90030-X

Roscoe, K., Diermanse, F., & Vrouwenvelder, T. (2015). System reliability with correlated components: Accuracy of the equivalent planes method. *Structural Safety*, *57*, 53–64. doi:10.1016/j.strusafe.2015.07.006

RWS. (2013). Guidelines for assessing infrastructure - Assessment of the structural safety of an existing infrastructure for renovation, use and rejection (Richtlijnen beoordeling kunstwerken – Beoordeling van de constructieve veiligheid van een bestaand kunstwerk bij verbouw, gebruik en afkeur). *Rijkswaterstaat, Report, Version 1.1, 27 May*. The Netherlands: Dutch Guideline.

Schacht, G., Bolle, G., & Marx, S. (2019). Load testing of concrete building constructions. In E.O.L. Lantsoght (Eds.), *Load testing of bridges: Proof load testing and the future of load testing* (pp. 107–138). Boca Raton, FL: CRC Press.

Schmidt, J. W., Thöns, S., Kapoor, M., Christensen, C. O., Engelund, S., & Sørensen, J. D. (2020). Challenges related to probabilistic decision analysis for bridge testing and reclassification. *Frontiers in Built Environment*, *6*, 14. doi:10.3389/fbuil.2020.00014

Schneider, R. (2020). *Time-variant reliability of deteriorating structural systems conditional on inspection and monitoring data*. PhD Thesis, Technical University of Munich, Bundesanstalt für Materialforschung und -prüfung (BAM), München, Germany.

Smith, D. J. (2005). Understanding terms and jargon. In D. J. Smith (Ed.), *Reliability, maintainability and risk* (7th ed., pp. 11–23). Oxford: Butterworth-Heinemann.

Spaethe, G. (1994). The influence of load testing on the structural safety (Die Beeinflussung der Sicherheit eines Tragwerks durch Probebelastung). *Bauingenieur*, *69*, 459–468.

Steenbergen, R. D. J. M., & Vrouwenvelder, A. C. W. M. (2010). Safety philosophy for existing structures and partial factors for traffic loads on bridges. *Heron*, *55*(2), 123–139.

Vrouwenvelder, T., & Scholten, N. (2010). Assessment criteria for existing structures. *Structural Engineering International*, *20*(1), 62–65. doi:10.2749/101686610791555595

Vu, K. A. T., & Stewart, M. G. (2000). Structural reliability of concrete bridges including improved chloride-induced corrosion models. *Structural Safety*, *22*(4), 313–333. doi:10.1016/S0167-4730(00)00018-7

Wang, N., Ellingwood, B. R., & Zureick, A. H. (2011). Bridge rating using system reliability assessment. II: Improvements to bridge rating practices. *Journal of Bridge Engineering*, *16*(6), 863–871. doi:10.1061/(ASCE)BE.1943-5592.0000171

Yuefei, L., Dagang, L., & Xueping, F. (2014). Reliability updating and prediction of bridge structures based on proof loads and monitored data. *Construction and Building Materials*, *66*, 795–804. doi:10.1016/j.conbuildmat.2014.06.025

Zarate Garnica, G. I., & Lantsoght, E. O. L. (2021). Stop criteria for proof load testing of reinforced concrete structures. *Proceedings of the 2021 Session of the 13th fib International PhD Symposium in Civil Engineering*, 195–202.

Zarate Garnica, G. I., Lantsoght, E. O. L., & Yang, Y. (2022). Monitoring structural responses during load testing of reinforced concrete bridges: A review. *Structure and Infrastructure Engineering*, *18*(10–11), 1558–1580. doi:10.1080/15732479.2022.2063906

Zhang, W. H., Lu, D. G., Qin, J., Thöns, S., & Faber, M. H. (2021). Value of information analysis in civil and infrastructure engineering: A review. *Journal of Infrastructure Preservation and Resilience*, *2*(1), 16. doi:10.1186/s43065-021-00027-0

D

Engineering Structures article

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M., Hendriks, M. A. N. & Naaktgeboren, M. (2025b). Structural reliability updating on the basis of proof load testing and monitoring data. *Engineering Structures, 330*, 119863. https://doi.org/10.1016/j.engstruct.2025.119863

Engineering Structures 330 (2025) 119863

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Structural reliability updating on the basis of proof load testing and monitoring data

R. de Vries [a,b,*], E.O.L. Lantsoght [a,c], R.D.J.M. Steenbergen [a,d], M.A.N. Hendriks [a,e], M. Naaktgeboren [f]

[a] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands
[b] Reliable Structures, Netherlands Organization for Applied Scientific Research (TNO), Delft, the Netherlands
[c] College of Sciences and Engineering, Universidad San Francisco de Quito (USFQ), Quito, Ecuador
[d] Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium
[e] Department of Structural Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[f] Rijkswaterstaat, Ministry of Infrastructure and Water Management, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Bayesian updating
Laboratory testing
Load testing
Proof load
Reliability

ABSTRACT

As infrastructure continues to age and traffic levels intensify, there is a growing need for efficient methods to verify the reliability of many existing structures. Field testing offers the possibility to assess the current condition of a structure. Specifically, in a proof load test, substantial loads are applied to evaluate the structure's resistance to future loads that could compromise structural safety. However, to prevent excessive test loads and their potential damage, it is desirable to assess structural reliability by monitoring the response under more moderate loads. This study merges laboratory and in-situ testing results through a Bayesian update of the structural reliability after each successful load application. Two case studies are presented where laboratory testing on structurally similar elements and analytical modelling provide ample evidence to justify test load reductions of 20 % and 25 %. The proposed method offers a systematic framework to link the structure's response during testing to structural reliability and address the uncertainties in resistance, loads and measurements. Nonetheless, the representativeness of the data in terms of structural similarity and uncertainties related to measurements continue to be significant factors. Despite these challenges, incorporating monitoring data during proof load testing is expected to reduce target loads in most cases.

1. Introduction

Buildings and civil engineering works are expected to meet specific reliability requirements throughout their entire design life. Reliability assessment of an existing structure becomes relevant when the structure displays performance issues, the loads have significantly increased, or its original design life has passed. A design that may have been sufficient in the past may not be adequate today. Over time, degradation may have taken place, and the traffic loads have predominantly increased. Typically, assessing reliability requires in-depth information about the structural model, failure mechanisms, the description of loads and their combination. Moreover, for existing structures, assumptions regarding the uncertainties made in the past may no longer be true today as the knowledge about resistance and load models has evolved. Therefore, the original design reliability, based on prior knowledge, should be updated to reflect the current state of the structure. In addition, the reliability requirements for the design of new structures are higher than those used for the assessment of existing structures, following from an economic motive [1,2]. In this article, comparisons will be made of the required target loads at reliability levels suitable for the assessment of existing structures

Inspections, structural assessments, and maintenance are essential to ensure sufficiently reliable bridges and viaducts. As the infrastructure ages and endures increased traffic loads and environmental challenges, accurate reliability assessment methods are needed to address these evolving conditions. In case no signs of deterioration are present, the typical desk study may confirm sufficient reliability if there are no reasons to suspect internal damage. However, wear is often present, and it is difficult to tell if it impairs structural reliability. Fortunately, tests can be carried out on the structure to gather supplementary data. Tests on reinforced concrete structures commonly entail measuring the

* Corresponding author at: Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands.
E-mail address: rein.devries@tudelft.nl (R. de Vries).

https://doi.org/10.1016/j.engstruct.2025.119863
Received 23 September 2024; Received in revised form 17 January 2025; Accepted 2 February 2025
Available online 11 February 2025

Nomenclature

List of abbreviations

CDF	Cumulative distribution function
COV	Coefficient of variation
DIC	Digital image correlation
FEM	Finite element method
FORM	First-order reliability method
LHS	Latin hypercube sampling
LM1	Load model 1 (Eurocode)
MCMC	Markov chain Monte Carlo
MCS	Monte Carlo simulation
PLT	Proof load test
SORM	Second-order reliability method
WIM	Weigh-in-motion

List of symbols

C_{0Q}	Time-independent variability of traffic load
E	Load effect
E_{PL}	Load effect during proof load testing
E_{s}	Young's modulus of reinforcing steel
f_{c}	Concrete compressive strength
f_{y}	Yield stress of reinforcement
$F_X(\cdot)$	CDF of random variable X
G_{DL}	Dead load effect
G_{SDL}	Superimposed dead load effect
I	Indicator of performance
$m_{Q,\mathrm{PL}}$	Mean proof load effect
m_X	Mean of resistance ratio
M	Moment, mean (sample)
M_G	Moment due to dead load
n	Sample size
N	Number of samples (MCS)
P	Load of an individual axle in proof load test
$p_X(\cdot)$	Probability density function of X
P_{f}	Failure probability
Q	Time-variant part of traffic load
$Q_{\mathrm{k,WIM}}$	Characteristic traffic load following from WIM data
$Q_{\mathrm{k,LM1}}$	Characteristic traffic load following from Eurocode LM1
Q_{PL}	Target load (proof load)
R	Resistance
s_X	Standard deviation of resistance ratio
S	Standard deviation (sample)
T	Student's t-distributed random variable
U	Standard normally distributed random variable
Var(·)	Variance (operator)
V_X	Coefficient of variation of resistance ratio
$w_{\max}$	Maximum nominal crack width
$w_{\max,\mathrm{w}}$	Maximum weighted nominal crack width
$\mathbf{x}$	Data vector (Bayesian)
X	Random variable, resistance ratio (R/E_{PL})
Z	Limit state function
Z_{PL}	Limit state function for proof load testing
$\boldsymbol{\alpha}$	Influence coefficients vector
β	Reliability index
ε_{s}	Steel strain
$\boldsymbol{\theta}$	Model parameters vector (Bayesian)
θ_E	Model uncertainty of the load effect
$\theta_{E,\mathrm{PL}}$	Model uncertainty of the proof load effect
θ_R	Model uncertainty of the resistance
μ	Mean (population)
ν	Degrees of freedom (Student's t-distribution)
σ	Standard deviation (population)
$\Phi(\cdot)$	Standard normal CDF

geometry, drilling cores or scanning reinforcement. In some instances, elaborate setups that subject the structure to significant loads may be employed to test its resistance. The latter is referred to as *proof load testing*; by resisting a large load, the structure can prove to have sufficient resistance. However, applying the often large loads is resource-intensive and imposes a risk on the structure, equipment and personnel. To avoid excessively large loads, all relevant information about the structure should be considered, even when uncertain. Given this uncertainty, employing probabilistic techniques is necessary. Utilising all information, rather than often conservative design rules, can avoid excessive target loads. By rationally selecting the appropriate load level, proof load testing can effectively demonstrate the structure's resistance to anticipated future traffic loads [3–7].

The current research on the probabilistic substantiation of proof load testing aims to develop a comprehensive structural reliability updating framework [8]. This article presents a novel reliability updating method that integrates information from two distinct sources: (1) the survival of the applied load during the proof load test and (2) the data following from monitoring relevant indicators during the test, coupled to laboratory experiments giving the uncertain relation between these indicators and structural resistance [9]. Highly representative tests can significantly enhance the state of information, and thereby reduce resistance model uncertainty. Alternatively, tests on less representative specimens can be used, but will result in greater uncertainty. While the proof load is carefully increased in controlled increments to avoid unnecessary damage, the structural performance assessment may be based on indicators such as displacements and crack widths. These indicators, while not immediately indicative of overall structural health, are interpreted in the light of structural behaviour observed in laboratory experiments on structural elements similar to those present in the proof loaded bridge. For example, strains may be interpreted via sectional analysis to identify a critical value. In this way, stop criteria can prevent unwanted damage [10], but the indicator value may also provide valuable information on structural performance.

To probabilistically interpret the information from indicators, first the theoretical background on structural reliability, statistical inference, and their relation to proof load testing is described. Then, the reliability updating method on the basis of proof load testing and monitoring data is presented. To illustrate the method's application, two case studies utilising laboratory measurements and analytical modelling are provided. The article rounds off with a discussion of the results, challenges, and conclusions.

2. Theoretical background

2.1. Structural reliability

The proposed method (Section 3) relies on the principle of structural reliability updating, which requires reliability assessment. A structural reliability assessment is based on a probabilistic model that includes a limit state function and the definition of random variables describing the load and resistance parameters and the modelling uncertainties. Reliability methods are used to calculate the reliability (or failure probability) of a structure or a structural component. The limit state function plays a central role as it expresses the boundary between safe and unsafe combinations of resistance and load effect. Negative values of the limit state function indicate failure, irrespective of the magnitude. An example of a typical limit state function involving just two random

variables is:

$$Z = g(\mathbf{X}) = g(R, E) = R - E \tag{1}$$

where R is the resistance and E is the load effect. The load effect may comprise the bending moment, shear force, axial load, and so on. Often, the combined effects of multiple loads are considered [11–13]. The state of information concerning the resistance and load effect varies between structures. Various sources of information can be utilised, ranging from literature to in-situ material testing and site-specific traffic data [8]. The Eurocode standards also allow for design assisted by testing and guidance is provided regarding the sampling process and statistical post-processing [14].

Commonly, the number of random variables is larger than just two, and the limit state function is more difficult to compute. There may be a very large number of random variables, and instead of having expressions in closed or analytic form available, complex finite element method (FEM) models may be needed to evaluate the limit state functions. For this reason, different reliability computation methods have been developed, and all have their pros and cons depending on the application. The calculation procedure for the proposed method (Section 3.5) makes use of the Markov chain Monte Carlo (MCMC) sampling [15], which in turn is based on Monte Carlo simulation (MCS). MCS is a straightforward method applicable to many reliability problems but is computationally expensive [16]. In an MCS, the random variables present in the problem formulation ($\mathbf{X}$) are repeatedly sampled and may be used to evaluate the limit state function. The probability of failure is obtained by calculating the fraction of failures that occur:

$$P_f = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[g(\mathbf{x}_i) < 0] \tag{2}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function; it is 1 when the condition within brackets is true and 0 otherwise. The number of samples is denoted by N, and the random vector containing the values of each sample is $\mathbf{x}_i$. The corresponding reliability index may be calculated via:

$$\beta = \Phi^{-1}(1 - P_f) = -\Phi^{-1}(P_f) \tag{3}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function (CDF) (i.e., mean = 0, standard deviation = 1). Reliability requirements are commonly formulated in terms of the reliability index and a reference period to account for the time-dependent nature of random processes. One way to unify reliability requirements for design and assessment is through the adoption of annual reliability targets [17,18].

When the failure probability is small, many samples are required to estimate the reliability index accurately. Improved sampling methods [19–23] or other reliability methods are beneficial in such circumstances. After simplification, common reliability methods can be used in the proposed reliability updating method (Section 3.5). For example, using the first-order reliability method (FORM), a computationally efficient method with easily understandable output: the reliability index (β) and influence coefficients ($\boldsymbol{\alpha}$), indicating the relative importance of the random variables [24,25]. Further improvement of the reliability index is possible with the second-order reliability method (SORM) [26,27] which utilises the second-order derivatives of the limit state function in the design point. After eigenvalue analysis or calculating determinants [28], they yield a correction factor to the FORM failure probability. Slightly different versions of the same idea were introduced by various authors, offering a small increase in accuracy [29,30].

2.2. Statistical inference

2.2.1. Principles

Deriving the statistical descriptions of the random variables within the limit state function is necessary to enable the reliability assessment. These random variables represent uncertainties in material properties, loads, and modelling approaches. When tests are performed in the laboratory, the number of specimens is usually limited due to cost, time, and material availability constraints. Therefore, the number of tests and resulting data points is usually small and cannot be expected to capture the inherent variability fully. This also applies to calculating the resistance ratio distribution based on laboratory tests (Section 3.3.1). Both frequentist and Bayesian approaches can be used to infer the statistical descriptions and account for the limited number of tests. The frequentist approach solely uses the sample statistics of the observed data. The prediction distribution allows for the inclusion of statistical uncertainty due to small sample sizes and known and unknown standard deviations. On the other hand, the Bayesian approach also allows for incorporating prior knowledge to deliver posterior distributions that incorporate both data and subjective knowledge. The calculation of the posterior probability distributions typically requires numerical methods, such as Markov chain Monte Carlo (MCMC) sampling [15].

2.2.2. Prediction distribution

Given n observations of an unknown random variable X, the prediction distribution describes the probability distribution of the next-to-be-observed value X_{n+1}. Generally, two situations are distinguished: one where the standard deviation is known, and the other where it must be estimated from the data. If the observations X_i come from the same normally distributed population, are independent and identically distributed (i.i.d.), it follows that [31,32]:

$$U = \frac{X_{n+1} - M}{\sigma\sqrt{1 + 1/n}} \sim \mathcal{N}(0, 1) \tag{4}$$

where U is a standard normally distributed random variable, $M = (X_1 + X_2 + \cdots + X_n)/n$ is the sample mean, and σ is the population standard deviation. Only when data is observed, random variable M becomes a realisation of the sample mean (denoted with lowercase m). The normalising term follows from considering the variance of the difference in the numerator, i.e. $\mathrm{Var}(X_{n+1} - M) = \mathrm{Var}(X_{n+1}) + \mathrm{Var}(M) = \sigma^2 + \sigma^2/n = \sigma^2(1 + 1/n)$. Taking its square root gives the denominator in Eq. (4). Intuitively, it may be understood as the standard deviation following from random variable X_{n+1} and the, typically smaller, standard deviation of the sample mean of values 1 to n. Solving Eq. (4) for X_{n+1} gives:

$$X_{n+1} = M + U\sigma\sqrt{1 + 1/n} \tag{5}$$

where the right-hand side may be interpreted as the prediction distribution of X and follows a normal distribution. It should be realised that X_{n+1} will not actually follow the prediction distribution, but Eq. (5) allows for incorporating the uncertainty about the mean, similar to the posterior predictive distribution in Bayesian inference (Section 2.2.3). The 'penalty' incurred by estimating the population mean with the sample mean is contained in the increased standard deviation and diminishes with an increasing number of observations n.

In case the standard deviation of X is not known and needs to be inferred from the data as well, Student's t-distribution emerges [33]. The t-distribution is wider than the normal distribution, reflecting the more significant uncertainty when the standard deviation is unknown. If, once again, the i.i.d. observations X_i from the same normally distributed population are considered, it follows that [31]:

$$T = \frac{X_{n+1} - M}{S\sqrt{1 + 1/n}} \sim t_{\nu=n-1} \tag{6}$$

where T is a (standard) t-distributed random variable with $\nu = n - 1$ degrees of freedom and S is the sample standard deviation including Bessel's bias correction for the variance:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2} \tag{7}$$

The same normalisation consideration holds for the denominator in Eq. (6) as in Eq. (4). Solving Eq. (6) for X_{n+1} gives:

$$X_{n+1} = M + TS\sqrt{1+1/n} \tag{8}$$

where the right-hand side may be interpreted as the prediction distribution of X and follows a shifted and scaled t-distribution. Compared to Eq. (5), another 'penalty' is introduced by using the random variable T instead of U, because the t-distribution is typically wider. In some cases, an intermediate approach is followed where n is artificially increased when (prior) information suggests a similar standard deviation as calculated from the dataset. In addition, knowledge about the type of distribution may also be incorporated. It is typical to assume a lognormal distribution for material properties and, thus, apply the above procedure to log-transformed values [14,34].

2.2.3. Bayesian inference

The Bayesian inference process enables the integration of prior knowledge to improve the statistical description of data. It is a systematic process in which *prior* beliefs are updated according to the available information (or evidence, or data), resulting in *posterior* beliefs. In a Bayesian inference context, usage of the t-distribution is equivalent to adopting a non-informative prior for the parameters of the normal distribution. However, in many cases, a non-informative prior is too generic and does not adequately represent prior knowledge or constraints known about the parameters, leading to overly conservative posterior estimates. In such scenarios, incorporating more informative priors by leveraging expert judgement and historical data leads to more realistic outcomes [35]. Bayesian inference can be used to update beliefs about important parameters in the structural reliability analysis. The reliability of the structure may be re-evaluated each time new information becomes available, resulting in Bayesian reliability updating [36–38]. In a sequential updating, scheme the posterior distribution following from the previously acquired data is used as a prior for the next iteration [39,40].

In Bayesian inference, the distribution of a random variable X itself may be updated, or the parameters of its assumed model. In the latter hierarchical model, the distribution parameters are modelled as random variables with specific distributions as well. The prior distributions may then be specified by providing values for the hyperparameters, e.g. the mean and standard deviation of the mean of X. In this scenario, the model parameters are the mean and standard deviation of X, collected in $\boldsymbol{\theta} = (\mu_X, \sigma_X)$, and may be updated through Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}} \tag{9}$$

where $\boldsymbol{\theta}$ is a vector containing the model parameters (random variables), $\mathbf{x}$ is a vector containing the data, $p(\boldsymbol{\theta}\,|\,\mathbf{x})$ is the posterior distribution, $p(\mathbf{x}\,|\,\boldsymbol{\theta})$ is the likelihood of observing the data given the model parameters, $p(\boldsymbol{\theta})$ is the prior distribution, and the denominator is called the marginal likelihood and acts a normalising constant. If prior information about the parameters to be inferred is not available, a non-informative prior may be used. In the typical case with the mean and the standard deviation as model parameters, the non-informative prior is obtained by the following (improper) probability density functions [41]:

$$p(\mu_X) = 1 \tag{10a}$$

$$p(\sigma_X) = 1/\sigma_X \tag{10b}$$

from which the prior distribution follows as $p(\boldsymbol{\theta}) = p(\mu_X)\,p(\sigma_X) = 1/\sigma_X$. Because the resulting prior distribution is improper, it cannot be sampled directly. Instead, a calculation procedure is required that draws its random values in a different way, such as Markov chain Monte Carlo (MCMC) [42,43] or Importance Sampling [19]. When the often multi-dimensional posterior distribution $p(\boldsymbol{\theta}\,|\,\mathbf{x})$ has been obtained, it can be used to update the probability distribution of X itself, i.e. the posterior predictive distribution or Bayes' distribution [44]:

$$p_X^{\text{Bayes}}(x) = \int p_X(x|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{x})\,\mathrm{d}\boldsymbol{\theta} \tag{11a}$$

$$F_X^{\text{Bayes}}(x) = \int F_X(x|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{x})\,\mathrm{d}\boldsymbol{\theta} \tag{11b}$$

where $p_X(\cdot)$ denotes the probability density function and $F_X(\cdot)$ the cumulative distribution function of random variable X.

2.3. Reliability following from proof load testing

In proof load testing, the so-called *target load* plays a central role, as it is specifically chosen to simulate today's and future usage conditions and account for a degree of uncertainty. The magnitude of the target load controls the desired reliability of the structure, as higher reliability demands require higher loads during testing. If a structure withstands the target load without signs of distress [45], it is deemed sufficiently reliable for continued operation after the test. Selecting the target load based on reliability considerations may be done in different ways, depending on how much information is available about the structure. For a reliability assessment of a bridge or viaduct, at minimum, a statistical description of expected traffic loads is required. By assuming that the resistance of the tested bridge or viaduct is at least equal to the permanent load effects and the target load effect ($R \geq G + Q_{\text{PL}}$), the limit state function, including model uncertainties, may be written as [5]:

$$Z = \theta_{E,\text{PL}} Q_{\text{PL}} - \theta_E C_{0Q} Q \tag{12}$$

where $\theta_{\text{E,PL}}$ is the model uncertainty of the load effect pertaining to the proof load testing situation (Section 3.2), Q_{PL} is the target load, θ_E is the model uncertainty of the load effect for the regular traffic load situation (correlated with $\theta_{\text{E,PL}}$), C_{0Q} accounts for the time-independent variability of the traffic load, and Q is the time-variant part of the traffic load. Evaluating the limit state function in Eq. (12) is referred to as the *lower-bound approach*, as it provides the most conservative estimate of the posterior resistance distribution [46]. A comparison with target loads obtained through this relatively straightforward method is presented in the case studies (Sections 4 and 5).

Alternatively, the distribution function of the resistance may be explicitly considered. This procedure effectively leads to truncating the left tail of the distribution to exclude the possibility that the resistance is lower than the load effect produced during the test [47]. However, this truncation is not abrupt but rather gradual, owing to some uncertainty about the actual load effect created by the applied load [48]. In principle, this resistance relates to the specific loading position and method of application. On a structural level, the effect of the applied load on a particular component or cross-section is more valuable. Therefore, the limit state, in principle, considers a resistance and load effect, not the applied load itself. The update of the resistance distribution may also be achieved via the application of Bayes' theorem, and an indicator likelihood function providing the value 0 for resistances lower than the target load and 1 otherwise. With the proof load effect described by a random variable, this procedure also results in an appropriate posterior distribution for the resistance. The prior distribution may then be formulated using a mean value based on the mean annual traffic load and a relatively large coefficient of variation to reflect the large degree of uncertainty about the resistance. When a resistance distribution is available, possibly updated by in-service proven strength, it is also possible to evaluate the reliability during the proof load testing situation [46].

D

3. Reliability updating method

3.1. Reliability updating using two information sources

After each load increment in the proof load test, the estimation of the resistance on the basis of measurements forms the first source of information (Fig. 1, point 1). The proposed method relates the observed in-situ response, through the use of indicator values, to the response observed in laboratory tests on similar structural elements or derived from suitable analytical models. The indicator value is measured in situ and on representative specimens in the lab. For example, strains derived from horizontal displacement measurements at the bottom of a beam or slab can serve as critical indicators for structural performance. By utilising the same indicators in both laboratory and in-situ testing, it is possible to determine the mean and standard deviation of the structure's resistance. Instead of a mean and standard deviation, the prediction distribution of R may be established via other statistical approaches, including Bayesian inference (Section 2.2).

If a structure can withstand a specific proof load, it also means that its resistance (R) is equal to or greater than the load effect during the test (Section 2.3). Considering the uncertainty of the proof load test, the gradual truncation of the resistance is the second source of information (Fig. 1, point 2). The information from the two sources is processed in the presented order and allows for variable load increments to determine the reliability of the structure. Generally, as the applied load increases, the structural reliability also tends to increase. A flowchart outlining the steps in the proposed proof load testing assessment method is provided in Fig. 2.

3.2. Probabilistic model for reliability updating

Structural reliability may be assessed by evaluating a limit state function (Section 2.1). The primary limit state considers the situation in which the regular traffic loads act on the bridge. The proposed function aligns with the guidelines from the Probabilistic Model Code [34] and fib Bulletin 80 [49]:

$$Z = \theta_R R - \theta_E(G_{\mathrm{DL}} + G_{\mathrm{SDL}} + C_{0Q}Q) \tag{13}$$

where θ_R is the uncertainty associated with resistance calculation, R is the resistance, θ_E is the model uncertainty of the load effect calculation, G_{DL} is the dead load effect, G_{SDL} is the superimposed dead load effect, C_{0Q} is the time-invariant part of the live load effect, and Q is the time-variant part of the live load effect (i.e. traffic load effect). The reliability of the bridge prior to proof load testing could be evaluated if distributions were assigned to all random variables, including R. Because of the low information state, the distribution of R would incorporate a large variability and thus result in low reliability. Instead of employing a conventional structural resistance model, the resistance R will be estimated by combining in-situ measurements with insights gained from laboratory experiments or analytical modelling. This also affects how the variability of θ_R, the uncertainty associated with resistance calculation, should be quantified (Section 3.4).

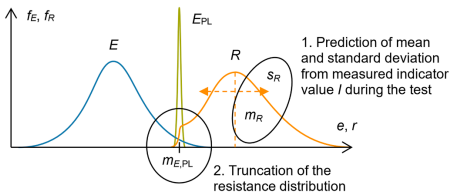


Fig. 1. General principle of updating the resistance distribution from two sources of information.

Instead of directly assigning a distribution to resistance R, it will be expressed as a function of the proof load effect via the factor X. Each time the structure withstands a new load level in a proof load test, the distribution function of R can be updated to reflect the information obtained via measurements, in addition to the truncation (Section 3.1). The revised distribution of R can be expressed as the product of a resistance ratio X and the load effect produced during the last successful load test cycle ($R = X\,E_{\mathrm{PL}}$). Given a certain observed indicator value I, the distribution of X may be obtained – as determined before proof load testing (Section 3.3). During a proof load test, the traffic load will be absent, and instead, the test load is present. Thus, the limit state function describing the proof load testing situation is:

$$Z_{\mathrm{PL}} = \theta_R X E_{\mathrm{PL}} - E_{\mathrm{PL}} = (\theta_R X - 1)E_{\mathrm{PL}} \propto \theta_R X - 1 \tag{14}$$

where $E_{\mathrm{PL}} = \theta_E(G_{\mathrm{DL}} + G_{\mathrm{SDL}}) + \theta_{E,\mathrm{PL}}Q_{\mathrm{PL}}$ with Q_{PL} denoting the load effect created by the proof load and $\theta_{E,\mathrm{PL}}$ its corresponding model uncertainty. By assuming that the proof load is withstood, it directly follows that $Z_{\mathrm{PL}} > 0$ and thus $\theta_R X > 1$. The conditionality can be satisfied in a Bayesian updating process by obtaining the joint posterior distribution of $\boldsymbol{\theta} = (\theta_R, X)$ as:

$$p(\boldsymbol{\theta}|Z_{\mathrm{PL}} > 0) \propto p(Z_{\mathrm{PL}} > 0|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{15}$$

where the likelihood $p(Z_{\mathrm{PL}} > 0 \mid \boldsymbol{\theta})$ acts as an indicator function, or potential [50,51], and $p(\boldsymbol{\theta})$ is the prior probability. In a Monte Carlo simulation, this process involves differentiating between samples that either withstand or fail the proof load test (Section 3.5). After this update, the marginal distributions should not be sampled independently because the interdependence of variable combinations significantly influences the outcomes. Returning to the original traffic load situation, Eq. (13), with $R = X\,E_{\mathrm{PL}}$ and denoting the updated variables as $\theta_R{}'$ and X' gives:

$$Z = \theta'_R X'\left[\theta_E\left(G_{\mathrm{DL}} + G_{\mathrm{SDL}}\right) + \theta_{E,\mathrm{PL}}Q_{\mathrm{PL}}\right] - \theta_E\left(G_{\mathrm{DL}} + G_{\mathrm{SDL}} + C_{0Q}Q\right) \tag{16}$$

and can be used to evaluate the structural reliability after a successful test cycle.

The chosen probabilistic formulation accounts for model uncertainties in both the load effect caused by regular loads (θ_E) and the load effect specific to the load applied in the proof load test ($\theta_{E,\mathrm{PL}}$). The statistical characterisation of these uncertainties is specific to the application. In particular, the model uncertainty pertaining to the proof load can address various factors, such as the method of load application, the number and configuration of tested positions and lanes, and the considered failure mode. It should be realised that the model uncertainties θ_E and $\theta_{E,\mathrm{PL}}$ are likely correlated, given that the same mathematical principles and models are employed to calculate both load effects.

3.3. Distribution of the resistance ratio (X)

3.3.1. Using laboratory test data

When laboratory data are available, the relationship between measurements or indicator values and the resistance ratio ($X = R/E$) can be inferred from these tests. The laboratory tests should be conducted on similar elements and in a configuration comparable to the in-situ proof load test. The laboratory measurements are processed for each load step to analyse the resistance ratio distribution as the indicator values increase. Each specimen has a resistance (R) that corresponds to the load effect at the moment that the limit state is reached (failure). During each load step, the load effect (E) can be calculated, resulting in a corresponding resistance ratio. Typically, an estimation of the self-weight is required to calculate the load effect from both permanent and applied loads (E). This procedure results in a resistance ratio versus indicator value (I) curve for each specimen (for example, see Fig. 5 in Section 4.2).

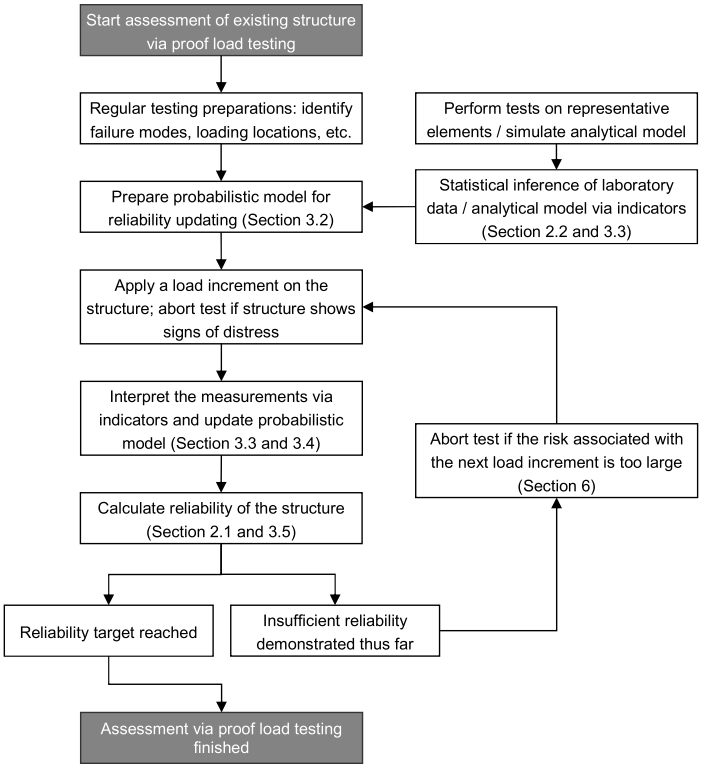


Fig. 2. Flowchart outlining the steps in the proposed proof load testing assessment method.

The resistance ratio is modelled as a random variable due to the inherent uncertainties in both the resistance and the indicator value at a given load level. Moreover, the uncertainty typically decreases as the indicator's value increases. To reduce noise and erratic responses, the maximum indicator value observed up to each load step is used. The same post-processing practice should be followed during in-situ tests for consistency. When the specimen has failed, the resistance ratio is 1, indicating that the resistance is at least equal to the current load effect but not higher.

Once the resistance ratio curves have been obtained for the laboratory tests, statistical post-processing may be performed to infer the statistical description for a range of indicator values. Given a certain indicator value, the data points (X_i) are obtained as the resistance ratios from each specimen. Interpolation of the resistance ratio curves is required to obtain intermediate values. Then, the data points can be analysed using established sample testing methodologies (Section 2.2). Because the prediction distribution of X is used, the actual variation is more significant than indicated by the standard deviation alone. Bayesian inference can also be employed instead of the prediction equation. However, this may be impractical since the inference process needs to be repeated for a range of indicator values. A better approach would also include the trend of model parameters with varying indicator values combined with a nuanced treatment of measurement error (Section 3.4). An application of the prediction equation and an exponential model for the trend in the resistance ratio is provided in Section 4.2.

3.3.2. *Using an analytical model*

In cases where laboratory measurements are unavailable, computer simulations can be utilised as an alternative. Instead of calculating the typical design resistance, the aim is to determine the resistance ratio distribution, which cannot be directly obtained through conventional methods. A regular resistance model is developed, but the parameters are random variables. To obtain the statistical distribution of the resistance ratio, it is necessary to integrate over all random variables included in the resistance model. There are several methods to accomplish this integration. The most straightforward method is Monte Carlo simulation (MCS) [16]. However, often a complex numerical (FEM) model is used for the resistance calculation. In these cases, the application of Latin hypercube sampling (LHS) can be beneficial as this method allows for a more efficient representation of the random space with fewer samples [23,52].

By using LHS, numerous resistance ratio versus measurement value (indicator) curves are generated. These curves may then be statistically

analysed for a range of indicator values. By assuming a normal distribution, a mean and standard deviation curve of the resistance ratio versus indicator value can be produced. Because the random space is directly integrated, there is no need to account for statistical uncertainty. Only if the number of simulations is small (e.g. fewer than 30) an approach similar to the post-processing of laboratory experiments should be followed (Section 3.3.1). The additional uncertainty inherent in analytical modelling, as compared to physical testing, can be reflected in a larger coefficient of variation for the resistance model uncertainty (θ_R; see Section 3.4). An application of this approach, in which analytical modelling of the bending resistance is performed, is presented in Section 5.

3.4. Resistance model and measurement uncertainty

In the described probabilistic framework, θ_R accounts for uncertainty in the resistance calculation introduced by the proposed method. This uncertainty is small when the laboratory specimens closely resemble the actual structure or when the mechanical model has been validated for accuracy. Conversely, significant uncertainty is expected if the structures differ substantially, the mechanical model is overly simplistic, measurement errors are significant, or there are inconsistencies in data post-processing. Practical applications of the proposed method will provide valuable insights into appropriate model uncertainty values as the application in real-world scenarios can highlight the difference between laboratory and in-situ observations [53].

When dealing with a small number of laboratory tests, statistical methods like Student's *t*-distribution or Bayesian inference effectively account for the inherent statistical uncertainty and variability in the data. The statistical uncertainty would be incorporated directly in the resistance R, thus separated from the modelling uncertainty θ_R. Student's *t*-distribution is particularly useful when the sample size is small, and the standard deviation is unknown. It is wider than the normal distribution, reflecting the increased uncertainty that comes with fewer laboratory measurements. Bayesian inference, on the other hand, offers a flexible way of incorporating prior knowledge and can account for measurement noise as well. To capture the noise, the data points may be viewed under an assumed distribution for the measurement error by adjusting the likelihood calculation. Assuming the likelihood model (X) and noise (ε) are both normally distributed and independent, their combined variance may be used to define a substitute random variable. The likelihood of observing the data points may then be calculated using the probability density function of the substitute random variable rather than X directly [54,55].

Measurement errors can significantly influence results, especially with small values, such as minor crack widths or small strains. As the magnitude of the values increases, the relative impact of measurement errors typically decreases. In addition, the error also depends on the parameter being measured and whether it can be directly measured or must be inferred. Typically, larger measurement errors are anticipated when estimating crack widths using digital image correlation (DIC) compared to direct strain measurements. Measuring strains on a concrete surface is more susceptible to errors than taking strain measurements directly on the reinforcement. The moderate load values in the proposed method will typically result in small indicator values, and thus, the large uncertainty should be appropriately accounted for. While Bayesian methods are well-suited for treating noise, other analytical approaches can also enhance model accuracy by incorporating physically expected trends with increasing indicator values, thereby providing a nuanced understanding of the data (see, for instance, Section 4.2).

3.5. Calculation procedure

In order to compute the structural reliability after a successful proof load test, the knowledge of surviving the applied load needs to be incorporated, see Eq. (15). To obtain the posterior distribution $p(\boldsymbol{\theta} \mid Z_{\mathrm{PL}} > 0)$, several calculation methods can be employed. In the direct Bayesian Monte Carlo (BMC) method, the prior distributions are directly sampled. Each sample is used in the simulation to determine if the random values ($\boldsymbol{\theta}$) result in the survival of the proof load test (likelihood evaluation). During the simulation, all parameter sets that produce the desired outcome are stored and collectively describe the posterior distribution [56,57]. In Markov chain Monte Carlo (MCMC), a Markov chain is constructed to obtain samples from the posterior distribution. Algorithms like Metropolis-Hastings [42,43] and Gibbs sampling [58] can generate such sequences (Section 2.2.3). In this work, the MCMC method is adopted because of its versatility. Owing to its Monte Carlo nature, the chain's current state is directly used to evaluate the reliability of the structure with the posterior distribution. For each chain state, $\boldsymbol{\theta} = (\theta_R, X)$, the sample is supplemented by random realizations of the remaining random variables to evaluate the limit state function, i.e. Eq. (16).

A computationally attractive alternative is using the SORM (Section 2.1), which can account for the non-linearity present in the limit state function. Because the survival condition, $Z_{\mathrm{PL}} > 0$, cannot be incorporated directly, an approximation must be made. This is achieved by introducing a substitute random variable $Y = \theta_R X$, which follows Student's *t*-distribution when X is described by a *t*-distribution and a lognormal distribution otherwise. The variances of the original random variables may be combined such that $\mathrm{Var}(Y) = \mathrm{Var}(\theta_R) + \mathrm{Var}(X)$. Then, the distribution of Y is left-truncated to the set $[1, \infty)$ to impose the survival of the proof load test. Applying this alternative procedure results in an error in the reliability index of approximately 0.1 within the range of common target values (as experienced in the case studies, Sections 4 and 5).

4. Shear resistance assessment supported by laboratory tests

4.1. Description

In order to illustrate the practical application of the method proposed in Section 3, the reliability of a hypothetical shear-critical reinforced concrete slab bridge is considered. The case exemplifies older Dutch slab bridges that lack shear reinforcement. For simplicity, the slab is designed to match the exact width of a single traffic lane (Fig. 3). Normally, a slab bridge would include several lanes, along with sidewalks and railings. The single-lane slab bridge, assumed to experience heavy traffic primarily from trucks, represents a relatively conservative scenario.

Deep beams representing sections of such a slab were tested in the laboratory to evaluate their shear resistance. The reliability of the bridge under consideration can be assessed using the resistance data and the measurements obtained from load tests. However, because the bridge is fictional and no actual in-situ measurements were performed, these values must be estimated to demonstrate the application of the proposed method. The laboratory measurements employed in this case study were initially designed to examine the shear behaviour of reinforced concrete beams lacking shear reinforcement [59]. The tests are a continuation of the study into the parameters that play a role in the transition between flexural and shear failure of reinforced concrete beams without reinforcement [60]. For this case study, H-variants (H121, H401, H403, H404, H602) from the test series were selected because their dimensions correspond to those of the studied concrete slab. The strips, or deep beams, tested in the laboratory had a length of 9 m, a width of 0.3 m and a height of 1.2 m. They were subjected to a load via a single jack positioned near the midpoint of the span, leading to shear failure near the supports (see Fig. 4). Given the specified lane width of 3.6 m, the slab comprises $3.6 / 0.3 = 12$ strips.

Because the experiments have already been conducted, providing a resistance distribution based on the test results (i.e. five V_u values) would be directly possible. Supplemented by the knowledge that the

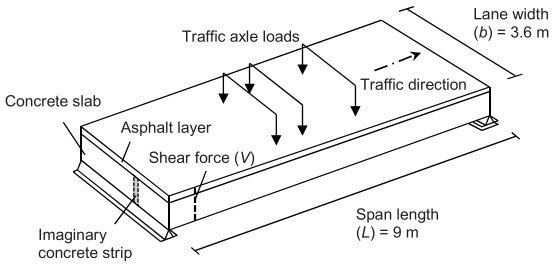
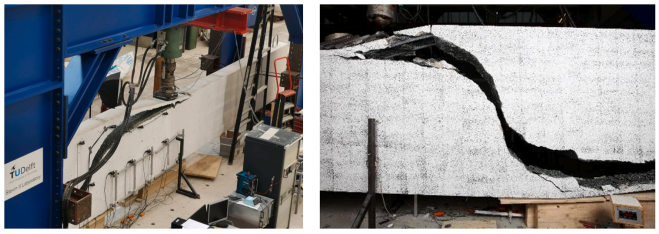


Fig. 3. Hypothetical reinforced concrete slab used as a case study [9].



(a) **(b)**

Fig. 4. Photos of (a) the test set-up and (b) the typical crack pattern at failure [61].

structure survives the proof load, this information would give an alternative strategy for reliability updating. However, this alternative procedure does not account for the in-situ observations. By using these observations, a distinction can be made between structures performing well, exhibiting small crack widths, and those performing poorly, showing larger crack widths. In addition, the use of a resistance ratio allows laboratory results from similar structural elements to be applied to other structures that are not entirely identical. However, the validity of this approach and the increase in resistance model uncertainty become relevant factors (Section 3.4).

4.2. Laboratory data post-processing

The digital image correlation (DIC) method was used to monitor and analyse the development of cracks in the concrete during the tests [62]. DIC allows for the determination of the nominal crack width at various locations. The nominal crack width results from the consolidation of several smaller cracks that are closely clustered, forming a single significant crack. The concept of virtual gauge length which is used in DIC, requires the area captured in photographs to be sufficiently large to facilitate the nominal crack width calculations. In this study, the optimal gauge length was found to be $0.8\,d$, where d represents the depth from the top of the beam to the centre of the longitudinal reinforcement [63]. At each selected location, the nominal crack width is calculated at the level of the reinforcement. To assign greater importance to the cracks near the supports, a weighted crack width calculation is introduced, multiplying the nominal crack width by the factor Vd/M, where V and M refer to the shear force and bending moment at the specific location, respectively. Both the maximum nominal crack width ($w_{\max}$) and the maximum weighted nominal crack width ($w_{\max,w}$) are computed. These maximum values refer to the (weighted) crack widths occurring at any point along the beam's length. Finally, the resistance ratio $X = R/E_{PL} = V_u/V$ for each indicator value was plotted against each load step (Fig. 5).

The data was post-processed to derive the sample mean and standard deviation for each indicator value (Fig. 6). It may be observed that the standard deviation is generally smaller for the maximum *weighted* nominal crack width indicator ($w_{\max,w}$), suggesting it as the preferable metric for subsequent modelling and field testing. Data points where the weighted nominal crack is less than 0.08 mm displayed noticeably high mean and standard deviation values. The high values are likely due to noise affecting the DIC measurements at very small displacements. Generally, the mean and standard deviation exhibited some variability across different values of the indicator $w_{\max,w}$. An exponential model was applied to the data to address the erratic behaviour (Fig. 6):

$$m_{V_u/V}(w_{\max,w}) = \begin{cases} 1.5\exp(-3w_{\max,w}) + 0.88 & w_{\max,w} < 0.85\ \text{mm} \\ 1 & w_{\max,w} \geq 0.85\ \text{mm} \end{cases} \tag{17a}$$

$$s_{V_u/V}(w_{\max,w}) = \begin{cases} 0.53\exp(-2.1w_{\max,w}) - 0.05 & w_{\max,w} < 1.1\ \text{mm} \\ 0 & w_{\max,w} \geq 1.1\ \text{mm} \end{cases} \tag{17b}$$

Eqs. (17a) and (17b) effectively capture the trend towards a mean value of 1 and a standard deviation of 0 as the load increases. It also becomes clear that for values of $w_{\max,w} \geq 1.1$ mm, the method offers no advantages compared to using the lower-bound approximation [5] as the resistance ratio becomes a deterministic value of 1. It should be noted that measurement uncertainty is not explicitly considered here (Section 3.4) since the possibly underestimated uncertainty (low standard deviation) is corrected by the adopted exponential model for small crack width values (Fig. 6). In addition, the use of Student's t-distribution is equivalent to the assumption of a non-informative prior for the mean and standard deviation. If Bayesian inference was performed,

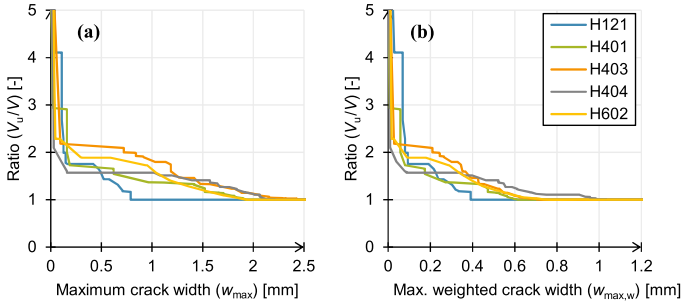


Fig. 5. Shear resistance ratio versus (a) the maximum nominal crack width and (b) the maximum weighted nominal crack width [9].

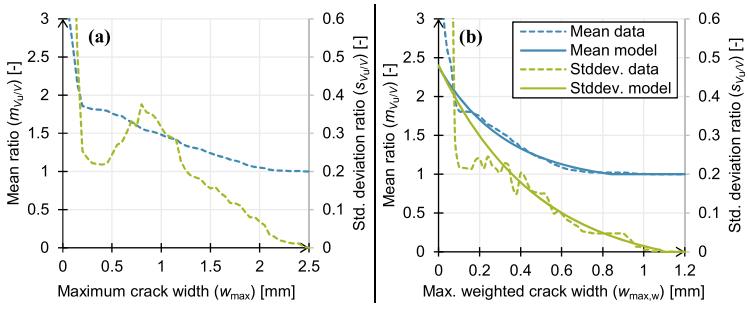


Fig. 6. Mean and standard deviation of the shear resistance ratio versus (a) the maximum nominal crack width and (b) the maximum weighted nominal crack width.

low-informative priors would be used, leading to somewhat lower dispersion.

4.3. Assumed load testing results

The bridge considered in this case study is theoretical and has not been tested. However, indicator values that would normally be acquired during testing are needed to apply the proposed method. Meaningful target loads can be based on the characteristic value of the traffic load multiplied by a certain factor. Given a target load, the tests performed in the laboratory may also be used to estimate likely indicator values (Table 1). In real-world applications, this step would not be required. The discussion (Section 6) describes the sensitivity to the assumed load testing results. It should be noted that measurements performed on the in-situ structure will already include the superimposed dead load (G_{SDL}) thus the measurements will begin from a different starting value. To compensate for this difference, the values should be increased by the crack widths provided in Table 1 for the case $G_{DL} + G_{SDL}$. Although the value is small in the current application, this step is important to align the measurements between the laboratory tests and the structure being monitored during the load test.

4.4. Probabilistic model and reliability analysis

To enable the update of structural reliability, a probabilistic model specific to the considered structure is required. The mean and coefficient of variation of the random variables in Table 2 are based on the Probabilistic Model Code [34] and fib Bulletin 80 [49]. The coefficient of variation used for model uncertainty of the shear resistance ($V_{\theta R} = 0.15$) is deemed appropriate for structures without shear reinforcement [64].

Table 1
Expected indicator readings given a proof load.

Test load level	Loads acting on structure	Expected shear force [kN]	Maximum crack width, weighted by position ($w_{max,w}$) [mm]					
			H121	H401	H403	H404	H602	Average
-	G_{DL}	30.0	0	0	0	0	0	0
-	$G_{DL} + G_{SDL}$	34.9	0.004	0.002	0.002	0.002	0.010	0.004
1	$G + 1.0\,Q_k$	73.0	0.069	0.058	0.019	0.010	0.016	0.034
2	$G + 1.2\,Q_k$	80.6	0.069	0.058	0.023	0.010	0.029	0.038
3	$G + 1.4\,Q_k$	88.2	0.069	0.071	0.026	0.016	0.058	0.048
4	$G + 1.6\,Q_k$	95.9	0.069	0.135	0.158	0.032	0.087	0.096
5	$G + 1.8\,Q_k$	103	0.074	0.174	0.244	0.052	0.256	0.160

Table 2
Overview of random variables in the limit state function.

Var.	Description	Distribution	Mean	COV
θ_R	Model uncertainty of the resistance	Lognormal	1	0.15
X	Resistance to current load effect ratio	Student's t	(varies)	(varies)
θ_E	Model uncertainty of the load effect	Lognormal	1	0.10
G_{DL}	Dead load effect	Normal	356 kN	0.05
G_{SDL}	Superimposed dead load effect	Normal	59.3 kN	0.10
C_{0Q}	Time-independent uncertainty of the traffic load, including dynamic effects	Lognormal	1.1	0.10
Q	Traffic load effect, annual maximum	Gumbel	390 kN	0.035
$\theta_{E,PL}$	Model uncertainty of the proof load effect; correlation $\rho(\theta_E, \theta_{E,PL}) = 0.7$	Lognormal	1	0.10
Q_{PL}	Load effect achieved by proof load	Normal	(varies)	0.02

The statistical description of the traffic load effect (Q) is based on weigh-in-motion (WIM, [65]) measurement data recorded in the Netherlands. This data was analysed to provide extreme value distributions for the annual maximum bending moment at midspan and shear force near the supports of a single-span structure [5]. The corresponding 1000-year characteristic load effect is obtained as $Q_{k,WIM} = F_Q^{-1}(0.999)$ according to the Eurocode where $F_Q^{-1}(\cdot)$ indicates the inverse CDF of Q. The mean test load levels ($m_{Q,PL}$) are related to the characteristic load effect via a factor (1.0–1.8) in the following analyses. The mean and standard deviation of the resistance ratio for each load level is calculated using Eqs. (17a) and (17b) based on the average indicator values from Table 1. The corresponding coefficient of variation is obtained as $V_X = s_X/m_X$ (Table 3). Finally, by making use of the probabilistic model specified in Table 2, reliability calculations are performed using the method outlined in Section 3. The calculated reliability indices are provided in Table 4.

The results presented in Table 4 clearly show that the proposed method provides higher annual reliability values compared to the lower-bound approach, i.e. Eq. (12). This difference signifies the advantage of integrating additional information from indicators that reflect structural performance. The value of this additional information is particularly significant when the test loads are relatively low, illustrated by an increase in the reliability index of about 1.5 with moderate loads. In addition, Table 4 provides the reliability during the proof load test, indicating the risk associated with applying the target load. A reliability index of 1.8, corresponding to a failure probability of approximately 0.036, suggests the structure is indeed likely to survive the applied load. The reliability during testing is also updated with each increment in the load level, enabling continuous assessment of risk during the test (Section 6).

Table 3
Mean and coefficient of variation of the resistance ratio (X) for each load test cycle.

WIM char. load factor ($m_{Q,PL,s}/Q_{k,WIM}$)	Mean PL effect, strip ($m_{Q,PL,s}$) [kN]	Mean PL effect, lane ($m_{Q,PL}$) [kN]	Indicator value ($w_{max,w}$) [mm]	Mean of resistance ratio (m_X)	COV of resistance ratio (V_X)
1.0	38.1	457	0.034	2.23	0.198
1.2	45.7	549	0.038	2.22	0.198
1.4	53.3	640	0.048	2.18	0.197
1.6	61.0	732	0.096	2.00	0.191
1.8	68.7	823	0.0160	1.81	0.182

Table 4
Calculated reliability indices given different levels of loading, during and after proof load testing (PLT).

WIM characteristic load factor ($m_{Q,PL}/Q_{k,WIM}$)	LM1 characteristic load factor ($m_{Q,PL}/Q_{k,LM1}$)	Reliability during PLT	Annual reliability after successful PLT	
			Proposed method	Lower-bound
1.0	0.78	1.80	2.99	0.53
1.2	0.93	1.79	3.77	1.89
1.4	1.09	1.78	4.55	3.04
1.6	1.24	1.71	5.23	4.03
1.8	1.40	1.61	5.83	4.85

5. Bending resistance assessment supported by a calculation model

5.1. De Beek viaduct

In this case study, proof load levels relating to the bending resistance of the reinforced concrete slab viaduct De Beek in the Netherlands will be examined. The viaduct from 1963 passes over the A67 highway near Ommel, but the viaduct itself is part of the secondary road network. The highway traffic passes beneath the viaduct's two central spans, each with a length of 15.4 m. Additional spans on either side have a length of 10.8 m. The cross-section height varies parabolically along the longitudinal direction from 0.47 m to 0.87 m, measured in the heart of the deck. The bridge is constructed as a continuous slab and thus experiences support moments above its middle three supports (Fig. 7). The sidewalks, added during a later phase of construction, are assumed to offer no significant contribution to the structural resistance (but can influence the stiffness). A pilot proof load test of the first span has already been performed for both bending and shear mechanisms, but only bending will be considered in this case study. Thus, in contrast to the previous case study (Section 4), in-situ test results and measurements are available [66]. A calculation model for the bending resistance will be used to interpret the measurements during testing because laboratory tests on similar bridge decks are not available.

In 2015, an inspection and assessment of the bridge took place and it was concluded that the bridge had insufficient resistance [67]. During the inspection, cracks at the bottom of the concrete slab were found. Afterwards, the detected cracks were filled to prevent water ingress. Later that year, further investigation and a pilot proof load test were performed, confirming that the resistance of span 1 was sufficient for two traffic lanes, both calculated and tested. For span 2, the resistance was only deemed sufficient if plastic redistribution would be allowed to take place, but this typically results in unwanted cracking. The bridge inspection and assessment led to a traffic restriction that reduced the original two lanes to one central lane, only allowing traffic from one direction at a time [68].

5.2. Traffic load effect

The viaduct was designed for two lanes, with traffic in opposite directions. The width of the carriageway is 7.44 m. After subtracting the width of the curbs, the remaining lane width is 3.5 m for each direction. To obtain the most accurate statistical description of the load effect, additional traffic simulations were performed using WIM data recorded in the Netherlands. In this way, the specific continuous slab configuration with traffic moving in opposite directions can be taken into account. The viaduct crosses the A67 highway but is part of the secondary road network. However, only WIM data recorded at highway locations is available in the Netherlands. Using the highway measurements likely results in overestimating the true load effect for this particular location. However, because the viaduct is situated in a rural area, tractors and slurry tanks also make use of it. Although the number of heavy

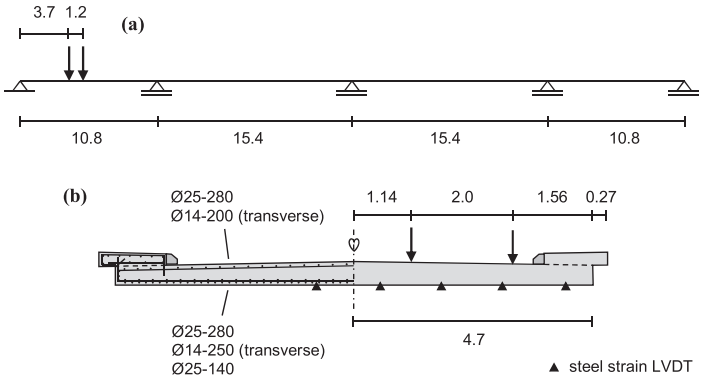


Fig. 7. Schematic (a) side view and (b) cross-section of the De Beek viaduct with the bending proof load test location indicated by the arrows. Reinforcement diameters and spacings provided in mm, other measurements provided in m.

agricultural vehicles is small, they may exert relatively large axle loads on the viaduct. Therefore, no reduction of the traffic load effect following the highway WIM data was applied.

From the WIM measurements, only the first lane data was used in the traffic simulations (i.e. where the trucks predominantly drive). The simulation uses a time-discretisation method in which the vehicle axle loads are placed on the bridge according to the time they were originally recorded. The linear load effect, i.e. the bending moment, was calculated through the superposition of the load effects caused by each individual axle. This calculation may be effectively performed using influence lines, which express the load effect at a certain location, given a unity axle load (Fig. 8). The influence lines were derived using a 1D finite element analysis utilising many small Euler-Bernoulli beam elements to account for the varying deck height in the longitudinal direction.

By analysing the bending moment caused by permanent and traffic loads it follows that the maximum combined load effect can be expected around $x = 4.5$ m. To identify the difference in the bending moment following from one lane and two lanes in opposite directions, both situations were analysed. The load effect in the two-lane configuration is not simply the one-lane load effect multiplied by a factor 2 because it is very unlikely that, in both directions, a heavy truck is present at the same time. However, the variability in the load effect increases considerably, as shown by the more gradually descending right tail of the distribution in the two-lane configuration (Fig. 9). The Gumbel distribution for the weekly maxima was found by fitting the right tail of the weekly extremes for the WIM highway locations (A16L, A27L, A50L and A67L) directly. The yearly distributions we found by modifying the Gumbel distribution location parameter as $\mu_{yr} = \mu_{wk} + \ln(52)\,\beta$ where 52 is the number of weeks in the year, and β is the scale parameter of the Gumbel distribution (which remains unchanged). The assessment of the first span will be performed for the two-lane configurations, for which the annual load effect is found to follow a Gumbel distribution with a mean value of 1301 kNm and a variation coefficient of 0.058.

5.3. Proof load test and measurements

In November 2015, proof load tests were performed on the first span of the viaduct De Beek using load levels described by relevant Dutch guidelines and standards [69–71]. The load was applied as the Eurocode LM1 tandem with two axles at a distance of 1.2 m with a wheelbase of 2 m and a wheel print size of 0.4 m × 0.4 m. The axle load was varied to achieve a load effect corresponding to various load levels (Table 5). The corresponding load effects caused by the tandem axle loads are calculated using the influence lines derived in Section 5.2. The bending moment caused by the LM1 tandem is obtained as $M = 3.75\,P$ where P is the load of an individual axle. The load test situation mimics the situation in which a heavy vehicle passes the bridge in a two-lane configuration. Due to the eccentric placement of the load, the bending moment

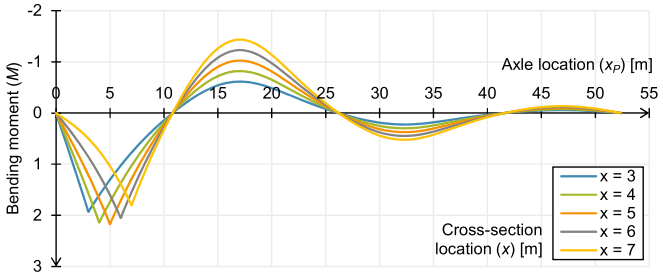


Fig. 8. Influence lines for the bending moment considering various cross-section locations, i.e., the moment for a cross-section located at x, resulting from a unit axle load ($P = 1$) positioned at x_P.

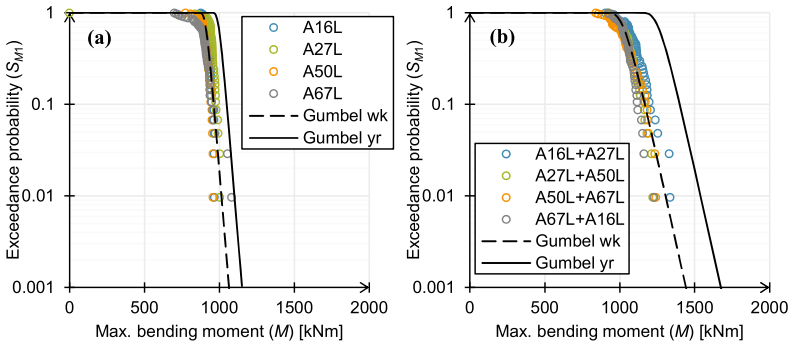


Fig. 9. Extreme value distributions of the maximum load effect in the first span for the (a) one-lane and (b) two-lane configurations.

Table 5
Overview of test loads, load effects and measured steel strains.

Cycle	Description	P_{tot} [kN]	P [kN]	M [kNm]	ε_s [10^{-6}]
1	Unfactored tandem	600	300	1125	165
2	Service level	1000	500	1875	340
3	RBK usage	1400	700	2625	570
4	Intermediate	1700	850	3188	790
5	Eurocode design	1750	875	3281	830

varies in the transverse direction of the deck. In this case study, the deck will be treated as a beam for simplicity, but the relative increase of strains near the edge will be taken into account (see Section 5.4.1).

During the test, various measurements were performed, including displacements, concrete surface strains and reinforcement strains. At a number of locations, the concrete cover was removed to inspect the reinforcement and attach strain gauges [68]. If the reinforcement cannot be accessed directly, stop-criteria expressed in concrete surface strains may prove beneficial [72]. In this case study, the maximum steel strain measured at 0.86 m from the edge of the deck (see Fig. 7) is of primary interest since yielding of the reinforcement is undesired for durability reasons. In the longitudinal direction, the strains are measured at a distance of 4.05 m from the support and are considered representative of the section between the axle loads (3.7–4.9 m), as the moment remains nearly constant throughout this range.

5.4. Analytical resistance model

5.4.1. Modified sectional analysis

If no tests have been performed on representative specimens, it is possible to use an analytical model to interpret measurements during the test. The resistance model may range from simple to complex, e.g. cross-sectional analysis [73,74], strip model [75,76], linear finite element model, non-linear finite element model [77,78]. In the considered case, the proof load test has already been performed, and therefore, the measurements can be used to correct a relatively simple sectional analysis model. This calculation is performed by gradually incrementing the strains across the section according to a linear strain distribution and finding the compression zone depth for which the axial force and bending moments in the section are in balance. The result is often presented as a moment-curvature (M-κ) diagram. In this case study, the bottom steel strain (ε_s) is utilised instead of the curvature (κ) because the steel strain was measured during testing.

For the sectional analysis of the De Beek bridge deck, a width of b = 9.4 m is used, along with an equivalent deck height of h = 0.47 m at the location where the maximum moment is deemed to occur (x = 4.5 m, Section 5.2). The reinforcement, as schematically presented in Fig. 7, listed from bottom to top, corresponds to the longitudinal reinforcement areas $A_{s,bot,1}$ = 3506 mm^2/m, $A_{s,bot,2}$ = 1753 mm^2/m and $A_{s,top}$ = 1753 mm^2/m. Thorenfeldt's model [79] is used for the concrete compressive stress-strain curve, and no concrete tensile strength contribution is assumed. The analysis is first performed using mean values for the geometry and material properties (Table 6) to establish the model correction.

Analysis of the measurement data reveals that the recorded strains must be adjusted to compare the numerically obtained response with the observed measurements. If a linear response up to and between load cycles 1 and 2 is assumed, the microstrain increment corresponding to 1875 − 1125 = 750 kNm is 340 − 165 = 175 (Table 5). Then, a microstrain of 1125/750 · 175 = 263 would be expected for the first cycle, which is about 100 higher than recorded. However, the corrected model is non-linear, as a result of Eq. (18), and the required adjustment has been determined as 85 iteratively.

The moment caused by the self-weight of the deck and the superimposed loads is estimated as M_G = 747 kNm [66]. The corrected cross-sectional analysis model results in a corresponding bottom steel strain $\varepsilon_{sG} = 154 \cdot 10^{-6}$. The calculated steel strains need to be lowered compared to the original model to align with the measured response (Fig. 10). This increase in stiffness could be caused by the contribution of the sidewalks. In addition, the measurements show a gradual decrease in stiffness that is not present in the modelled response. Therefore, a power law is included to increase the steel strain proportionally to the reinforcement's yield strain (ε_y; mean value $\varepsilon_{ym} = f_{ym} / E_{sm} = 1417 \cdot 10^{-6}$). The adopted expression for the modified steel strain is:

$$\varepsilon_s^* = \left[c_1 + c_2 \left(\varepsilon_s / \varepsilon_y\right)^{c_3}\right] \varepsilon_s \tag{18}$$

Table 6
Overview of the random variables in the mechanical model.

Var.	Description	Distribution	Mean	COV
c	Concrete cover thickness	Gamma	30 mm	0.17
h	Height of the cross-section	Lognormal	470 mm	0.10
E_s	Young's modulus of reinforcing steel	Lognormal	205 GPa	0.02
f_c	Concrete compressive strength	Log-t (n = 6)	57.5 MPa[a]	0.10[a]
f_y	Yield stress of the reinforcement	Log-t (n = 3)	290.5 MPa[a]	0.034[a]
c_2	Non-linearity of moment-strain relation	Lognormal	0.41	0.10

[a] Sample statistics are reported. The log-t prediction distribution accounts for the small number of samples and will effectively lead to higher variability.

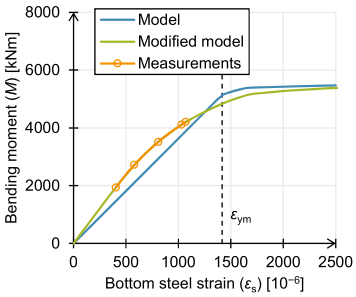


Fig. 10. Comparison of the moment-strain diagram following from modelling and measurements.

where c_1 accounts for the difference in stiffness, c_2 controls the degree of non-linearity and c_3 specifies the shape of the curve. An optimal fit for this case study is obtained with $c_1 = 0.75$, $c_2 = 0.41$ and $c_3 = 4.1$. As a result, the modified model predicts a lower (local) yield moment, but the ultimate resistance remains the same. In this way, the model accounts for eccentric loading with increased strains towards the side of the bridge deck where the load is applied.

5.4.2. Sampling of the mechanical model

Since no laboratory measurements are available, the uncertainty regarding the structural response must be included in an alternative manner. To this end, the mechanical model can be set up using random variables for the parameters (Table 6). The distribution types and coefficients of variation (COV) follow fib [49] and JCSS [34]. A larger COV value is adopted for the height of the cross-section since it is an equivalent value for the actual height, which varies along the longitudinal and transverse directions. The steel area of the reinforcement is not included because its variability is minimal. Material testing was performed on the concrete and the reinforcing steel [67]. It is expected that concrete compressive strength and yield stress of the reinforcement would follow a lognormal distribution when many measurements are available. Therefore, the prediction distribution, Eq. (8), is used to describe the logarithm of these material properties. The resulting log-t distributed random variables for the concrete compressive strength (f_c) and the yield stress of the reinforcement (f_y) are:

$$f_c = \exp\left(4.05 + 0.106\, T_{\nu=6-1}\sqrt{1+1/6}\right) \quad [\text{MPa}] \tag{19a}$$

$$f_y = \exp\left(5.76 + 0.0346\, T_{\nu=3-1}\sqrt{1+1/3}\right) \quad [\text{MPa}] \tag{19b}$$

where random variable T_ν follows Student's t-distribution with ν degrees of freedom. The sample mean and standard deviation values have been calculated from the log-transformed measurement data (i.e. taking the natural logarithm of each value). The uncertainty about the non-linearity of the modified sectional analysis model is included in the probabilistic description by using a lognormal distribution for the parameter c_2 in Eq. (18), which controls the degree of non-linearity.

Latin hypercube sampling [23,52] was used to simulate responses according to the probabilistic model (Table 6). Compared to the more common Monte Carlo simulation [16], Latin hypercube sampling better represents the output distribution when using a small number of samples. Although the mechanical model is not computationally demanding in this case study, a more complex (FEM) model will require considerable computation time. The calculated moment-strain curves and the resistance ratios for 100 samples are displayed in Fig. 11. Because the yielding of reinforcement is undesired, the yield moment resistance (M_y) is used instead of the ultimate bending resistance (M_u) to calculate the ratio $X = M_y/M$. The sample yield moment is obtained by calculating the sample yield stress ($\varepsilon_y = f_y / E_s$) and subsequently interpolating the moment-strain curve to find the corresponding moment. In Fig. 12, the calculated mean and standard deviation of the resistance ratio are plotted against the bottom steel strain. In contrast to the approach in Section 4.2, the statistics of the resistance ratio X can be directly obtained from the curves. The mean and standard deviation relations are discretised using a strain interval of $50 \cdot 10^{-6}$ and can be linearly interpolated in the subsequent reliability analysis.

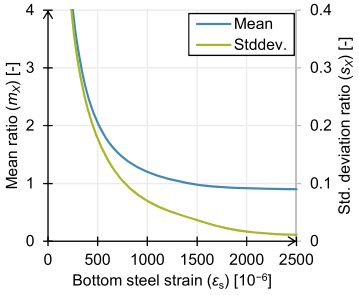


Fig. 12. Mean and standard deviation of the simulated resistance ratio.

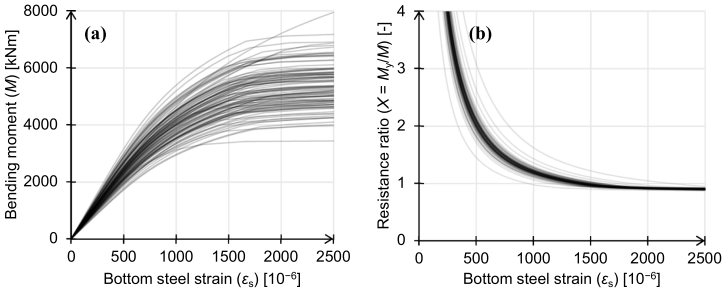


Fig. 11. Simulation result displayed as moment-strain (a) and resistance ratio-strain curves (b).

5.5. Probabilistic model and reliability analysis

The probabilistic model described in Section 3.2 is now employed to perform the reliability analysis. The mean values and variation coefficients of the random variables are specified in Table 7 and are based on the Probabilistic Model Code [34] and fib Bulletin 80 [49]. A statistical description of the traffic load effect was obtained in Section 5.2. The statistics for the resistance ratio depend on the indicator value – as discussed in Section 5.4.2. In this case, the indicator is the steel strain in the bottom reinforcement near the edge of the deck (ε_s). In contrast to the previous case study (Section 4), the resistance ratio (X) follows a normal distribution, not Student's t-distribution, because the resistance model is based on a stochastic model. The coefficient of variation used for the model uncertainty of the resistance ($V_{\theta R} = 0.15$) is twice the value generally used for bending [64]. This increased value reflects the greater uncertainty associated with the use of a basic sectional analysis model for a bridge deck subjected to eccentric loading (Section 5.4.1).

For each of the load levels used in the proof load test, the mean and coefficient of variation of the resistance ratio are calculated (Table 8). The load levels may also be related to the WIM and LM1 characteristic traffic load effects to enable comparisons with the previous case study (Section 4). For each of these load levels, the reliability during and after surviving the proof load test has been calculated (Table 9). Similar to the previous case study, using the indicator data results in markedly higher reliability indices at low load levels than the lower-bound method, i.e. Eq. (12). The reliability during testing is relatively high but decreases rapidly with higher test loads, underlining the need for risk assessment during the proof load test between load steps (Section 6).

6. Discussion

In the data post-processing of the shear resistance case study (Section 4.2), the model deviates from the calculated values for $w_{\max,w} < 0.08$ mm. This region is important for the subsequent reliability calculations and greatly influences the outcomes. As described in Section 3.4, measurement errors play a significant role with small indicator values and various strategies are discussed to account for them. The power-law model was chosen because analytical resistance models show a similar decreasing trend (Section 5.4). Further data inspection reveals very high resistance ratios in the small-value region with a positive distribution skew. Thus, the adoption of a conservative model in this region effectively mitigates measurement-error issues. Additional calculations using Bayesian inference to statistically describe the resistance ratio resulted in higher reliability indices (up to 0.5), thereby confirming the effectiveness of the chosen approach.

Attention must be given to ensuring the similarity between the in-situ tested structure and the results from laboratory tests or analytical models. In the shear resistance case study, it is assumed that the response of the bridge slab is similar to the response of the strips tested in the laboratory. However, still a relatively large coefficient of variation was assigned to the model uncertainty to account for the remaining differences (e.g., transverse load distribution, boundary conditions, edge and size effects). With analytical models, it is crucial to evaluate whether they can reliably provide correct indicator values (such as crack widths, strains, etc.), as these models often primarily consider the resistance of structural members. Experience with the proposed method in practice can establish suitable model uncertainty values.

In reality, the tested structure may exhibit a very different response than assumed in both case studies. If the structure's condition is above average, smaller indicator values are anticipated and therefore the reliability increases. Conversely, for structures that show large indicator values under small loads, lower reliability is expected. The results include reliability indices that will likely be observed in practical applications, but may differ from case to case. The same holds for the relation between the target load, the statistical description of the traffic load effect and the load effect calculated from standards. For example, the required target load for the bending resistance assessment of the viaduct is relatively low when expressed by the unfactored LM1 characteristic load (Section 5.5). This result stems from the detailed statistical analysis of the traffic load effect, indicating that the LM1 characteristic load may be rather conservative for the two-lane situation with opposite driving directions.

The proposed method also provides the opportunity to calculate the reliability during the proof load test. Although no standards or guidelines specify minimum reliability levels for load testing, case-specific risk analysis can help determine optimal values. Such an analysis can take into account the load-testing conditions (e.g., absence of traffic, restricted areas), allowing for informed decisions on whether to continue with higher load levels or stop the test. This approach offers a risk-optimal alternative to pre-determined stop criteria but requires additional analysis and calculation effort.

The value of the current study lies in the proposed method for proof load testing rather than the calculated reliability values and target loads. Future research and practical applications can further refine the

Table 7
Overview of random variables in the limit state functions.

Var.	Description	Distribution	Mean	COV
θ_R	Model uncertainty of the resistance	Lognormal	1	0.15
X	Resistance to current load effect ratio	Normal	(varies)	(varies)
θ_E	Model uncertainty of the load effect	Lognormal	1	0.10
G_{DL}	Dead load effect	Normal	604 kNm	0.05
G_{SDL}	Superimposed dead load effect	Normal	143 kNm	0.10
C_{0Q}	Time-independent uncertainty of the traffic load, including dynamic effects	Lognormal	1.1	0.10
Q	Traffic load effect, annual maximum	Gumbel	1301 kNm	0.058
$\theta_{E,PL}$	Model uncertainty of the proof load effect; correlation $\rho(\theta_E, \theta_{E,PL}) = 0.7$	Lognormal	1	0.10
Q_{PL}	Load effect achieved by proof load	Normal	(varies)	0.02

Table 8
Mean and coefficient of variation of the resistance ratio (X) for each load test cycle.

Cycle	Description	Measured steel strain [10^{-6}]	Indicator value (ε_s) [10^{-6}]	Mean of ratio (m_X)	COV of ratio (V_X)
1	Unfactored tandem	165	419	2.42	0.090
2	Service level	340	594	1.76	0.082
3	RBK usage	570	824	1.36	0.068
4	Intermediate	790	1044	1.17	0.056
5	Eurocode design	830	1084	1.15	0.054

Table 9
Calculated reliability indices given different levels of loading, during and after proof load testing (PLT).

WIM characteristic load factor ($m_{Q,PL}/Q_{k,WIM}$)	LM1 characteristic load factor ($m_{Q,PL}/Q_{k,LM1}$)	Reliability during PLT	Annual reliability after successful PLT	
			Proposed method	Lower-bound
0.67	0.46	4.78	3.58	−1.70
1.12	0.76	3.18	3.99	1.95
1.57	1.06	1.79	5.00	4.12
1.91	1.29	0.91	5.78	5.27
1.96	1.33	0.80	5.92	5.40

framework, providing valuable insights into model uncertainties and measurement techniques. Real-world validation is recommended to establish the robustness of the adopted models and data post-processing procedures. Large-scale implementation would benefit from research into generally applicable indicators, resistance ratio curves, identification of application criteria and possibly (FEM) modelling guidelines.

7. Conclusions

The proposed proof load testing method involves updating the resistance distribution using two sources of information: (1) the observed in-situ response, which is related to the resistance via indicator values and associated resistance ratios, and (2) the survival of the applied load during the proof load test. Although the prediction of the resistance on the basis of the in-situ response is associated with considerable uncertainty, this information is still valuable and can be accounted for probabilistically. A Bayesian procedure for updating the resistance distribution given the two information sources is presented, along with a method for calculating the posterior reliability using the Markov chain Monte Carlo (MCMC) technique. In addition, the probability of failure during the test is calculated using the most current data, offering a risk-optimal alternative to pre-determined stop criteria.

The application of the proposed method was demonstrated through two case studies. The first case study considered a hypothetical shear-critical concrete bridge using laboratory data, while the second examined the bending resistance of an existing viaduct in the Netherlands using an analytical model. These case studies illustrated how the framework could be applied in real-world scenarios, and the potential gains in reliability when monitoring data is included in the assessment. The two case studies in which in-situ measurement data was included allowed for test load reductions of 20 % and 25 %. The proposed method's versatility was highlighted by using laboratory experiments in the first case study and an analytical model in the second to establish the relationship between indicator values and resistance ratios. In cases with complex mechanical behaviour where tests or analytical models are less representative of the in-situ structure, smaller test load reductions should be expected due to the increased uncertainty.

Advanced probabilistic models and calculation techniques were utilised to account for uncertainties in resistance, material properties, load effects, and measurement errors. Laboratory data post-processing, including the consideration of weighted crack widths, provided insight into the resistance ratios for different target loads. The impact of model uncertainty and measurement errors was addressed, particularly for small indicator values. Alternatively, analytical models can be used to derive similar insights, ensuring the model closely matches the in-situ conditions.

The suggested method offers a more comprehensive and accurate approach to evaluating existing infrastructure using proof load testing. Using in-situ measurements, the procedure also enables the calculation of failure probability during the test, allowing for risk-based decisions on whether to proceed or stop. Practical applications of the method can determine whether similar reductions in test loads, as found in the current study, are feasible. If so, proof load testing can become more economically attractive and less time-consuming, minimising the traffic disruption involved in bridge testing.

CRediT authorship contribution statement

Hendriks M.A.N.: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Steenbergen R.D.J.M.:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Naaktgeboren M.:** Writing – review & editing, Funding acquisition, Conceptualization. **Lantsoght E.O.L.:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **de Vries R.:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

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

Acknowledgements

The authors wish to express their gratitude and sincere appreciation to the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat) for financing the research work. In addition, the fruitful discussions with Sonja Fennis of Rijkswaterstaat, Ton Vrouwenvelder of TNO and Yuguang Yang from Delft University of Technology have been of great help.

Data Availability

Data will be made available on request.

References

[1] Steenbergen RDJM, Vrouwenvelder ACWM. Safety philosophy for existing structures and partial factors for traffic loads on bridges. Heron 2010;55(2):123–39.
[2] Vrouwenvelder T, Scholten N. Assessment criteria for existing structures. Struct Eng Int 2010;20(1):62–5.
[3] Faber MH, Val DV, Stewart MG. Proof load testing for bridge assessment and upgrading. Eng Struct 2000;22:1677–89.
[4] Casas JR, Gómez JD. Load rating of highway bridges by proof-loading. KSCE J Civ Eng 2013;17(3):556–67.
[5] De Vries R, et al. Proof load testing method by the American association of state highway and transportation officials and suggestions for improvement. Transp Res Rec 2023.
[6] Frangopol DM, et al. *Reliability-based analysis and lifecycle management of load tests.* Book chapter. In: Lantsoght EOL, editor. Load testing of bridges: Proof load testing and the future of load testing; 2019. p. 263–92.
[7] Lantsoght EOL. Load testing of bridges: Proof load testing and the future of load testing. In: Frangopol DM, editor. Structures and Infrastructures. London, UK: CRC Press / Balkema - Taylor & Francis Group; 2019.
[8] De Vries R., Reliability assessment of existing reinforced concrete bridges and viaducts through proof load testing Proceedings of the 11th International Conference on Bridge Maintenance, Safety and Management (IABMAS), Barcelona, Spain; 2022.
[9] De Vries R., Structural reliability updating using monitoring data from in-situ load testing and laboratory test results. Proceedings of the 13th International Conference on Bridge Maintenance, Safety and Management (IABMAS), Copenhagen, Denmark; 2024. p. 409-417..
[10] Zarate Garnica GI, Lantsoght EOL. Stop criteria for proof load testing of reinforced concrete structures. Proc 2021 Sess 13th fib Int PhD Symp Civ Eng 2021:195–202.
[11] Ditlevsen O, Madsen HO. Book. Structural reliability methods. John Wiley & Sons; 1996.
[12] Madsen HO, Krenk S, Lind NC. Methods of structural safety. Englewood Cliffs, New Jersey: Prentice Hall; 1986. p. 403.
[13] Der Kiureghian A. Structural and system reliability. Cambridge University Press; 2022.
[14] CEN, Eurocode 0: Basis of structural design. Standard, EN 1990+A1+A1/C2:2019, European Committee for Standardization (CEN), Brussels, Belgium; 2019.
[15] Wasserman L. All of statistics: A concise course in statistical inference. Book, Springer; 2004.
[16] Metropolis N, Ulam S. The Monte Carlo method. J Am Stat Assoc 1949;44(247):335–41.
[17] De Vries R, Steenbergen RDJM, Maljaars J. Annual reliability requirements for bridges and viaducts. Heron 2023;68(2).
[18] Melhem M.M., Caprani C.C., Stewart M.G., Zhang S. Bridge Assessment Beyond the AS 5100 Deterministic Methodology Research Report AP-R617-20, Austroads, Sydney, Australia; 2020.
[19] Kloek T, Van Dijk HK. Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. Econometrica 1978;46(1):1–19.
[20] Bjerager P. Probabilistic integration by directional simulation. J Eng Mech 1988;114(8):1285–302.
[21] Waarts PH. Structural reliability using finite element analysis - An appraisal of DARS: Directional Adaptive Response surface Sampling. Delft University of Technology; 2000.
[22] Moustapha M, Marelli S, Sudret B. Active learning for structural reliability: survey, general framework and benchmark. Struct Saf 2022;96:102174.

[23] McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 1979;21(1):239–45.
[24] Hasofer AM, Lind NC. Exact and invariant second-moment code format. J Eng Mech Div 1974;100(1):111–21.
[25] Rackwitz R, Fiessler B. Structural reliability under combined random load sequences. Comput Struct 1978;9. 498-494.
[26] Fiessler B, Neumann HJ, Rackwitz R. Quadratic limit states in structural reliability. Eng Mech 1979;105(4):661–76.
[27] Breitung K. Asymptotic approximations for multinormal integrals. Eng Mech 1984;110(3):357–66.
[28] Phoon KK. Numerical recipes for reliability analysis – a primer. in Reliability-Based Design in Geotechnical Engineering: Computations and Applications. Taylor & Francis; 2008. p. 75.
[29] Hohenbichler M, et al. New light on first- and second-order reliability methods. Struct Saf 1987;4:267–84.
[30] Tvedt, L., Two second order approximations to the failure probability. Technical report, RDIV/20-004-83, Det Norske Veritas; 1983.
[31] Geisser S. Predictive inference: An introduction. New York: Chapman & Hall; 1993.
[32] Box GE, Hunter JS, Hunter WG. Statistics for Experimenters: Design, Innovation, And Discovery. 2nd edition. Hoboken, New Jersey, USA: Wiley-Interscience; 2005.
[33] Gosset WS. The probable error of a mean. Biometrika 1908;6(1):1–25.
[34] JCSS ; 2015, Available from: ⟨https://www.jcss-lc.org/jcss-probabilistic-model-code/⟩.Probabilistic model code⟨https://wwwjcss-lc.org/jcss-probabilistic-model-code/⟩..
[35] Ditlevsen O, Vrouwenvelder A. "Objective" low informative priors for Bayesian inference from totally censored Gaussian data. Struct Saf 1994;16:175–88.
[36] Straub D, Papaioannou I. Bayesian updating with structural reliability methods. Eng Mech 2015;141(3).
[37] Ghose MK, Rajagopalan S. Bayesian reliability assessment for discrete data—a case study. Reliab Eng 1985:27–36.
[38] Peterka V. Bayesian system identification. Automatica 1981;17(1):41–53.
[39] Lauritzen SL. Sequential Bayesian updating – BS2 statistical inference, lectures 14 and 15. Lect Notes, Univ Oxf 2009.
[40] Alam J, et al. Sequential Bayesian updating for time-variant reliability analysis of ageing structures. Mech Syst Signal Process 2023;204(110774).
[41] Jeffreys H. Theory of probability, 3rd ed Book, Oxford University Press; 1961.
[42] Metropolis N, et al. Equations of state calculations by fast computing machines. Chem Phys 1953;21(6):1087–91.
[43] Hastings WK. Monte Carlo sampling methods using Markov Chains and their applications. Biometrika 1970;57(1):97–109.
[44] Benjamin JR, Cornell CA. Probability, statistics, and decision for civil engineers. Book, McGraw-Hill; 1970.
[45] AASHTO The manual for bridge evaluation Standard, 3rd Edition. Washington, D. C., USA; 2018.
[46] De Vries R, et al. Time-dependent reliability assessment of existing concrete bridges with varying knowledge levels by proof load testing. Struct Infrastruct Eng 2023;20(7-8):1053–67.
[47] Lin TS, Nowak AS. Proof loading and structural reliability. Reliab Eng 1984;8:85–100.
[48] Brüske H.Structural test design with value of information PhD Thesis (Report 401), DTU, Kongens Lyngby, Denmark; 2018.
[49] fib Partial factor methods for existing concrete structures. Bulletin 80, Recommendation, Task Group 3.1, Fédération internationale du béton, Lausanne, Switzerland; 2016.
[50] Abril-Pla O, et al. PyMC: a modern, and comprehensive probabilistic programming framework in Python. PeerJ Comput Sci 2023;9.
[51] Lauritzen SL, et al. Independence properties of directed Markov fields. Networks 1990;20:491–505.
[52] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Saf 2003;81:23–69.
[53] Sykora M, Holicky M. Assessment of uncertainties in mechanical models. Appl Mech Mater 2013;378:13–8.
[54] MacKay DJC. Bayesian methods for adaptive models. Pasadena, California, U.S.A.: California Institute of Technology; 1991.
[55] Bishop CM. Pattern recognition and machine learning. New York, U.S.A.: Springer; 2006.
[56] Hornberger GM, Spear RC. Eutrophication in peel inlet—I. The problem-defining behavior and a mathematical model for the phosphorus scenario. Water Res 1980;14(1):29–42.
[57] Dilks DW, Canale RP, Meier PG. Development of Bayesian Monte Carlo techniques for water quality model uncertainty. Ecol Model 1992;62(1):149–62.
[58] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 1984;6(6).
[59] Yang Y, Van der Ham HWM, Naaktgeboren M. Shear capacity of RC slab structures with low reinforcement ratio – An experimental approach. Lisbon, Portugal: fib Symposium; 2021.
[60] Koekkoek R.T., Yang Y. Measurement report on the transition between flexural and shear failure of RC beams without shear reinforcement Delft University of Technology, Stevin Report 25.5-16-04, Delft, The Netherlands; 2016.
[61] Zhang F. Acoustic emission-based indicators of shear failure of reinforced concrete structures without shear reinforcement. The Netherlands: Delft University of Technology, Delft; 2022.
[62] Jones EM, Iadicola MA. A good practices guide for digital image correlation. International Digital Image Correlation Society; 2018.
[63] Zarate Garnica G.I., Vries R.De, Lantsoght E.O.L. Analysis report of reinforced concrete slabs for stop criteria Stevin Report 25.5-22-02, Delft University of Technology, Delft, The Netherlands; 2022.
[64] Sýkora M, et al. Uncertainties in resistance models for sound and corrosion-damaged RC structures according to EN 1992-1-1. Material and Structures 2015;48(10):3415–30.
[65] FHWA, Weigh-in-motion pocket guide - Part 1: WIM technology, data acquisition, and procurement guide. Federal Highway Association, FHWA-PL-18-015, Washington D.C., USA; 2018.
[66] Koekkoek R.T., Analysis report for the assessment of viaduct De Beek by proof loading Stevin Report 25.5-16-01, Delft University of Technology, Delft, The Netherlands; 2016.
[67] Iv-Infra INPA120717-17-RAP-0043, Papendrecht, The Netherlands (in Dutch).51H-304-01 - De Beek - Herberekening brugdek (Recalculation bridge deck) Report; 2015.
[68] Lantsoght EOL, et al. Pilot proof-load test on viaduct De Beek: case study. J Bridge Eng 2017;22(12).
[69] NEN Beoordeling van de constructieve veiligheid van een bestaand bouwwerk bij verbouw en afkeuren - Grondslagen (Assessment of the structural safety of an existing structure for renovation and rejection - Basic requirements). Standard, NEN 8700:2011, Nederlands Normalisatie-instituut, Delft, The Netherlands; 2011.
[70] RWS Richtlijnen beoordeling kunstwerken - Beoordeling van de constructieve veiligheid van een bestaand kunstwerk bij verbouw, gebruik en afkeur (Guidelines for assessing structures - Assessment of the structural safety of an existing structure for renovation, use and rejection).Rijkswaterstaat, Report, Version 1.1, 27 May; 2013.
[71] CEN Standard, EN 1991-2+C1:2015, European Committee for Standardization (CEN), Brussels, Belgium, Eurocode 1: Actions on structures - Part 2: Traffic loads on bridges; 2015.
[72] Lantsoght EOL, et al. Stop criteria for flexure for proof load testing of reinforced concrete structures. Front Built Environ 2019;5.
[73] Mörsch E. Der Eisenbetonbau: seine Theorie und Anwendung (Reinforced concrete: theory and practice), Konrad Wittwer, Stuttgart, Germany; 1908.
[74] Wight JK, MacGregor JG. Reinforced concrete: Mechanics and design. 6th edition. New Jersey, USA: Pearson Education, Inc; 2012.
[75] Lantsoght EOL, et al. Extended strip model for reinforced concrete slabs under concentrated loads. Acids Struct J 2017;114(2):565–74.
[76] Lantsoght E.O.L., Ospina C.E., Alexander S.D.B. and , Punching capacity of spread footings using ACI 318-19 and the strip model ACI Special Issue SP-357: Punching shear of concrete slabs: insights from new materials, tests, and analysis methods, American Concrete Institute, Farmington Hills, USA; 2023.
[77] De Borst R, et al. Non-linear finite element analysis of solids and structures. 2nd edition. West Sussex, UK: John Wiley & Sons, Ltd; 2012.
[78] Lantsoght EOL, et al. Optimizing finite element models for concrete bridge assessment with proof load testing. Front Built Environ 2019;5.
[79] Thorenfeldt E., Tomaszewicz A., Jensen J.J. and , Mechanical properties of high-strength concrete and applications in design Proceedings of the Symposium on the Utilization of High-Strength Concrete, Stavanger, Norway; 1987.

E

Structural Safety article

De Vries, R., Lantsoght, E. O. L., Steenbergen, R. D. J. M., Hendriks, M. A. N. & Fennis, S. A. A. M. (2026). Addressing spatial correlation and system reliability in proof load testing of reinforced concrete bridges and viaducts. *Structural Safety, Under review.*

Addressing spatial correlation and system reliability in proof load testing of reinforced concrete bridges and viaducts

R. de Vries[a,b,*], E. O. L. Lantsoght[a,c], R. D. J. M. Steenbergen[b,d], M. A. N. Hendriks[a,e], S. A. A. M. Fennis[f]

[a] *Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands*
[b] *Reliable Structures, Netherlands Organisation for Applied Scientific Research (TNO), Delft, The Netherlands*
[c] *College of Sciences and Engineering, Universidad San Francisco de Quito (USFQ), Quito, Ecuador*
[d] *Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium*
[e] *Department of Structural Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*
[f] *Rijkswaterstaat, Ministry of Infrastructure and Water Management, Utrecht, The Netherlands*

Abstract

As the infrastructure ages, structural assessments may be necessary, especially when signs of deterioration are present or traffic loads have increased. Proof load testing provides a means of demonstrating sufficient reliability; however, such tests are challenging for structures with many components, critical sections, and multiple possible failure mechanisms. This study addresses these challenges through a system-reliability updating framework that accounts for spatial correlation and component similarity. A hierarchical Bayesian model is developed, in which test outcomes from a limited number of components can be used to demonstrate the reliability of the entire structure, provided that higher test loads are applied. This method is implemented through a Bayesian Monte Carlo procedure that simulates variability in resistance and load effects, while updating probability distributions based on the proof load testing outcomes. Two case studies demonstrate the approach: a simply-supported slab bridge with a varying number of spans, and a continuous two-span slab bridge, accounting for bending and shear failure. In the first case study, results are presented as transfer factors; indicating the increased target load if not all components are tested. In the second case, a model considering multiple critical sections and failure modes is used to develop optimal load testing strategies, including test locations, target loads, and vehicle configurations. The presented methodology enables efficient testing procedures and enhances decision-making in assessing existing infrastructure.

Keywords: Bayesian updating, bridges, correlation, load testing, proof load, system reliability

1. Introduction

Every day, we rely on roads, bridges and other infrastructure to reach our destinations, usually without considering their structural integrity. Safe and dependable infrastructure has become a cornerstone of modern society. However, as structures age, their ability to meet current reliability criteria should be periodically assessed. Ageing, signs of deterioration, increased loads, and outdated practices often prompt such reviews. In the assessment of existing structures, the original design reliability, based on prior knowledge, should be updated to reflect our current knowledge and the current state of the structure [1–3]. To judge if the calculated reliability is sufficient, it is essential to consider the differences in risk-based requirements between new and existing structures. For new structures, reliability targets are typically set higher than what is strictly necessary for human safety, reflecting economic and societal preferences. For existing

*Corresponding author
Email address: rein.devries@tudelft.nl (R. de Vries)

structures, reliability targets balance safety with (environmental) replacement costs, with human safety as a minimum. [4–6].

Accurate assessment methods are desired to avoid the replacement of bridges and viaducts that are, in fact, sufficiently reliable. A desk study may be performed in which the original design, possibly including vintage detailing, is assessed using the current standards, also considering possibly increased traffic loads. Where no signs of deterioration are observed, the typical desk study may confirm sufficient reliability if there are no reasons to suspect internal damage. However, such studies rely on limited information about the actual strength of a structure, which makes them generally conservative. Additionally, wear is often present, and it is difficult to tell whether it impairs structural reliability. Fortunately, tests can be carried out on the structure to gather supplementary data. In addition to material tests, diagnostic load tests involving small to moderate loads can be used to quantify structural properties. Typically, a finite element model of the structure is updated with the test results to align with the measured response and reduce model uncertainty [7,8]. However, such tests are usually limited to the linear domain and do not guarantee agreement in the non-linear range at higher load levels, limiting their value in reliability assessments. In proof load tests, larger loads are exerted to demonstrate a structure's ability to resist future loading situations [9–11]. However, applying the often large loads is typically resource-intensive and poses a risk to the structure, equipment, and personnel. To avoid excessively large loads, all relevant information about the structure should be considered—even if subject to considerable uncertainty. Given such uncertainty, probabilistic techniques are required, with Bayesian methods being particularly well suited [12–15].

Although proof load testing is a highly suitable method when doubts exist about a structure's capacity, practical and conceptual difficulties arise when a structure is viewed as a system of interrelated components. These components include not only physical elements such as the bridge deck and supports, but also cross-sections and connections. A single component may be involved in multiple potential failure mechanisms, each of which must be evaluated. This raises the question: *to what extent can one or more load tests demonstrate the reliability of the entire structure*? The proof load testing situation differs significantly from code-based structural design, where elements are typically verified at the component level. It is assumed that if each component satisfies the prescribed reliability criteria, the structure as a whole is compliant. However, the actual reliability of the system may differ significantly given its configuration [16]. Questions persist about whether structural design and assessments should adopt a system-reliability perspective and what reliability levels would be appropriate. Specifically, system-level reliability targets, often associated with severe consequences, are difficult to justify for components that only lead to local failures [3]. Nonetheless, system-reliability assessments are becoming more prevalent and address the spatial variations that are also of interest to proof load testing [17].

This research forms part of a broader effort to integrate proof load testing into a comprehensive framework for updating structural reliability, addressing time dependence, varying knowledge levels, in-situ monitoring data, spatial correlation, and system reliability [18]. The article begins by outlining the necessary background on reliability assessment through load testing, including the natural variability in shear and bending resistance, as well as the use of random fields to model spatial variations. Next, a system-reliability updating method is presented, accounting for spatial correlation and component similarity. The method is first applied to simply-supported slab bridges with a varying number of spans, and a second case study addresses a complex two-span bridge with various critical sections and multiple failure modes. Finally, key findings and unresolved challenges are discussed, followed by the conclusions.

2. Background

2.1. Reliability updating through proof load testing

The use of load testing to assess structural performance in relation to reliability was already recognised in the 1980s [19–21]. In these early studies, the outcome of a successful proof load test was incorporated through a 'cut-off' of the resistance distribution function. In contrast, proof load testing is now being considered within a broader framework of structural reliability updating [13,15], which includes time-dependent effects and in-service demonstrated strength [22,23]. During proof load testing, the reliability is generally low due to the high applied load. However, if the test is successful, it results in a significant increase in structural reliability [24]. By incorporating trends in traffic loading and resistance degradation, proof load tests can be strategically scheduled to extend a structure's service life optimally [25].

When viewed from a broader reliability updating perspective, Bayesian methods provide a coherent and flexible formulation that incorporates the information gained from surviving one or more proof load tests. Rather than focusing solely on resistance, the emphasis shifts to the limit state function, recognising that the observation provides inequality-type information [13]. In the case of proof loading, a successful test corresponds to:

$$Z_{\mathrm{PL}} = g_{\mathrm{PL}}(\mathbf{X}) > 0 \tag{1}$$

where $g_{\mathrm{PL}}(\cdot)$ indicates the limit state function and $\mathbf{X}$ is a vector containing the random variables, including (statistical) model uncertainties. This general expression accommodates various aspects that may have contributed to the survival of the test load: unexpectedly low permanent loads, modelling inaccuracies in resistance or load effects, and others. The posterior distribution is obtained by applying Bayes' rule to perform the reliability update:

$$p(\boldsymbol{\theta} \mid Z_{\mathrm{PL}} > 0) \propto p(Z_{\mathrm{PL}} > 0 \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{2}$$

where the likelihood $p(Z_{\mathrm{PL}} > 0 \mid \boldsymbol{\theta})$ is typically an indicator function, or potential [26,27], and $p(\boldsymbol{\theta})$ is the prior probability. The model parameters ($\boldsymbol{\theta}$) may directly correspond to the random variables ($\mathbf{X}$), or their parameters, depending on the modelling approach. A reliability analysis considering the regular traffic load effect can be conducted, now utilising the posterior distribution $p(\boldsymbol{\theta} \mid Z_{\mathrm{PL}} > 0)$, to calculate the updated reliability.

2.2. Transfer factors

A key concern in proof load testing is whether the tested components, cross-sections or spans can adequately represent the performance of similar components within the same structure. Selecting representative components and determining the number to test is crucial. For example, one span may be chosen to represent an entire bridge, potentially yielding conservative results if it has unfavourable geometry, detailing or condition [28]. Where components exhibit similar demand-to-capacity ratios, a more systematic approach can provide a statistical basis for addressing this uncertainty.

While proof load testing typically focuses on the performance of an individual component, spatial correlation and other forms of inter-component variability introduce complexities more commonly associated with conventional sample testing. The translation of this additional statistical uncertainty into characteristic or design values is addressed in Annex D of EN 1990 [29]. Specifically for proof load testing, a transfer factor (γ_{t}) is introduced in the most recent German guideline for load tests on concrete structures [30,31] to extrapolate test results to similar, untested components or structures. The factor increases the target load to compensate for the added uncertainty of not testing all components:

$$Q^{*}_{\mathrm{PL}} = \gamma_{\mathrm{t}} Q_{\mathrm{PL}} \tag{3}$$

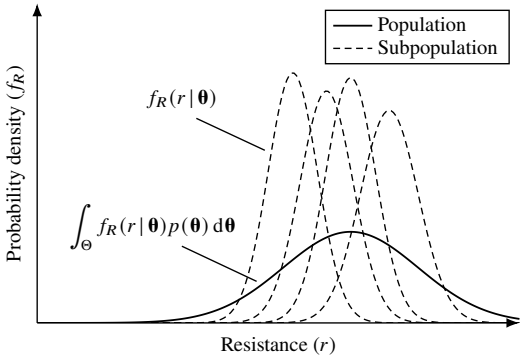


Figure 1: Hierarchical representation of resistance uncertainty: subpopulations capture the natural variability, while the population reflects both epistemic and aleatory.

where γ_t is the transfer factor, and Q_{PL} is the target load that applies to one component. The value of the transfer factor depends on the number of components tested (n), the total number of components (N), and the variation coefficient of the resistance (V_R). In the German guideline for proof load testing [30], tables are provided for common combinations of these parameters. ACI 437R-19 [32], which addresses the strength evaluation of existing concrete buildings, also recognises the need for transferring load test results to untested components, but more generally recommends statistical methods for interpreting test results. The extent to which the tested components are considered representative of the broader population is left to the engineer's judgement.

2.3. Natural variability of the resistance

Structural members exhibit variability in their resistance due to natural, mostly random, differences in composition, manufacturing processes, and interaction with other structural components. It is the variability that would persist if a manufacturer were asked to produce the same structural component twice. This type of uncertainty is also referred to as aleatory, or inherent, and it is suitable for modelling through random fields (Section 2.4). It is considered separate from model, or epistemic, uncertainty, which is attributed to a lack of knowledge. Their separation can become a topic of philosophical debate, as a perfect understanding of the natural world would allow an all-knowing entity to predict the outcomes of future events entirely [33]. In practice, with limited knowledge and resources, some natural, irreducible variability will typically remain. In many applications, an incorrect separation does not significantly influence the results—provided both are incorporated. However, the separation becomes important when some form of scaling is performed, e.g. spatially (system reliability) or in a time-variant context [34].

Within one structure, resistances are typically dependent when viewed within the overall population, for instance, due to a shared manufacturer, source materials or environmental factors. Such dependencies reduce uncertainty when extrapolating test results to untested components [19,35]. A simple way to model them is through basic correlation values and equicorrelation matrices. Alternatively, hierarchical Bayesian modelling allows these dependencies to be represented more naturally [36]. This approach has also been used to establish a *track record* of structural performance from in-service proven strength of buildings sharing specific floor detailing [37]. Hierarchical Bayesian modelling (Section 3.2) enables a clear separation of uncertainties: in this application high-level parameters capture epistemic uncertainty, while lower-level variables describe natural variability. Fig. 1 illustrates this: the solid line is the overall population (prior predictive) distribution, combining epistemic and aleatory uncertainty, while dashed lines show

Table 1: Coefficient of variations (COVs) following from repeat tests of RC beams *without* shear reinforcement.

Reference	Test specimens	Number of tests	Mean compr. strength [MPa]	COV
Bentz and Buckley [38]	SBB2.1 – SBB2.3	3	34.3	0.03
	SBB3.1 – SBB3.3	3	36.1	0.03
Bernander [39]	A1, A2, A4, A5	4	36.9	0.06
Chana [40]	2.1a - 2.3b	6	45.4	0.04
	3.3a, 3.3b, D3	3	41.6	0.11
	5.1a-5.2b	4	40.0	0.09
Collins and Kuchma [41]	B100, B100B, B100-R	3	37.0	0.10
Hallgren [42]	B90 SB5, SB6, SB9, SB10	4	39.8	0.05
	B91 SD1 – SD6	6	75.0	0.06
Leonhardt and Walther [43]	D3/1, D3/2l, D3/2r	3	45.5	0.06
	D4/1, D4/2l, D4/2r	3	41.7	0.04
Lubell et al. [44]	AT-3/N1 – AT-3/T2	4	37.4	0.04
Regan [45]	R1, R3, R7, R29	4	34.2	0.10
Sarkhosh et al. [46]	S1B1 – S1B6	6	38.2	0.05
	S3B1 – S5B5	10	45.9	0.04
Sherwood [47]	S-10N1+2, S-20N1+2, S50N1+2	6	41.4	0.03

Table 2: Coefficient of variations (COVs) following from repeat tests of RC beams *with* shear reinforcement.

Reference	Test specimens	Number of tests	Mean compr. strength [MPa]	COV
Ahmad et al. [48]	NHW-3, NHW-3a, NHW-4	3[a]	92.7	0.07
Bernhardt and Fynboe [49]	S7, S8	4[a]	78.0	0.08
Kong and Rangan [50]	S1	6	58.5	0.12
Krefeld and Thurston [51]	29a-2, 29b-2, 29f-2	3[b]	38.1	0.06
Özden [52]	T6, T7, T9	3	29.9	0.06

[a] Slightly varying shear reinforcement.
[b] Specimens with very low concrete compressive strength omitted.

subpopulation (conditional) distributions, capturing only natural variability. A subpopulation resistance may be assigned a random field to model spatial variability (Section 2.4) or used to produce realisations of resistances within the structure.

To determine suitable variation coefficients for subpopulations intended to represent the natural variability, relevant *repeat tests* were identified in the literature. Such tests are commonly included in extensive testing campaigns to verify monitoring equipment and testing procedures, and are expected to display differences only due to natural variability. The DAfStb-ACI dataset on the shear resistance of structural concrete members by committee 445-0D contains numerous test series, both with and without shear reinforcement [53–55]. For the first time, variation coefficients are derived from these data to quantify the natural variability. An overview of the suitable test series and the calculated variation coefficients is provided in Tables 1 and 2. Only test series with at least three specimens were considered since a coefficient of variation is required. The mean compressive strengths are reported to indicate the range covered by the experiments. Due to the lack of series in which all specimens were tested under identical parameters, series with slightly varying shear reinforcement have also been included. From these tables, it can be concluded that the coefficient of variation varies across cases, with a representative value of approximately 0.10, regardless of the presence of shear reinforcement. The uncertainty associated

with this value will be incorporated into the Bayesian hierarchical model (Section 3). To complete the statistical description, spatial variability must also be considered and will be addressed next.

2.4. Correlation and random fields

To complete the required background for the treatment of spatial variability and system effects, the modelling of spatial correlation through stochastic processes, or random fields, is introduced. Stochastic processes and fields mathematically describe random quantities that share dependencies, allowing them to be incorporated into structural models. Typically, one-dimensional time-dependent processes are referred to as stochastic processes, while descriptions involving spatial or multidimensional variability are referred to as random fields. These fields can represent the spatial correlation of loads, material properties, or directly their resulting resistance (Section 3.4). Gaussian processes model such phenomena by assuming that the observations are normally distributed, and fully characterised by a mean and covariance function [56]. The equations in this section are expressed as functions of time but also apply to one spatial dimension. Although some natural processes are better captured by non-normal distributions, Gaussian processes remain instrumental, as their realisations can always be related through transformation, for example using the inverse transform method [57]. The covariance function may be derived for a real-valued Gaussian process $X(t)$ by evaluating the covariance between two time points separated by lag τ, denoted as the random variables $X_1 = X(t)$ and $X_2 = X(t+\tau)$:

$$\begin{aligned} K_{XX}(\tau) &= \mathrm{Cov}(X_1, X_2) = \mathrm{E}[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= \mathrm{E}[X_1 X_2] - \mu_1\mu_2 = R_{XX}(\tau) - \mu_1\mu_2 \end{aligned} \tag{4}$$

where $\mu_1 = \mu_X(t)$ and $\mu_2 = \mu_X(t+\tau)$ refer to the mean values at the two time points, and $R_{XX}(\tau)$ denotes the autocorrelation function. The autocorrelation function provides the most complete description of the process, including information about the mean and variance. This function can be normalised to give the Pearson autocorrelation function, which is often simplified assuming constant mean and variance:

$$\rho(\tau) = \frac{R_{XX}(\tau) - \mu_1\mu_2}{\sigma_1\sigma_2} = \frac{R_{XX}(\tau) - \mu^2}{\sigma^2} \tag{5}$$

where $\sigma_1 = \sigma_X(t)$ and $\sigma_2 = \sigma_X(t+\tau)$ refer to the standard deviations at the time points. In a structural reliability context, random processes and fields are typically described through mean values and standard deviations (or variation coefficients) in conjunction with Eq. (5), which is plainly referred to as the autocorrelation function. In addition, the prefix 'auto' is often omitted, since the focus lies on the dependence, regardless of the considered quantity [34].

Various autocorrelation functions have been described in the literature and applied to describe diverse phenomena. The exponential function arises naturally in one dimension, as it corresponds to a Markov process. In higher dimensions this property is lost, but it is still frequently used, particularly in geo-statistical applications [58]. The Gaussian (or squared exponential) kernel, appears naturally in diffusion-type processes and is widely used for its smoothness, allowing infinite differentiation, and its ability to approximate more complex descriptions. The Matérn kernel offers greater flexibility by introducing a smoothness parameter (ν):

$$\text{Exponential:}\quad \rho(\tau) = \exp\left(-\frac{|\tau|}{l}\right) \tag{6a}$$

$$\text{Matérn:}\quad \rho(\tau) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{|\tau|}{l}\right)^{\nu} K_\nu\left(\sqrt{2\nu}\frac{|\tau|}{l}\right) \tag{6b}$$

$$\text{Gaussian:}\quad \rho(\tau) = \exp\left(-\frac{\tau^2}{2l^2}\right) \tag{6c}$$

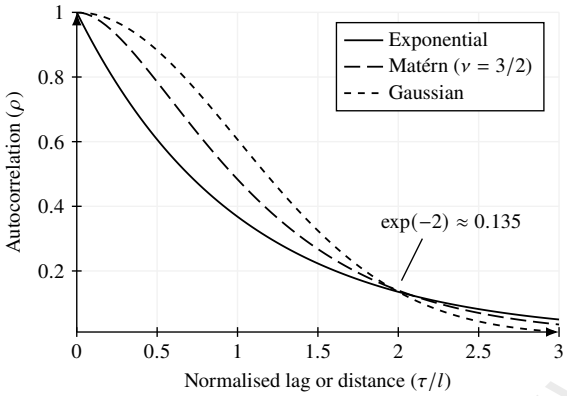


Figure 2: Comparison of the exponential, Matérn and Gaussian autocorrelation functions.

where τ is the lag, or distance between points, l is the correlation length, ν is a smoothness parameter, $\Gamma(\cdot)$ is the gamma function, and $K_\nu(\cdot)$ is the Bessel function of the second kind. The exponential and Gaussian autocorrelation functions represent two limiting cases in terms of smoothness, respectively corresponding to Matérn kernels with $\nu = 1/2$ and $\nu \to \infty$. Closed-form expressions of the Matérn kernel are available for $\nu = p + 1/2$, where p is a natural number, making these variants more common. Fig. 2 shows the normalised autocorrelation functions and highlights the intermediate smoothness of the Matérn kernel with $\nu = 3/2$. Although other kernels exist, the three described here are typical for structural reliability analyses. It is worth noting that, in some texts, the Gaussian kernel of Eq. (6c), appears without the 2 in the denominator—thus reported correlation lengths must be interpreted with care [35,59,60].

3. System-reliability evaluation method

3.1. Reliability updating given system performance

The reliability updating procedures incorporating proof load testing, as described in Section 2.1, may also be performed within a system reliability context. A series system is assumed, meaning that system failure occurs if any component fails. The analysis accounts for the survival of multiple components (or cross-sections) under a given proof load effect. Each component is characterised by its own limit state function, involving component-specific random variables; however, inter-component correlations are typically present. If the survival of n components is used as evidence to assess the failure probability of a system comprising N components, the system failure probability is expressed as:

$$P_{\mathrm{f,sys}} = \mathrm{P}\left(\bigcup_{i=1}^{N} Z_i < 0 \;\middle|\; \bigcap_{j=1}^{n} Z_{\mathrm{PL},j} \geq 0 \right) \tag{7}$$

The failure probability of a particular component, given the knowledge of a successful proof load test that involved n components, is expressed as:

$$P_{\mathrm{f},i} = \mathrm{P}\left(Z_i < 0 \;\middle|\; \bigcap_{j=1}^{n} Z_{\mathrm{PL},j} \geq 0 \right) \tag{8}$$

The component-level reliability is relevant, as it aligns with codified assessment procedures where, in principle, structural components are verified individually. The corresponding reliability indices

may be calculated via the inverse standard normal CDF, i.e. $\beta = -\Phi^{-1}(P_f)$. Evaluating Eqs. (7) and (8) generally requires numerical methods to integrate over the conditional probability space (see Section 3.3).

For bridge structures, the limit state functions during regular traffic and proof load testing conditions are formulated following JCSS [35] and fib [61] recommendations. The load effect is split into the contributions from the dead load (G_{DL}), the superimposed dead load (G_{SDL}) and the variable traffic load (Q). Both the resistance and load effect are associated with model uncertainty (θ_R and θ_E). Specific to the variable traffic load is the time-invariant uncertainty (C_{0Q}):

$$Z = g(\mathbf{X}) = \theta_R R - \theta_E(G_{\mathrm{DL}} + G_{\mathrm{SDL}} + C_{0Q}Q) \tag{9}$$

In the proof load testing situation, the variable traffic load is replaced by the load effect created by the applied test load:

$$Z_{\mathrm{PL}} = g_{\mathrm{PL}}(\mathbf{X}) = \theta_R R - [\theta_E(G_{\mathrm{DL}} + G_{\mathrm{SDL}}) + \theta_{E,\mathrm{PL}} Q_{\mathrm{PL}}] \tag{10}$$

The model uncertainty $\theta_{E,\mathrm{PL}}$, relating the proof load to the corresponding load effect, is specific to the proof load testing situation but is correlated with the regular model uncertainty θ_E, as the loads are similar and generally utilise the same calculation (FEM) model. When applied in a system-reliability context, the limit state functions in Eqs. (9) and (10) and their random variables may be annotated with the subscript i (or j) to indicate component-specific quantities; for clarity this notation is omitted.

3.2. Hierarchical Bayesian model

In an informative probabilistic setting, the resistance distribution could be inferred from existing documentation and supplementary test data. However, such detailed information is not always available, and deterioration can significantly affect the resistance. In such cases, low-informative priors may be formulated based on basic, physically justified information about the structure [62]. A hierarchical Bayesian model enables separate treatment of the high uncertainty surrounding the mean resistance and the more refined knowledge regarding its coefficient of variation. If represented by a single broad-banded distribution, these nuances would be lost.

In the proposed hierarchical Bayesian model, the resistance (R) of the considered component is assumed to follow a lognormal distribution, parametrised through a mean (m_R) and coefficient of variation (V_R) [37]:

$$R \sim \mathrm{Lognormal}(m_R, V_R) \tag{11}$$

The mean and standard deviation of the underlying normal distribution are computed as $\mu_{\ln R} = \ln(m_R^2/x)$ and $\sigma_{\ln R} = \sqrt{2\ln(x/m_R)}$, where $x = \sqrt{m_R^2 + (V_R m_R)^2}$. The parameters m_R and V_R themselves are random variables as well, also following lognormal distributions:

$$m_R \sim \mathrm{Lognormal}(m_{m_R}, V_{m_R}) \tag{12a}$$

$$V_R \sim \mathrm{Lognormal}(m_{V_R}, V_{V_R}) \tag{12b}$$

The hyperparameters $m_{m_R}, V_{m_R}, m_{V_R}$, and V_{V_R} are assigned appropriate values based on expert judgement and case-specific information. The parametrisation via mean and coefficient of variation does introduce a dependency between the parameters of the underlying normal distribution. However, this dependency is not considered problematic, as it reflects domain knowledge, and the Bayesian updating process introduces such dependencies as well. When multiple components (or cross-sections) are considered, each is assigned two random variables representing its mean and coefficient of variation.

The estimated mean resistance (m_{m_R}) can be informed by data on permanent and traffic-induced load effects, for example through WIM measurements and load effect simulations. A low-informative prior is constructed by scaling the mean load effects by a factor of 1.5, reflecting

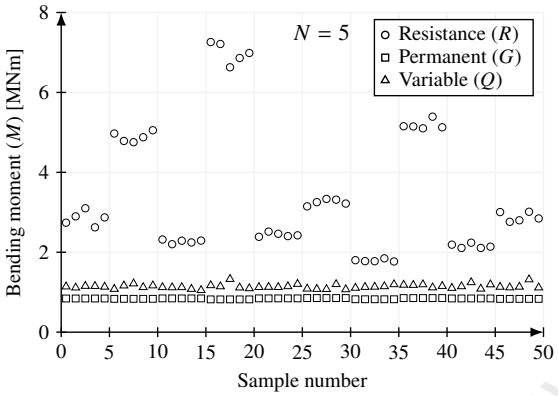


Figure 3: Example of samples obtained using a hierarchical modelling approach for the bending-moment resistance, together with permanent and variable load effects following from a single distribution.

the assumption that resistance exceeds typical annual demands. In practice, resistances tend to exceed annual load effects by a much larger margin (typically 3–4 for new designs). A high degree of uncertainty typically surrounds this estimate (e.g. $V_{m_R} = 0.5$), reflecting the low-informative nature of the prior for m_R. In contrast, the parameters associated with the coefficient of variation (V_R) may define a moderately informative distribution. The moment resistance is primarily dependent on the yield stress of the reinforcement and geometry, while other parameters, such as concrete properties, play a less significant role [63]. Thus, if bending failure governs, the mean (m_{V_R}) may be set to a low value such as 0.05, representing the typically small variation in yield stress and geometric properties. For shear failure, a higher value such as 0.10 is more appropriate (Section 2.3). Reflecting engineering knowledge of the natural variability inherent to the failure mechanism, the corresponding coefficient of variation may be relatively small (e.g. $V_{V_R} = 0.1$). An example of such a hierarchical sampling scheme is illustrated in Fig. 3, where 10 structures with $N = 5$ components were generated, resulting in 50 samples. The example illustrates how distinct groups of resistances form, with variability present within each group.

3.3. Calculation procedure

Various methods can be used to calculate the posterior distribution, incorporating observed system performance and the resulting reliability of the system or its components. Given the large number of random variables and distribution types, these methods are typically numerical in nature. To obtain the joint posterior distribution, several approaches may be employed, such as the Bayesian Monte Carlo (BMC) method [64,65] or Markov-chain Monte Carlo (MCMC) [66,67]. In this application, the BMC method is employed due to its simplicity. Since all prior distributions are moderately informative, no difficulties arise from improper or overly broad distributions. While there are concerns about the efficiency and accuracy of BMC [68], no such issues were experienced because the procedure was implemented in C++, a programming language designed for high-performance computing [69].

Using the Bayesian Monte Carlo method, the accepted (posterior) samples may be directly used to evaluate the reliability of the structure. This simplifies the calculation, such that effectively just one integral needs to be evaluated. For example, the failure probability of component i, given a successful proof load test of the system, may be evaluated via:

$$P_{\mathrm{f},i} = \int_{\boldsymbol{\Theta}} \mathbb{I}[g_i(\boldsymbol{\theta}) < 0] p\left(\boldsymbol{\theta} \;\middle|\; \bigcap_{j=1}^{n} g_{\mathrm{PL},j}(\boldsymbol{\theta}) \geq 0\right) \mathrm{d}\boldsymbol{\theta} \tag{13}$$

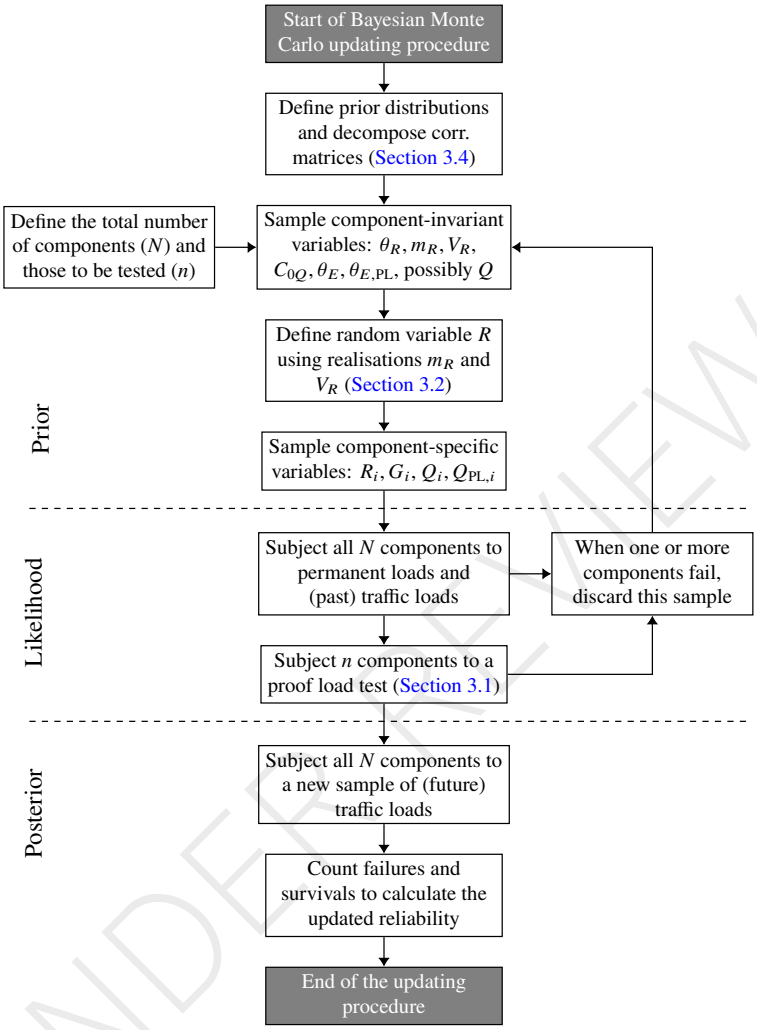


Figure 4: Flowchart outlining the steps in the proposed Bayesian Monte Carlo procedure to update the system reliability.

where $\mathbb{I}[\cdot]$ is the indicator function, $p(\boldsymbol{\theta} \mid \cdot)$ is the posterior distribution, and $\boldsymbol{\theta} = (\theta_R,\ m_R,\ V_R,\ R_1,\ R_2,\ \ldots)$ carries the entire state of the system, including unconditioned predictive variables such as the future traffic loads. If the system failure probability is desired, the indicator function would consider the minimum of all limit state functions $g_i(\cdot)$. To acquire a more realistic posterior distribution, the survival of one year of traffic load may also be included. This inclusion can be viewed as sequential Bayesian updating [70], in which the posterior obtained after accounting for the survival of past traffic loads serves as the prior for the survival update of the proof load test. A flowchart outlining the Bayesian Monte Carlo calculation procedure, assuming the same statistical description for all components N of which n are tested, is provided in Fig. 4.

3.4. Simulation of correlated variables and random fields

When evaluating system reliability, it is necessary to consider the correlation between quantities used in the probabilistic description of a structure. As described in Section 2.4, these quantities may be modelled as random fields but must be discretised for use in evaluating local limit state functions. This applies to resistance, permanent loads, and variable (traffic) loads. Typically, the (statistical) model uncertainties ($\theta_R, \theta_E, C_{0Q}$) can be assumed to be fully correlated across components, as they stem from a common schematisation or modelling approach [71]. The correlation between resistances depends on the failure mechanism considered. Regarding shear resistance, the material properties of concrete govern and may be assumed to be uncorrelated, except when cross-sections are very close. For bending resistance, the degree of correlation depends primarily on whether the same reinforcement bars are present, as their strength is the dominant influence (Section 5).

The discretised correlations, derived from the evaluation of autocorrelation functions or otherwise, can be represented using correlation matrices, for example:

$$\mathbf{R}^{(2)} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}, \quad \mathbf{R}^{(3)} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \tag{14}$$

where ρ_{ij} denote the Pearson correlation coefficients between variables. Correlation matrices are symmetric, reflecting the symmetry of autocorrelation functions. To sample correlated variables, a correlation matrix may be decomposed such that multiplication by a vector of independent standard normally distributed variables, $\mathbf{u}$, yields the desired correlated vector $\mathbf{u}_c$. If the correlation matrix is positive definite, it can be decomposed using the Cholesky decomposition [72]:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^{\mathrm{T}}, \quad \mathbf{u}_c = \mathbf{L}\mathbf{u} \tag{15}$$

where $\mathbf{L}$ is the lower-triangular matrix resulting from the decomposition. The final step involves transforming the standard normally distributed variables to the desired distributions, for example, via the inverse transform method [34].

Although alternative decomposition procedures exist, such as eigenvalue decomposition, the Cholesky method has the advantage of producing a lower-triangular matrix. Each element in the correlated vector thus depends only on preceding elements, which facilitates hierarchical sampling procedures, such as separate sampling of global and local correlations. This approach offers the advantage that only smaller, more manageable correlation matrices are used. In the proposed Bayesian Monte Carlo procedure, this computational efficiency is highly beneficial. The application of such a scheme is detailed in Section 5.

3.5. Load correlation decay with increasing reference period

In analysing the load effects derived from weigh-in-motion (WIM) data [73], it was found that the calculated Pearson correlation coefficients between month maxima were systematically lower than those between week maxima. A similar reduction was observed when comparing weekly with daily maxima, and this trend persisted across different locations and load effects (bending moment and shear). These observations are consistent with extreme value theory, which states that dependencies between block maxima generally weaken with increasing block size [74,75].

For structural reliability calculations, reference periods longer than just weeks or months are typically considered. Therefore, it is necessary to predict how correlation decays as the reference period increases. To this end, Monte Carlo simulations were conducted by sampling many pairs of correlated standard normal random variables, i.e. $(U_1, \rho U_1 + \sqrt{1-\rho^2}U_2)$, where U_1 and U_2 are independent standard normal variables. These pairs were transformed using an extreme value distribution, after which the maximum per block (and per variable) of n values was extracted. The resulting maxima were then transformed back to standard normal variables to compute the remaining correlation. These simulations employed the generalised extreme value

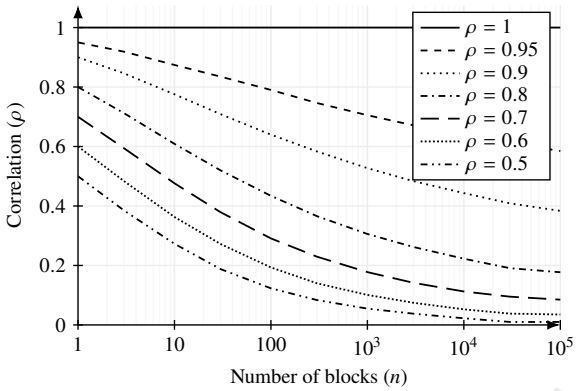


Figure 5: Decay of the Pearson correlation coefficient between normalised block maxima as a function of the number of blocks, for various initial correlation values.

distribution, with varied parameters—although these variations did not change the result. A graph demonstrating the decay of correlation with increasing reference period was produced by considering multiple initial correlation coefficients and various block sizes (Fig. 5). For example, when assessing annual reliability with load effects based on weekly maxima, the number of blocks is $n = 52$.

In the proposed methodology for addressing spatial variability in proof load testing, the results described above apply to the statistical analysis phase of the load effect. Although the reference period may be chosen arbitrarily, a week represents a suitable short cycle in which traffic conditions generally repeat, and the maxima can be assumed to be independent. The following procedure is suggested to acquire a spatially dependent description of the load effect:

1. Obtain the weekly maxima for various locations and relevant failure mechanisms, such as bending and shear, from load effect simulations based on WIM data, or artificially generated vehicle streams.
2. Fit extreme value distributions to the maxima, with particular attention to the right tail, to ensure that the distributions accurately represent the load effects for small exceedance probabilities.
3. Calculate the Pearson correlation coefficient between all relevant load effects after transforming the weekly maxima to standard normal distributions (with accurate fits), or directly by filtering outliers and assuming normality of the maxima.
4. Verify the correlation coefficients by comparing scatter plots of one load effect versus the other, using both observed and simulated data based on the fitted extreme value distributions and the computed correlation.
5. Apply correlation reduction using the graph of Fig. 5, where the number of blocks corresponds to the target reference period (e.g. $n = 52$ when scaling up from week to year maxima).

To sample the spatially dependent load effect description, the correlation coefficients may be assembled into correlation matrices. These matrices can then be decomposed, e.g. via Cholesky or eigendecomposition methods, to produce matrices suitable for generating correlated vectors of standard normal variables. Finally, these standard normal vectors must be transformed to the appropriate extreme value distributions, i.e. those obtained from the statistical fitting procedure.

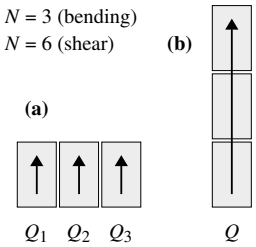


Figure 6: Schematic top view of slab bridges composed of three spans, with arrows indicating the traffic lanes: (a) parallel to each other (uncorrelated), and (b) sequentially across spans (correlated).

4. Simply-supported slab bridges with a varying number of spans

4.1. Description and objectives

To illustrate the application of the proposed method (Section 3), a case study on simply-supported slab bridges is presented. Such bridges and viaducts, often with short spans, are common in the Netherlands and many other countries [76,77]. In this case study, bridges composed of multiple separate slabs are considered. These configurations are less common than continuous-span bridges (Section 5), but provide a clear basis. The reinforced concrete slabs have a length of 10 m and a width of 3.5 m, representing a conservative single-lane scenario without load sharing between lanes. Assuming that the substructure has sufficient capacity, only the bending moment and shear resistance require evaluation. Given the simply-supported configuration, the maximum bending moment typically occurs at mid-span, while the maximum shear force is expected near the supports. When the critical cross-sections are treated as components, two shear-critical sections exist per span, which is relevant for defining the number of components (N) and sample size (n). Note that the quantification of the target load is not the objective here, but to determine the transfer factor (Section 2.2) when only a subset of components is tested.

Two variable load configurations are considered in the following analysis: one in which the variable load effect is independent across components, and a configuration where the load effects are dependent (i.e. fully correlated). Applied to the simply-supported slab bridges, these configurations correspond to (a) multiple parallel spans and (b) a bridge with sequential spans (Fig. 6). If the traffic load effect originates from multiple lanes, the correlation will generally decrease as vehicles travel at different speeds and change lanes. The situations considered in the following analyses, i.e. independent and fully correlated, represent the two extremes.

4.2. Probabilistic model

The resistance is incorporated as a random variable (R), modelled hierarchically with a low-informative prior for its mean, and a moderately informative prior distribution for the variation coefficient. The prior distribution parameters are assigned in line with the reasoning described in Section 3.2. The mean values of the resistance and permanent load effect were normalised to the mean traffic load effect since no absolute target loads, but rather factors, are calculated (Table 4).

Given the simply supported schematisation, the maximum bending moment occurs at mid-span, so each span is treated as one component. Additional cross-sections within a slab are unnecessary, as moment resistance is highly correlated where the same reinforcement bars are utilised. The moment resistance is regarded uncorrelated between spans, since they are different structural members. This independence refers to natural variability, represented by the low-level resistance variable R in the hierarchical model. Correlation arising from shared production processes, manufacturers, or source materials is captured through component-invariant model parameters at the highest level of the hierarchical model. The maximum shear force occurs near

Table 3: Overview of random variables for the simply-supported slab bridges case study, indicating their distribution, mean value, coefficient of variation (COV), and the Pearson correlation coefficient between components.

Var.	Description	Distribution	Mean	COV	Comp. corr.
θ_R	Model uncertainty of the resistance	Lognormal	1.0	[a]	1
R	Resistance (hierarchical Bayesian model)	Lognormal	m_R	V_R	0
m_R	Mean of the resistance	Lognormal	[a]	0.5	1
V_R	COV of the resistance	Lognormal	[a]	0.1	1
θ_E	Model uncertainty of the load effect	Lognormal	1.0	0.1	1
G	Permanent load effect	Normal	[a]	0.07	0.7
C_{0Q}	Time-independent uncertainty of the traffic load	Lognormal	1.0	0.1	1
Q	Traffic load effect, annual maximum	Gumbel	1.0	[a]	[a]
$\theta_{E,\mathrm{PL}}$	Model uncertainty of the proof load effect $\rho(\theta_E, \theta_{E,\mathrm{PL}}) = 0.7$	Lognormal	1.0	0.1	1
Q_{PL}	Load effect achieved by proof load (target load)	Normal	(varies)	0.02	0

[a] Values depend on the failure mechanism and variable load dependency; see Table 4.

Table 4: Situation-specific parameters depending on the failure mechanism and variable-load correlation for the simply-supported case.

Situation	V_{θ_R}	m_{m_R}	m_{V_R}	m_G	V_Q	ρ_Q
Bending, parallel	0.05	2.55	0.05	0.7	0.025	0
Bending, sequential	0.05	2.55	0.05	0.7	0.025	1
Shear, parallel	0.1	2.7	0.1	0.8	0.035	0
Shear, sequential	0.1	2.7	0.1	0.8	0.035	1

E

the supports, where each end is taken as one component. Although small correlation lengths of concrete properties [60] would motivate evaluating multiple sections, the shear force decreases rapidly in short spans, limiting the shear-critical region. Consequently, only one cross-section is considered at each end, with natural variability uncorrelated and epistemic correlations again represented through the hierarchical model.

Based on a previously studied case [25], the normalised mean permanent bending moment is 822 kNm / 1150 kNm ≈ 0.7, and the permanent shear force is 340 kN / 420 kN ≈ 0.8. Following JCSS [35], a between-component correlation of 0.7 applies, which can be incorporated via a $N \times N$ equicorrelation matrix. Ideally, a hierarchical modelling scheme would also be adopted for the permanent load effect to separate epistemic from aleatory uncertainty (see Section 6). The statistical descriptions of the traffic load effect were derived from load effect simulations based on WIM data from the Netherlands. The resulting coefficients of variation are relatively low, owing to the short span length (10 m) and the high frequency of heavy trucks [78]. The parameters of the remaining random variables in the probabilistic model follow the recommendations of JCSS [35] and fib [61]. No bias to account for the dynamic effects is included in factor C_{0Q} since its inclusion leads to lower transfer factors. An overview of the random variables for the current case study is provided in Table 3, which refers to Table 4 for specific parameter variations depending on the failure mechanism and variable load dependency.

4.3. Results

Given a particular situation (Table 4), number of components (N), sample size (n), and target load ($m_{Q,\mathrm{PL}}$), the system reliability of a bridge comprising multiple spans may be evaluated (Section 3.3). The annual reliability target adopted in this case study is $\beta = 4$, which aligns with the requirements for existing consequence class 3 (CC3) structures in the Netherlands [23]. Modifying this reliability target has little impact on the resulting transfer factors, as these are

Table 5: Transfer factors calculated for the simply-supported case study, considering bending failure.

Number of comp. (N)	Sample size (n)						
	1	2	3	4	5	6	8
1	1	-	-	-	-	-	-
2	1.13	1	-	-	-	-	-
3	1.17	1.07	1	-	-	-	-
5	1.21	1.11	1.07	1.03	1	-	-
8	1.23	1.14	1.11	1.08	1.05	1.03	1
10	1.25	1.16	1.12	1.09	1.07	1.05	1.02
20	1.28	1.19	1.15	1.13	1.11	1.10	1.07

Table 6: Transfer factors calculated for the simply-supported case study, considering shear failure.

Number of comp. (N)	Sample size (n)						
	1	2	3	4	5	6	8
1	1	-	-	-	-	-	-
2	1.50	1	-	-	-	-	-
3	1.58	1.30	1	-	-	-	-
5	1.66	1.40	1.27	1.16	1	-	-
8	1.73	1.47	1.34	1.27	1.21	1.15	1
10	1.76	1.49	1.38	1.30	1.25	1.20	1.12
20	1.83	1.56	1.44	1.38	1.33	1.29	1.24

defined as ratios. Since the target load cannot be derived analytically, a discrete set of target load values is evaluated, and the resulting reliability indices are interpolated to identify the load level that satisfies the reliability target. The corresponding transfer factor may then be calculated by dividing the system target load by the load level that would be required for just a single component [30].

It is found that the inclusion of variable-load dependency has a negligible effect on the reliability index for both failure mechanisms, indicating factors increased by 0.01 in some cases when assuming independent variable loads. Therefore, the resulting transfer factors are presented in Tables 5 and 6 for bending and shear, respectively, without distinguishing between correlated and uncorrelated traffic loads. As expected, transfer factors for shear are generally higher, reflecting the larger variability associated with this failure mechanism. In the considered case study, each span includes two critical cross-sections, effectively doubling the number of components (N). As a result, the transfer factor increases rapidly when the number of test positions is limited, illustrating the reduced representativeness of the test results.

For reference, the results are compared with values from the German guideline [30], which provides transfer factors for generic situations, focusing solely on resistance. For bending, the factors calculated in this case study are slightly lower than those reported in Table 2 of the German guideline. For shear, the values are comparable if the presence of shear reinforcement is assumed. In this study, no significant difference in the variation coefficient was found between cases with or without shear reinforcement (Section 2.3). Although the numerical values are similar, the underlying methodology differs substantially. Notably, the current approach explicitly considers the survival of one year of traffic loading, which significantly reduces the resulting transfer factors. Furthermore, this case study focuses on short-span slabs, their specific traffic loading, and the associated permanent-to-variable load ratio. The presented results are thus bridge and configuration-dependent.

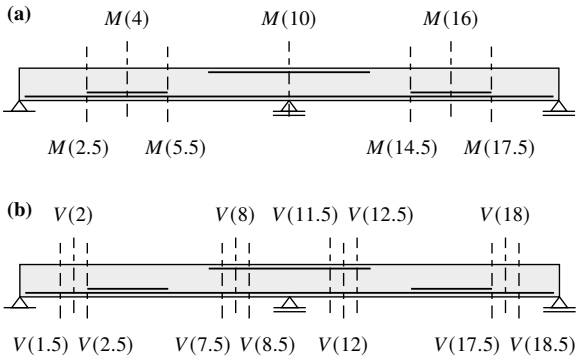


Figure 7: Schematic side view of the continuous slab bridge, displaying the main reinforcement and the (a) bending and (b) shear cross-sections of interest.

5. Multiple spans, cross-sections and failure modes

5.1. Description and objectives

This section presents a more complex application of the proposed method (Section 3), assessing a continuous slab bridge while accounting for spatial variability across multiple cross-sections and failure modes. Unlike the previous case study, no generalisation in terms of the number of components (N) and tested samples (n) is applicable. The objective is to quantify system-level reliability for this specific bridge configuration in bending and shear through proof load testing, considering uncertainties in material properties, reinforcement layout, and traffic loading. By treating its cross-sections as a series system, optimal testing strategies are developed, including the configuration of a dedicated load-testing vehicle.

A hypothetical concrete slab bridge with two equal spans of 10 meters is considered, comprising a single traffic lane. The critical cross-sections are defined at multiple positions, considering bending moments and shear forces (Fig. 7). The mid-span cross-sections at $x = 4$ m and 16 m are known to contain additional reinforcement (as could be determined from scans), while the rebar properties (diameter, grade, and condition) remain uncertain. For this reason, the cross-sections where the additional reinforcement ends are of particular interest since they introduce a discontinuity in the moment resistance.

5.2. Resistance modelling

Following Section 3.2, a hierarchical model is adopted for the definition of the resistances. The random variables $m_{M_{Ri}}$, V_{M_R}, $m_{V_{Ri}}$, and V_{V_R} represent parameters that are unknown due to incomplete knowledge. Once values are assumed for these parameters, considering prior probabilities, lower-level random variables M_{Ri} and V_{Ri} can be defined to model the aleatory uncertainty. Since these variables describe natural variations, spatial correlations apply to them. Three bending resistances and two shear resistances are defined. The moment resistances correspond to cross-sections with and without additional reinforcement, as well as the support moment resistance. Shear resistances near the end supports are distinguished from those near the middle support, as the latter may benefit from increased dowel action provided by the top reinforcement (Table 7). The prior distribution parameters are specified according to the rationale presented in Section 3.2. For simplicity, the mean shear resistance is assumed to remain constant over 1.5 m, although in reality a slight decrease might occur due to the increasing bending moment [79].

In the current numerical implementation, the random fields of moment and shear resistance (M_{Ri} and V_{Ri}) are discretised to obtain cross-sectional values (Table 8). No correlations are specified between moment and shear resistances, as they are assumed to stem from different

Table 7: Overview of random variables for the continuous slab bridge case study, indicating their distribution, mean value, coefficient of variation (COV), and the Pearson correlation coefficient between components.

Var.	Description	Distribution	Mean	COV	Comp. corr.
θ_{M_R}	Model uncertainty of the moment resistance	Lognormal	1.0	0.05	1
θ_{V_R}	Model uncertainty of the shear resistance	Lognormal	1.0	0.1	1
M_{Ri}	Bending moment resistance (hierarchical Bayesian model)	Lognormal	$m_{M_{Ri}}$	V_{M_R}	[b]
V_{Ri}	Shear resistance (hierarchical Bayesian model)	Lognormal	$m_{V_{Ri}}$	V_{V_R}	[b]
$m_{M_{R1}}$	Mean of the span moment resistance, incl. additional reinforcement	Lognormal	2031 kNm	0.5	1
$m_{M_{R2}}$	Mean of the span moment resistance, excl. additional reinforcement	Lognormal	1777 kNm	0.5	1
$m_{M_{R3}}$	Mean of the support moment resistance (absolute)	Lognormal	2561 kNm	0.5	1
$m_{V_{R1}}$	Mean of the shear resistance near the end supports	Lognormal	767 kN	0.5	1
$m_{V_{R2}}$	Mean of the shear resistance near the middle support	Lognormal	1157 kN	0.5	1
V_{M_R}	COV of the bending moment resistance	Lognormal	0.05	0.1	1
V_{V_R}	COV of the shear resistance	Lognormal	0.1	0.1	1
θ_E	Model uncertainty of the load effect	Lognormal	1.0	0.1	1
G	Permanent load effect	Normal	[a]	0.07	0.7
C_{0Q}	Time-independent uncertainty of the traffic load, including bias for dynamic load effect	Lognormal	1.1	0.1	1
Q	Traffic load effect, annual maximum	Gumbel	[a]	[a]	[b]
$\theta_{E,\mathrm{PL}}$	Model uncertainty of the proof load effect; $\rho(\theta_E, \theta_{E,\mathrm{PL}}) = 0.7$	Lognormal	1.0	0.1	1
Q_{PL}	Load effect achieved by proof load (target load)	Normal	(varies)	0.02	0

[a] Various values apply depending on the cross-section location and failure mechanism; see Table 8.
[b] The spatial correlations of the resistance traffic load effects are detailed in Sections 5.2 and 5.3.

physical mechanisms. Following the recommendations of JCSS [35], high correlation is assumed between moment resistances at locations sharing reinforcement within one member. A correlation coefficient of $\rho = 0.9$ is assumed where identical rebars are present. Between other cross-sections within the same member but with different bar diameters, a coefficient of $\rho = 0.4$ applies. Since span cross-sections with and without additional reinforcement share several bars, an intermediate value of $\rho = 0.7$ is used.

To avoid expensive computations involving large correlation matrices, a hierarchical correlation scheme is employed (Section 3.4). Standard normal variables are first sampled for sufficiently distinct 'global' cross-sections, with local variations generated thereafter. By treating the shear resistances as uncorrelated (Section 3.4), only a correlation matrix for vector $[M_R(4),\ M_R(10),\ M_R(16)]$ is required:

$$\mathbf{R}_R^{(\mathrm{g})} = \begin{bmatrix} 1 & (\mathrm{sym.}) & \\ 0.4 & 1 & \\ 0.4 & 0.4 & 1 \end{bmatrix} \tag{16}$$

In both spans, the local variations described by vectors $[M_R(4),\ M_R(2.5),\ M_R(5.5)]$ and $[M_R(16),\ M_R(14.5),\ M_R(17.5)]$ are sampled using the following local correlation matrix:

$$\mathbf{R}_R^{(\mathrm{l})} = \begin{bmatrix} 1 & (\mathrm{sym.}) & \\ 0.7 & 1 & \\ 0.7 & 0.9 & 1 \end{bmatrix} \tag{17}$$

Table 8: Definition of parameters per component (cross-section) for the continuous slab bridge case study. Bending moments are provided in kNm and shear forces in kN.

Comp.	Resist.	m_G	m_Q	V_Q	$Q_{\mathrm{k,WIM}}$	$Q_{\mathrm{k,LM1}}$
$M(2.5)$	M_{R2}	405	789	0.025	886	1100
$M(4)$	M_{R1}	453	901	0.025	1012	1270
$M(5.5)$	M_{R2}	356	819	0.025	920	1169
$M(10)$	$-M_{R3}$	−810	−897	0.02	−986	−906
$M(14.5)$	M_{R2}	356	819	0.025	920	1169
$M(16)$	M_{R1}	453	901	0.025	1012	1270
$M(17.5)$	M_{R2}	405	789	0.025	886	1100
$V(1.5)$	V_{R1}	146	365	0.03	419	505
$V(2)$	V_{R1}	113	329	0.03	378	455
$V(2.5)$	V_{R1}	81	298	0.03	342	406
$V(7.5)$	$-V_{R2}$	−243	−387	0.02	−425	−568
$V(8)$	$-V_{R2}$	−275	−425	0.02	−467	−608
$V(8.5)$	$-V_{R2}$	−308	−463	0.02	−509	−645
$V(11.5)$	V_{R2}	308	463	0.02	509	645
$V(12)$	V_{R2}	275	425	0.02	467	608
$V(12.5)$	V_{R2}	243	387	0.02	425	568
$V(17.5)$	$-V_{R1}$	−81	−298	0.03	−342	−406
$V(18)$	$-V_{R1}$	−113	−329	0.03	−378	−455
$V(18.5)$	$-V_{R1}$	−146	−365	0.03	−419	−505

E

5.3. Load effects

Permanent load effects for bending (M_G) and shear (V_G) are defined for each cross-section. Their mean values follow from geometric properties of the bridge combined with an assumed concrete density of 2400 kg/m^3 and an effective lane width of 3 m. In line with the previous case study, a variation coefficient of 0.07 is used and a correlation between cross-sections of 0.7, which may be introduced through an equicorrelation matrix. As for the resistances, also for the permanent load effect, a hierarchical sampling strategy is adopted to reduce the computational effort.

The statistical description of the variable load effects (M_Q and V_Q) are obtained from WIM-based traffic simulations. They are found to be well-approximated with a Gumbel distribution [78], of which the parameters (mean and variation coefficient) vary per cross-section (Table 8). The characteristic traffic load effects ($Q_{\mathrm{k,WIM}}$), used for comparisons later, are derived from the 0.999 fractile of the fitted Gumbel distributions describing annual maxima. Eurocode values ($Q_{\mathrm{k,LM1}}$) are also considered; they are found to be 20–30% higher than the WIM-based estimates for most sections, but approximately 10% lower for the support moment. As a result, target load factors based on Eurocode characteristic load effects will show disproportionately high values for assessing the support moment, compensating for the poor representation of actual traffic loads.

Spatial traffic load correlations were also determined using the WIM data, accounting for the decay of correlation with increasing reference period (Section 3.5). No significant correlations were identified involving the middle support moment and shear forces near the end supports. Therefore, the global correlation matrix only considers [$M_Q(4)$, $M_Q(16)$, $V_Q(8.5)$, $V_Q(11.5)$]:

$$\mathbf{R}_Q^{(\mathrm{g})} = \begin{bmatrix} 1 & & & (\mathrm{sym.}) \\ 0.5 & 1 & & \\ 0.35 & 0.5 & 1 & \\ 0.5 & 0.35 & 0.35 & 1 \end{bmatrix} \tag{18}$$

The local variations in bending moments, [$M_Q(4)$, $M_Q(2.5)$, $M_Q(5.5)$] and [$M_Q(16)$, $M_Q(14.5)$, $M_Q(17.5)$], as well as shear forces, [$V_Q(1.5)$, $V_Q(2)$, $V_Q(2.5)$], [$V_Q(8.5)$, $V_Q(8)$, $V_Q(7.5)$], and so on, are achieved via:

$$\mathbf{R}_{M_Q}^{(\mathrm{l})} = \begin{bmatrix} 1 & & (\text{sym.}) \\ 0.85 & 1 & \\ 0.85 & 0.82 & 1 \end{bmatrix}, \quad \mathbf{R}_{V_Q}^{(\mathrm{l})} = \begin{bmatrix} 1 & & (\text{sym.}) \\ 0.85 & 1 & \\ 0.82 & 0.85 & 1 \end{bmatrix} \tag{19}$$

5.4. Proof load testing strategies and results

5.4.1. Separate tests

With the completed model description of the resistance and regular load effects, the proof loading strategy can be defined. In a classical testing approach, each critical section would be considered individually through specific load placement and corresponding target load. The Eurocode LM1 tandem, comprising two axle loads of 300 kN spaced 1.2 m apart, is commonly used in Europe to mimic heavy vehicles [10,11]. Other axle configurations and spacings can also be employed to achieve desired load distributions. To verify that each critical region has sufficient resistance, this tandem load can be applied at an appropriate location identified via structural analysis or engineering judgement. The separate tests are then executed, gradually incrementing the load until the target is reached.

By using the developed model, the tandem locations and target loads for these separate tests can be determined. This is an iterative process because the Bayesian Monte Carlo method requires the target load as input. In this basic strategy, testing starts with the tandem at (1.5, 2.9) m from the left support to assess shear capacity. The tandem is then moved to (3, 4.2) m for the second test, to check bending resistance, and subsequently to (6.8, 8) m to assess shear resistance near the middle support. A factor of 1.45 applied to the LM1 tandem loads is sufficient to reach an annual reliability level of $\beta = 4$, including survival of one year of traffic as in-service proven strength (Section 3). In the fourth test for the support moment, it is preferable to use two tandems at (5.8, 7) m and (13, 14.2) m, with a lower tandem factor of 1.25. Tests five to seven are then conducted on the second span, with testing locations mirrored to the first span. The corresponding target load factors range from about 1.1 to 1.5, considering WIM data ($m_{Q,\mathrm{PL}}/Q_{\mathrm{k,WIM}}$), depending on the cross-section. If compared to Eurocode LM1 characteristic loads ($m_{Q,\mathrm{PL}}/Q_{\mathrm{k,LM}}$) they range from 0.9 to 1.3—except for the support moment, where the factor is 1.5 (see Section 5.3).

The calculated reliability indices are presented in Table 9, showing values after one year of traffic loads, after traffic plus individual tests, and after all tests have been survived. After the first test (β_1), the indices for cross-sections $V(1.5)$, $V(2)$, and $V(2.5)$ increase significantly, as intended. Due to system reliability analysis, including correlations, other cross-sections also benefit. After the second test (β_2), the reliability of $M(2.5)$, $M(4)$, and $M(5.5)$ improves notably, and so on. It should be noted that test loads have been chosen such that component reliability values are adequate—in accordance with current practice—while the system reliability is lower (also see Section 6).

The reliability values after all proof load tests have been survived are slightly lower than those from individual tests, which may seem counterintuitive. This phenomenon is caused by the system-wide conditioning of the hierarchical resistance model. The posterior distributions from individual tests typically still include irrational realisations of random fields that would not pass proof loading at other locations. Therefore, the posterior distribution obtained through system-wide updating is more realistic than those from isolated tests, and less sensitive to the shape of the prior distribution. The probability of failure during testing can also be calculated, but with low-informative distributions it is generally high. Formulating more informative resistance priors can significantly lower this failure probability [25].

Table 9: Reliability indices for the continuous slab bridge using the separate tests strategy. Reliability index β_0 after one year of traffic loads; β_i after individual tests, and β_{all} after surviving all tests.

Comp.	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_{all}
$M(2.5)$	2.7	3.9	**4.1**	2.8	2.7	2.7	3.6	3.5	4.0
$M(4)$	2.6	2.7	**4.6**	2.6	2.6	2.6	4.0	2.7	4.4
$M(5.5)$	2.8	3.9	**4.0**	3.0	2.9	2.9	3.7	3.6	4.0
$M(10)$	2.6	2.6	2.6	2.6	**4.2**	2.6	2.6	2.6	4.0
$M(14.5)$	2.8	3.6	3.7	2.9	2.9	3.0	**4.0**	3.9	4.0
$M(16)$	2.6	2.7	4.0	2.6	2.6	2.6	**4.6**	2.7	4.4
$M(17.5)$	2.7	3.5	3.6	2.7	2.7	2.8	**4.1**	3.9	4.0
$V(1.5)$	2.5	**4.7**	3.3	2.6	2.6	2.6	2.9	3.4	4.5
$V(2)$	3.1	**4.1**	4.1	3.2	3.2	3.2	3.6	4.0	4.4
$V(2.5)$	3.9	**4.8**	> 5	4.0	4.0	4.0	4.4	4.8	> 5
$V(7.5)$	3.7	3.7	3.8	**4.4**	5.1	4.4	3.7	3.7	> 5
$V(8)$	3.2	3.2	3.2	**5.0**	4.4	3.8	3.2	3.2	4.7
$V(8.5)$	2.8	2.8	2.8	**4.3**	3.7	3.4	2.8	2.8	4.1
$V(11.5)$	2.8	2.8	2.8	3.4	3.7	**4.3**	2.8	2.8	4.1
$V(12)$	3.2	3.2	3.2	3.8	4.4	**5.0**	3.2	3.2	4.9
$V(12.5)$	3.7	3.7	3.7	4.4	5.1	**4.4**	3.8	3.7	4.9
$V(17.5)$	3.9	4.7	4.5	4.0	4.0	4.0	> 5	**4.8**	> 5
$V(18)$	3.1	4.1	3.6	3.2	3.2	3.2	4.2	**4.1**	4.4
$V(18.5)$	2.5	3.4	2.9	2.6	2.6	2.6	3.3	**4.7**	4.6
System	1.7	2.1	2.2	1.9	1.9	1.9	2.2	2.1	3.5

5.4.2. Optimised locations and target loads

Following the rationale of Section 4, the number of tests can be reduced by accepting higher target loads—at the cost of a greater risk of failure or damage during testing. If it is undesired to test both spans, raising the LM1 tandem factor from 1.45 to 1.8 ($\gamma_t = 1.24$) provides sufficient reliability. In this scenario, the factor for assessing the support moment capacity remains the same (1.25).

For complex bridge configurations, the minimum number of trucks and load cases that simultaneously meet the target load requirements of all critical sections can be determined via optimisation [80]. Applying such a strategy for the current case study, an optimum at (2.5, 3.7) m was identified that can verify both the span moment and shear resistance near the left support. A modest increase of the tandem load factor from 1.45 to 1.55 is required ($\gamma_t = 1.07$) to yield sufficient reliability in both failure modes. Next, two tandems at (6.3, 7.5) m and (12.5, 13.7) m can be used to verify support moment and shear resistance on either side of the support. Here, increasing the load factor from 1.25 to 1.35 ($\gamma_t = 1.08$) eliminates the need for separate shear tests. The third test is conducted on the second span, with testing locations mirrored to the first span. Thus, with just three tests and slightly higher proof loads, sufficient reliability can be demonstrated (Table 9).

5.4.3. Moving vehicle

Instead of complex test set-ups involving loading frames, hydraulic jacks and counterweights, it is preferable to use a single load testing vehicle. As the vehicle gradually moves over the bridge, many loading positions are examined, and several passes can be made to capture loading eccentricities. A simulation can then determine the resulting load effect envelope for the relevant cross-sections. Two suitable vehicle configurations have been identified:

Table 10: Reliability indices for the continuous slab bridge following from optimised location testing and vehicle runs. Definitions as in Table 9, now including β_{tip} tipper truck, and β_{tra} semi-low trailer results.

Comp.	β_0	β_1	β_2	β_3	β_{all}	β_{tip}	β_{tra}
$M(2.5)$	2.7	**4.8**	2.7	4.3	5.0	4.2	4.6
$M(4)$	2.6	**4.3**	2.6	3.9	4.2	4.0	4.5
$M(5.5)$	2.8	**4.8**	2.9	4.4	5.0	4.3	4.7
$M(10)$	2.6	2.6	**4.2**	2.6	4.1	4.4	4.1
$M(14.5)$	2.8	4.4	2.9	**4.8**	5.0	4.2	4.8
$M(16)$	2.6	3.9	2.6	**4.3**	4.2	4.0	4.5
$M(17.5)$	2.7	4.3	2.7	**4.8**	5.0	4.1	4.7
$V(1.5)$	2.5	**4.1**	2.6	3.3	4.0	4.2	4.0
$V(2)$	3.1	**4.9**	3.2	4.0	4.9	4.7	4.3
$V(2.5)$	3.9	**4.7**	4.0	4.9	5.0	> 5	4.8
$V(7.5)$	3.7	3.7	$>$ **5**	3.7	> 5	> 5	> 5
$V(8)$	3.2	3.2	**4.8**	3.2	4.7	4.8	4.8
$V(8.5)$	2.8	2.8	**4.1**	2.8	4.0	4.3	4.7
$V(11.5)$	2.8	2.8	**4.1**	2.8	4.0	4.6	4.8
$V(12)$	3.2	3.2	**4.8**	3.2	4.7	> 5	4.9
$V(12.5)$	3.7	3.7	$>$ **5**	3.7	> 5	> 5	> 5
$V(17.5)$	3.9	4.6	4.0	**4.7**	5.0	> 5	4.8
$V(18)$	3.1	4.0	3.2	**4.9**	4.9	4.5	4.3
$V(18.5)$	2.5	3.3	2.6	**4.1**	4.0	4.2	4.0
System	1.7	2.3	1.9	2.3	3.6	3.6	3.7

- **Tipper truck**: Commonly used on construction sites for transporting materials like sand or gravel. Various configurations exist, differing in axle count and spacing. For the current case study, the preferred truck has two tandems, each with axle loads of 1.3×300 kN $\approx$ 40 t. Tandem spacing is 1.3 m, with 6 m between tandems.
- **Semi-low trailer**: To reduce axle loads, a semi-low trailer can also be used. The ideal set-up has eight axles, each loaded at 0.8×300 kN $\approx$ 25 t, spaced at 1.5 m.

The reliability evaluation results for both configurations are summarised in Table 9. The tipper truck performs better in verifying shear capacity due to higher reliability values, while the trailer is more effective for bending. However, the semi-low trailer's gross weight exceeds 206 t, which is significantly higher than the tipper truck's 167 t. Such high weights could lead to overloading of the supports, which may be just as critical as the deck slab.

6. Discussion

The load levels in Section 5 were selected to ensure that each cross-section meets individual reliability requirements. As a result, the overall system reliability is lower than normative targets, since all elements contribute to failure in the assumed series model. This procedure aligns with current practice, where components are verified individually. However, system-level reliability assessment can provide more accurate results, particularly within risk-based approaches addressing safety, economy, and the environment [3]. This study assumes fully correlated model uncertainties across components, appropriate for elements within the same structure. For applications involving multiple structures, this assumption requires re-evaluation, although the methodology remains unchanged [37]. Likewise, direct application of the derived

transfer factors and target loads to other contexts requires caution. In particular, the influence of dynamic effects is case-specific and should always be appropriately quantified [81,82].

It should be noted that, if components are too dissimilar, there comes a point where no magnitude of the transfer factor can demonstrate sufficient reliability. In such cases, their variability is too substantial and appropriate subpopulations should be identified to reduce epistemic uncertainty. At the other end of the spectrum are very similar components. For example, if cross-sections in concrete members are closely spaced, correlations should be introduced using random field descriptions (Section 2.4). However, the resistances and load effects will be very highly correlated. Then, the extra cross-sections will not necessarily lead to additional accuracy, as the random dimension does not increase significantly—only resulting in extra computational effort [56,83]. Practical factors can also guide subpopulation identification; for instance, concrete cast in summer may differ markedly concrete cast in winter in case of long construction periods.

The analysis of the natural variability of shear resistance (Section 2.3) and comparisons with values in the literature reveal that within-structure variability is generally lower than population-level values. This distinction is addressed via a hierarchical Bayesian model, which may also be appropriate for other parameters, such as permanent loads. In this context, the use of equicorrelation matrices is physically unsatisfactory, and hierarchical modelling offers a more consistent representation of the true spatial variability.

Due to their general applicability, Bayesian procedures such as adopted in this work, can accommodate a wide range of information sources and observations. By linking the structural response observed during testing to prior laboratory data, continuous reliability updates can justify substantial reductions in the required test load [84]. Additionally, Bayesian probability theory offers the opportunity to consider load testing itself within a value-of-information framework. It enables informed decision-making under considerable uncertainties regarding testing risks and associated costs [85,86]. While not the primary objective of the present study, the procedures developed herein aid risk evaluation prior to and during proof load testing.

7. Conclusions

The current study proposes a probabilistic framework to quantify the reliability of bridges through proof load testing, explicitly accounting for spatial correlation and system behaviour. Recognising that existing infrastructure is typically assessed under significant uncertainty, a Bayesian procedure is suggested in which low-informative priors represent basic structural information. Specific attention is paid to the natural variability of shear resistance, as it is relevant to many concrete bridges and viaducts. The proposed reliability updating method also incorporates the survival of past traffic loads to account for the in-service proven strength of existing bridges. The methodology is applied to two case studies, offering guidance on proof load testing where spatial similarities significantly influence the assessment.

For simply-supported slab bridges with a varying number of spans, the increase in target loads required to compensate for untested components can be systematically represented through transfer factors. The magnitude of these factors depends on the failure mode, number of components, and sample size. Shear-critical components display higher factors compared to bending due to larger inherent variability. Despite methodological differences, the calculated factors are comparable to those in the German guideline for proof load testing [30]. For a continuous slab bridge, separately testing each critical section is considered as a basic strategy to verify its reliability. An optimised testing scheme, involving fewer locations and moderately increased load levels, achieves the same reliability with lower test effort. Testing strategies using moving vehicles were also investigated, leading to the specification of tipper truck and semi-low trailer configurations suitable for proof load testing.

The results demonstrate how advanced probabilistic analysis can support the development of optimal proof load testing strategies, where the number of test positions and target loads are balanced. The proposed methodology is flexible, allowing for the integration of bridge-specific

information to enable tailored structural assessment or the derivation of tabulated factors for common, repetitive structures.

CRediT authorship contribution statement

R. de Vries: Writing – original draft, Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Visualisation. **E. O. L. Lantsoght:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualisation. **R. D. J. M. Steenbergen:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualisation. **M. A. N. Hendriks:** Writing – review & editing, Supervision, Funding acquisition, Conceptualisation. **S. A. A. M. Fennis:** Writing – review & editing, Funding acquisition, Resources, Conceptualisation.

Declaration of competing interest

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

Acknowledgements

The authors wish to express their gratitude and sincere appreciation to the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat) for financing the research work. In addition, the fruitful discussions with Marius Naaktgeboren of Rijkswaterstaat, Ton Vrouwenvelder of TNO and Yuguang Yang from Delft University of Technology have been of great help. Special thanks are also due to ACI Committee 445-0D and the DAfStb for providing access to their comprehensive shear database.

Data availability

Data will be made available on request.

References

[1] M. H. Faber. Reliability based assessment of existing structures. *Progress in Structural Engineering and Materials*, 2(2):247–253, 2000. https://doi.org/10.1002/1528-2716(200004/06)2:2<247::AID-PSE31>3.0.CO;2-H

[2] M. Holický. New European document on assessment of existing structures and building stock. *IOP Conference Series: Earth and Environmental Science*, 290(1):012133, 2019. https://doi.org/10.1088/1755-1315/290/1/012133

[3] D. Diamantidis, P. Tanner, M. Holický, H. O. Madsen, and M. Sykora. On reliability assessment of existing structures. *Structural Safety*, 113:102452, 2025. https://doi.org/10.1016/j.strusafe.2024.102452

[4] M. G. Stewart, D. V. Rosowsky, and D. V. Val. Reliability-based bridge assessment using risk-ranking decision analysis. *Structural Safety*, 23(4):397–405, 2001. https://doi.org/10.1016/S0167-4730(02)00010-3

[5] R. D. J. M. Steenbergen and A. C. W. M. Vrouwenvelder. Safety philosophy for existing structures and partial factors for traffic loads on bridges. *Heron*, 55(2):123–139, 2010.

[6] M. Sýkora, D. Diamantidis, M. Holický, and K. Jung. Target reliability for existing structures considering economic and societal aspects. *Structure and Infrastructure Engineering*, 13(1):181–194, 2017. https://doi.org/10.1080/15732479.2016.1198394

[7] B. Bhattacharya, D. Li, M. Chajes, and J. Hastings. Reliability-based load and resistance factor rating using in-service data. *Journal of Bridge Engineering*, 10(5):530–543, 2005. https://doi.org/10.1061/(ASCE)1084-0702(2005)10:5(530)

[8] S. Alampalli, D. M. Frangopol, J. Grimson, D. Kosnik, M. Halling, E. O. L. Lantsoght, and Y. E. Zhou. Primer on bridge load testing. Transportation Research Circular E-C257, Transportation Research Board, Washington, D.C., USA, 2019.

[9] M. H. Faber, Val D. V., and M. G. Stewart. Proof load testing for bridge assessment and upgrading. *Engineering Structures*, 22:1677–1689, 2000. https://doi.org/10.1016/S0141-0296(99)00111-X

[10] E. O. L. Lantsoght, editor. *Load testing of bridges: Current practice and diagnostic load testing.* CRC Press, 2019. https://doi.org/10.1201/9780429265426.

[11] E. O. L. Lantsoght, editor. *Load testing of bridges: Proof load testing and the future of load testing.* CRC Press, 2019. https://doi.org/10.1201/9780429265969.

[12] J. R. Benjamin and C. A. Cornell. *Probability, statistics, and decision for civil engineers.* Dover, Mineola, USA, 2014. ISBN 978-0486780726.

[13] D. Straub and I. Papaioannou. Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, 141(3), 2015. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000839

[14] N. de Koker, C. B. Viljoen, R. Lenner, and S. W. Jacobsz. Updating structural reliability efficiently using load measurement. *Structural Safety*, 84:101939, 2020. https://doi.org/10.1016/j.strusafe.2020.101939

[15] M. Kapoor. Optimal structural health information approaches for the efficient classification and management of structural systems. PhD thesis, Department of Civil Engineering, Technical University of Denmark (DTU), 2021.

[16] O. Ditlevsen. Narrow reliability bounds for structural systems. *Journal of Structural Mechanics*, 7 (4):453–472, 1979. https://doi.org/10.1080/03601217908905329

[17] S. Celati, A. Natali, W. Salvatore, I. Björnsson, and S. Thöns. Spatial and time-dependent reliability analysis for post-tensioned concrete decks subjected to multiple failure modes. *Structural Safety*, 117:102634, 2025. https://doi.org/10.1016/j.strusafe.2025.102634

[18] R. de Vries, E. O. L. Lantsoght, R. D. J. M. Steenbergen, and S. A. A. M. Fennis. Reliability assessment of existing reinforced concrete bridges and viaducts through proof load testing. *Proceedings of the 11th IABMAS conference, Barcelona, Spain*, pages 467–475, 2022. https://doi.org/10.1201/9781003322641-54

[19] M. Grigoriu and W. B. Hall. Probabilistic models for proof load testing. *Journal of Structural Engineering*, 110(2):260–274, 1984. https://doi.org/10.1061/(ASCE)0733-9445(1984)110:2(260)

[20] T. S. Lin and A. S. Nowak. Proof loading and structural reliability. *Reliability Engineering*, 8:85–100, 1984. https://doi.org/10.1016/0143-8174(84)90057-X

[21] R. Rackwitz and K. Schrupp. Quality control, proof testing and structural reliability. *Structural Safety*, 2:239–244, 1985. https://doi.org/10.1016/0167-4730(85)90030-X

[22] N. Wang, B. R. Ellingwood, and A. H. Zureick. Bridge rating using system reliability assessment. II: Improvements to bridge rating practices. *Journal of Bridge Engineering*, 16(6):863–871, 2011. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000171

[23] R. de Vries, R. D. J. M. Steenbergen, and J. Maljaars. Annual reliability requirements for bridges and viaducts. *Heron*, 68(2), 2023.

[24] G. Spaethe. Die Beeinflussung der Sicherheit eines Tragwerks durch Probebelastung (The influence of proof load testing on the structural reliability). *Bauingenieur*, 69:459–468, 1994.

[25] R. de Vries, E. O. L. Lantsoght, R. D. J. M. Steenbergen, and S. A. A. M. Fennis. Time-dependent reliability assessment of existing concrete bridges with varying knowledge levels by proof load testing. *Structure and Infrastructure Engineering: IABMAS 2022 Special Issue*, 20(7-8):1053–1067, 2023. https://doi.org/10.1080/15732479.2023.2280712

[26] O. Abril-Pla, V. Andreani, C. Carroll, L. Dong, C. J. Fonnesbeck, M. Kochurov, R. Kumar, J. Lao, C. C. Luhmann, O. A. Martin, M. Osthege, R. Vieira, T. Wiecki, and R. Zinkov. PyMC: a modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*, 9: e1516, 2023. https://doi.org/10.7717/peerj-cs.1516

[27] S. L. Lauritzen, A. P. Dawid, B. N. Larsen, and H.-G. Leimer. Independence properties of directed Markov fields. *Networks*, 20(5):491–505, 1990. https://doi.org/10.1002/net.3230200503

[28] Y. E. Zhou and M. R. Guzda. Bridge load rating through proof load testing for shear at dapped ends of prestressed concrete girders. *Frontiers in Built Environment*, 6:117, 2020. https://doi.org/10.3389/fbuil.2020.00117

[29] CEN. Eurocode 0: Basis of structural design. Standard EN 1990+A1+A1/C2:2019, European Committee for Standardization, Brussels, Belgium, 2019.

[30] DAfStb. DAfStb-Richtlinie: Belastungsversuche an Betonbauwerken (DAfStb guideline: Load tests on concrete structures). Deutscher Ausschuss für Stahlbeton, Berlin, Germany, 2020.

[31] S. Marx, J. Grünberg, and G. Schacht. Methods to evaluate test results based on small sample sizes. *Civil Engineering Design*, 1(2):74–84, 2019. https://doi.org/10.1002/cend.201900012

[32] ACI. Strength evaluation of existing concrete buildings. Number 437R-19, American Concrete Institute, Farmington Hills, USA, 2019.

[33] P.-S. Laplace. *Essai philosophique sur les probabilités (A philosophical essay on probabilities)*. Courcier, Paris, France, 1814.

[34] A. Der Kiureghian. *Structural and system reliability*. Cambridge University Press, 2022. https://doi.org/10.1017/9781108991889.

[35] JCSS. Probabilistic model code. Joint Committee on Structural Safety, 2015. https://www.jcss-lc.org/jcss-probabilistic-model-code/

[36] K. Nishijima and M. H. Faber. Bayesian approach to proof loading of quasi-identical multi-components structural systems. *Civil Engineering and Environmental Systems*, 24(2):111–121, 2007. https://doi.org/10.1080/10286600601159172

[37] R. de Vries, R. D. J. M. Steenbergen, and A. C. W. M. Vrouwenvelder. Bayesian structural reliability updating using a population track record. *Reliability Engineering & System Safety*, 225:110644, 2025. https://doi.org/10.1016/j.ress.2024.110644

[38] E. C. Bentz and S. Buckley. Repeating a classic set of experiments on size effect in shear of members without stirrups. *ACI Structural Journal*, 102(6):832–838, 2005. https://doi.org/10.14359/14791

[39] K. Bernander. An investigation of the shear strength of concrete beams without stirrups or diagonal bars. *Proceedings of the RILEM Symposium, Stockholm, Sweden*, 1:211–214, 1957.

[40] P. S. Chana. Some aspects of modelling the behaviour of reinforced concrete under shear loading. Technical report 543, Cement and Concrete Association, Wexham Springs, UK, 1981.

[41] M. P. Collins and D. Kuchma. How safe are our large, lightly reinforced concrete beams, slabs and footings? *ACI Structural Journal*, 96(4):482–490, 1999. https://doi.org/10.14359/684

[42] M. Hallgren. Shear tests on reinforced high and normal strength concrete beams without stirrups. KTH Royal Institute of Technology, Stockholm, Sweden, 1994.

[43] F. Leonhardt and R. Walther. Schubversuche an einfeldrigen Stahlbetonbalken mit und ohne Schubbewehrung (Shear tests on simply supported reinforced concrete beams with and without shear reinforcement). Heft 151, DAfStb, Berlin, Germany, 1962.

[44] A. S. Lubell, E. G. Sherwood, E. C. Bentz, and M. P. Collins. Safe shear design of large wide beams. *Concrete International*, 26(1):66–78, 2004.

[45] P. E. Regan. Shear in reinforced concrete - An experimental study. CIRIA-Report, Imperial College of Science and Technology, Department of Civil Engineering, London, UK, 1971.

[46] R. Sarkhosh, J. Walraven, and J. Den Uijl. Shear-critical reinforced concrete beams under sustained loading – Part I: Experiments. *Heron*, 60(3):181–206, 2015.

[47] E. G. Sherwood. One-way shear behaviour of large, lightly-reinforced concrete beams and slabs. PhD thesis, University of Toronto, Department of Civil Engineering, Toronto, Canada, 2008.

[48] S. H. Ahmad, Y. Xie, and T. Yu. Effectiveness of shear reinforcement for normal and high-strength concrete beams. *Structural Engineering Review*, 8(1):65–78, 1996.

[49] C. J. Bernhardt and C. C. Fynboe. High strength concrete beams. *Nordic Concrete Research*, 5: 19–26, 1986.

[50] P. Kong and B. V. Rangan. Reinforced high strength concrete (HSC) beams in shear. *Australian Civil Engineering Transactions*, 39(1):43–50, 1997.

[51] W. J. Krefeld and C. W. Thurston. Studies of the shear and diagonal tension strength of simply supported reinforced concrete beams. *Journal of the American Concrete Institute*, 63(4), 1966. https://doi.org/10.14359/7633

[52] K. Özden. An experimental investigation on the shear strength of reinforced concrete beams: Tests performed at structural research laboratory, technical university of Denmark. Technical University of Istanbul, Istanbul, Turkey, 1967.

[53] K.-H. Reineck, E. C. Bentz, B. Fitik, D. A. Kuchma, and O. Bayrak. ACI-DAfStb database of shear tests on slender reinforced concrete beams without stirrups. *ACI Structural Journal*, 110(5):867–876, 2013. https://doi.org/10.14359/51685839

[54] K.-H. Reineck, E. C. Bentz, B. Fitik, D. A. Kuchma, and O. Bayrak. ACI-DAfStb databases for shear tests on slender reinforced concrete beams with stirrups. *ACI Structural Journal*, 111(5):1147–1156, 2014. https://doi.org/10.14359/51686819

[55] ACI-DAfStb. "Shear databases" for structural concrete members. https://dafstb.de/aci-dafstb.html, 2025. Accessed 12 March 2025.

[56] E. H. Vanmarcke. *Random fields: Analysis and synthesis*. MIT Press, Cambridge, USA, 1983. ISBN 9780262720458.

[57] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic processes*. McGraw-Hill, New York, USA, 4th edition, 2002. ISBN 9780071226615.

[58] P. Guttorp and T. Gneiting. Studies in the history of probability and statistics xlix: On the Matérn correlation family. *Biometrika*, 93(4):989–995, 2006. https://doi.org/10.1093/biomet/93.4.989

[59] S. Geyer, I. Papaioannou, and D. Straub. Spatial modeling of concrete strength based on data. *Structural Safety*, 103:102345, 2023. https://doi.org/10.1016/j.strusafe.2023.102345

[60] W. Botte, E. Vereecken, and R. Caspeele. Random field modelling of spatial variability in concrete – a review. *Structure and Infrastructure Engineering*, 21:1047–1060, 2023. https://doi.org/10.1080/15732479.2023.2248102

[61] Partial factor methods for existing concrete structures. Bulletin 80, International Federation for Structural Concrete (*fib*), Lausanne, Switzerland, 2016. https://doi.org/10.35789/fib.BULL.0080

[62] O. Ditlevsen and A. Vrouwenvelder. "Objective" low informative priors for Bayesian inference from totally censored Gaussian data. *Structural Safety*, 16(3):175–188, 1994. https://doi.org/10.1016/0167-4730(94)00020-Q

[63] R. Lu, Y. Luo, and J. P. Conte. Reliability evaluation of reinforced concrete beams. *Structural Safety*, 14(4):277–298, 1994. https://doi.org/10.1016/0167-4730(94)90016-7

[64] G. M. Hornberger and R. C. Spear. Eutrophication in peel inlet–I. The problem-defining behavior and a mathematical model for the phosphorus scenario. *Water Research*, 14(1):29–42, 1980. https://doi.org/10.1016/0043-1354(80)90039-1

[65] D. W. Dilks, R. P. Canale, and P. G. Meier. Development of Bayesian Monte Carlo techniques for water quality model uncertainty. *Ecological Modelling*, 62(1):149–162, 1992. https://doi.org/10.1016/0304-3800(92)90087-U

[66] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1091, 1953.

[67] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. https://doi.org/10.1093/biomet/57.1.97

[68] S. S. Qian, C. A. Stow, and M. E. Borsuk. On Monte Carlo methods for Bayesian inference. *Ecological Modelling*, 159:269–277, 2003.

[69] B. Stroustrup. *The C++ programming language*. Pearson Education, 4th edition, 2013. ISBN 978-0-321-56384-2.

[70] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman & Hall / CRC Press, New York, USA, 3rd edition, 2013. https://doi.org/10.1201/b16018.

[71] R. E. Melchers and A. T. Beck. *Structural reliability analysis and prediction*. John Wiley & Sons, Chichester, UK, 3rd edition, 2018. ISBN 978-1-119-26399-9.

[72] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, USA, 4th edition, 2013. ISBN 978-1-4214-0794-4.

[73] FHWA. Weigh-in-motion pocket guide – Part 1: WIM technology, data acquisition, and procurement guide. FHWA-PL-18-015, Federal Highway Administration (FHWA), Washington D.C., USA, 2018.

[74] M. Sibuya. Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics*, 11(2): 195–210, 1959.

[75] B. T. Russell and W. K. Huang. Modeling short-ranged dependence in block extrema with application to polar temperature data. *Environmetrics*, 32:e2661, 2021. https://doi.org/10.1002/env.2661

[76] E. Lantsoght, C. Veen, A. Boer, and D. Hordijk. Proof load testing of reinforced concrete slab bridges in the Netherlands. *Structural Concrete*, 18(4):597–606, 2017. https://doi.org/10.1002/suco.201600171

[77] C. O. Christensen, F. Zhang, G. I. Zarate Garnica, E. O. L. Lantsoght, P. Goltermann, and J. W. Smith. Identification of stop criteria for large-scale laboratory slab tests using digital image correlation and acoustic emission. *Infrastructures*, 7(3), 2022. https://doi.org/10.3390/infrastructures7030036

[78] R. de Vries, E. O. L. Lantsoght, R. D. J. M. Steenbergen, and M. Naaktgeboren. Proof load testing method by the American association of state highway and transportation officials and suggestions for improvement. *Transportation Research Record*, 2677(11):245–257, 2023. https://doi.org/10.1177/03611981231165026

[79] *fib. Model code for concrete structures 2010*. Ernst & Sohn, Berlin, Germany, 2013. ISBN 978-3-433-03061-5.

[80] X. Zheng, T.-H. Yi, D.-H. Yang, and H.-N. Li. Multisection optimization-based target proof load determination method for bridge load testing. *Journal of Bridge Engineering*, 28(6):04023025, 2023. https://doi.org/10.1061/JBENF2.BEENG-6073

[81] A. González, A. Žnidarič, J. R. Casas, and E. J. O'Brien. Recommendations on dynamic amplification allowance in assessment of bridges. Technical report, ARCHES Deliverable D10, FEHRL, European Commission, Brussels, Belgium, 2009. http://hdl.handle.net/10197/6238

[82] L. Deng, Y. Yu, Q. Zou, and C. S. Cai. State-of-the-art review of dynamic impact factors of highway bridges. *Journal of Bridge Engineering*, 20(5):04014080, 2015. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000672

[83] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: A spectral approach*. Springer, New York, USA, 1991. https://doi.org/10.1007/978-1-4612-3094-6.

[84] R. de Vries, E. O. L. Lantsoght, R. D. J. M. Steenbergen, M. A. N. Hendriks, and M. Naaktgeboren. Structural reliability updating on the basis of proof load testing and monitoring data. *Engineering Structures*, 330:119863, 2025. https://doi.org/10.1016/j.engstruct.2025.119863

[85] J. W. Schmidt, S. Thöns, M. Kapoor, C. O. Christensen, S. Engelund, and J. D. Sörensen. Challenges related to probabilistic decision analysis for bridge testing and reclassification. *Frontiers in Built Environment*, 6, 2020. https://doi.org/10.3389/fbuil.2020.00014

[86] W.-H. Zhang, D.-G. Lu, J. Qin, S. Thöns, and M. H. Faber. Value of information analysis in civil and infrastructure engineering: A review. *Journal of Infrastructure Preservation and Resilience*, 2, 2021. https://doi.org/10.1186/s43065-021-00027-0

This dissertation investigates how proof load testing can be used to assess the reliability of existing structures. A flexible Bayesian methodology is proposed and applied to reinforced concrete road bridges and viaducts. Key contributions include:

- A demonstration of how proof load testing affects annual reliability, showing reductions during testing, but substantial gains after surviving target loads.
- An in-depth look at the conservative lower-bound estimation of structural reliability, revealing its assumptions and limitations.
- Combining in-situ monitoring and laboratory data for Bayesian updating during testing, enabling substantial reductions in required test loads.
- Hierarchical Bayesian modelling addressing spatial correlation and system reliability, enabling optimal testing strategies with a low number of tests, and the configuration of load testing vehicles.

This research positions proof load testing at the core of a Bayesian reliability-updating methodology, thereby providing a uniquely accurate procedure for assessing existing infrastructure.

Electronic copy available at: **https://repository.tudelft.nl**
Underlying data and code at: **https://data.4tu.nl**