Acta Polytechnica CTU Proceedings **36**:109–118, 2022
https://doi.org/10.14311/APP.2022.36.0109

Published by the Czech Technical University in Prague

ON THE TREATMENT OF MEASUREMENT UNCERTAINTY IN STOCHASTIC MODELING OF BASIC VARIABLES

Stefan Küttenbaum[a,*], Stefan Maack[a], Alexander Taffe[b], Thomas Braml[c]

[a] *Bundesanstalt für Materialforschung und -prüfung (BAM), Division 8.2: Non-Destructive Testing Methods for Civil Engineering, Unter den Eichen 87, 12205 Berlin, Germany*

[b] *HTW Berlin (UAS), Division Building materials sciences and diagnostics, 12459 Berlin, Germany*

[c] *Universität der Bundeswehr München, Inst. of Struct. Engineering, 85577 Neubiberg, Germany*

[*] corresponding author: stefan.kuettenbaum@bam.de

Abstract.

The acquisition and appropriate processing of relevant information about the considered system remains a major challenge in assessment of existing structures. Both the values and the validity of computed results such as failure probabilities essentially depend on the quantity and quality of the incorporated knowledge. One source of information are onsite measurements of structural or material characteristics to be modeled as basic variables in reliability assessment. The explicit use of (quantitative) measurement results in assessment requires the quantification of the quality of the measured information, i.e., the uncertainty associated with the information acquisition and processing. This uncertainty can be referred to as measurement uncertainty. Another crucial aspect is to ensure the comparability of the measurement results.This contribution attempts to outline the necessity and the advantages of measurement uncertainty calculations in modeling of measurement data-based random variables to be included in reliability assessment. It is shown, how measured data representing time-invariant characteristics, in this case non-destructively measured inner geometrical dimensions, can be transferred into measurement results that are both comparable and quality-evaluated. The calculations are based on the rules provided in the guide to the expression of uncertainty in measurement (GUM). The GUM-framework is internationally accepted in metrology and can serve as starting point for the appropriate processing of measured data to be used in assessment. In conclusion, the effects of incorporating the non-destructively measured data into reliability analysis are presented using a prestressed concrete bridge as case-study.

Keywords: Existing structures, FORM, measurement uncertainty, nondestructive testing, reliability assessment.

1. Introduction

Considering that absolute certainty is practically not attainable and further not worth striving for, since its theoretical achievement implies the consumption (or dissipation, respectively) of infinite resources [1], every decision is associated with a higher or smaller degree of uncertainty. This also includes decisions about the reliability of existing structures. Uncertainties in reliability assessment may be characterized process inherent (such as future wind and seismic loading) and arise generally from the persistently existing lack of knowledge regarding the considered system, which is described by the characteristics, exposition, behavior, condition, etc. of the structure. Accordingly, it is vitally important to find a computation model that appropriately represents the structural system and its environment, and, for this purpose, to gather and treat relevant, additional information in a suitable way. This serves to sufficiently fill the knowledge gap regarding the system of interest and to increase the level of approximation of a model used for the assessment.

In most cases many types of information are available or at least obtainable that can (and, where appropriate, should) be utilized in reliability and condition assessment of existing structures to reduce uncertainties and to identify both biases and errors in models and assumptions. These include information from design that can be extracted from reports or drawings as well as information from field experience and observed data, respectively [2]. The evaluation of the condition of individual information to be used in assessment is crucial, amongst others because the inclusion of imprecise or even incorrect (in the sense of biased) information may lead to unfavorable decisions on reliability that may have serious consequences. It should be noted that the various conceivable sources of information provide diversely structured data that may not necessarily be treated in the same way.

The significance of on-site observations has been shown, e.g., in studies on the appreciation of advanced measurement techniques in condition assessment [3] and on the use of monitoring-data in reliability anal-

ysis [4, 5]. The refinement of computation models by including additional information facilitates the targeted planning of actions such as use restrictions, maintenance, and reconstruction, as well as improved life time predictions, the optimization of resource consumption and overall more realistic assessment results that can lead to better decisions. Whereas measurements can be seen as a tool to generate knowledge, it should be kept in mind that observed data in most cases only represent a quantity of interest, e.g., a physical characteristic, with uncertainty. This uncertainty can be understood as a measure to quantify the quality of the measurement result and its accuracy, respectively, shall be expressed in a suitable, practically applicable way and has to be principally considered when using measured information in assessment, see i. a. [2, 6].

This paper attempts to outline the necessity and the advantages of measurement uncertainty calculations in measured data-based modeling of random variables to be used as basic variables in reliability assessment. Stochastic processes and random fields are delimited. The rules provided in the Guide to the Expression of Uncertainty in Measurement (GUM) [7], which are suitable for the calculation of measurement uncertainty in many cases in metrology and at the same time computationally simple, are summarized in order to shed light on an internationally accepted approach. This concept could be applied for the comparable modeling of uncertainty that is related to information acquisition and processing in measuring data-supported reliability assessment. How measurement results can then be incorporated into a time-invariant component reliability analysis and what effects this can have on reliability is demonstrated using inner geometrical dimensions, which have been measured non-destructively on a prestressed concrete bridge in northern Germany.

2. Necessity and calculation of measurement uncertainties

The set of basic variables included in a reliability analysis and their mathematical relationship form the "entire input information" to the model used for an assessment [8]. Fundamental challenges in stochastic modeling of basic variables to be used to compute small probabilities such as failure probabilities in many engineering problems include the treatment of model uncertainties, the tail-sensitivity problem, and the quantification of correlation [9]. To address, for instance, the tail-sensitivity problem, [8] and others recommend the standardization of types of distribution functions of basic variables for certain groups of structural problems. In addition, all relevant *types of uncertainty* should be covered in a stochastic model of a basic variable [8]. Frequently mentioned types are the intrinsic physical or mechanical uncertainty, the statistical uncertainty, the model uncertainty [8], and furthermore the measurement uncertainty [2], on which this paper focuses. This uncertainty describes the precision of measured information provided that systematic errors have been corrected appropriately and serves to express the measurement result in a comparable way. This is significant in reliability analysis in that calculated values need to be comparable to certain target values.

Measurement uncertainty can be defined as a parameter to quantify the dispersion of the values assigned to the quantity to be measured (the measurand) based on the incorporated information [10]. "From the metrological point of view, a measured value to which no measurement uncertainty has been assigned is useless. The calculation of measurement uncertainty serves to establish confidence in measurement, to ensure the comparability of measurement results and to express the quality, that is, trueness and precision, of the information measured about a characteristic. In the context of modeling basic variables to be used in assessment, two central requirements on stochastic models can be met by adequate measurement uncertainty considerations: verifiability and comparability. Moreover, a measurement result is required to be unambiguously expressed and transparently documented. Thus, the objectivity is assured in the sense that the calculated results as well as the models, input quantities, and assumptions underlying the measurement uncertainty considerations are deniable." [11]. "A good or rather useful measured data-based probabilistic model should cover the uncertainty associated with information acquisition and processing besides the uncertainty quantifying the inherent natural variability of the considered characteristic. The measurement uncertainty describes the limits of an interval containing the (generally unknown) true value of the measurand with a certain probability, and is epistemic, provided that an alternative exists to obtain the information (different testing methods, etc.). A stochastic model that has been created based on observations on site and that does not cover the uncertainty to be attributed to the information acquisition and processing appears to be equally useless as a measurement value to which no measurement uncertainty has been attributed to." [11].

In the following, it is shown how measurement uncertainties can be calculated according to the GUM. The explanations refer to the main document [7]. The application of Monte Carlo simulation to the propagation of distributions and the computation of multiple output quantities is treated in the supplements [12, 13]. The basic idea is to find a model of the measurement consisting of different input quantities X_i that may be either necessary to compute the measurement result or influence the outcome of the experiments in most cases unfavorably (and hence contribute to measurement uncertainty unless they are modeled deterministically). The functional relationship of these input quantities can be often expressed in form of an explicit model equation and is used to determine the output quantity

Y, that is, the quantity of interest (the measurand):

$$Y = f(X_i) \tag{1}$$

The input quantities X_i have to be identified using, e.g., the knowledge about the measuring process, and evaluated with respect to their individual relevance.

***Example**: A considerable number of planners and owners requested the nondestructive localization of the reinforcement and tendons inside of concrete components in the past. The aim was not necessarily to reconstruct missing or incomplete, but also questioned as-built drawings. An example is the localization of longitudinal tendons inside existing bridge webs to safely drill cores and subsequently pretension anchor blocks transversely against each other for external strengthening actions. Let us thus define the depth position of a (posttensioned bonded) tendon in relation to the concrete surface, which serves as the measuring area in ground penetrating radar (GPR) inspection, as the measurand . The principle of GPR is to derive the distance between an antenna and an object of interest using (a) the observed times of flight (TOF) needed for an impulse to travel the respective distance forwards and backwards and (b) the propagation velocity of the electromagnetic wave inside the specific investigated volume. The equation*

$$Y = d_{Sp} = f(X_i) = VT/2, \tag{2}$$

with T being the observed TOF and V being the velocity, can therefore serve as starting point to develop the individual model function in the case of echo arrangement (see Figure 2 in section 3) of the GPR antennas.

Subsequently, the relevant input quantities need to be quantified. Since they are considered to be random variables in most cases, quantification means finding a stochastic model appropriately representing the individual phenomena. For this purpose, the evaluation of measuring series with statistical methods, the type A evaluation, or the use of other information such as data from calibration certificates, physical reasoning, and experts' judgements, the type B evaluation, can lead to suitable distributions of the input quantities.

The type A evaluation can be applied if a measuring series consists of a sufficiently large number of identically distributed, independent values that were observed under constant conditions. The distribution parameters of an input quantity can then be estimated with statistical methods. It should be noted that the application of statistical methods may yield less reliable results in comparison to the type B evaluation if the number of observations is too small. Particularly precise knowledge about a quantity may in turn obviate the need for experiments and type A evaluation. In type A, the sample mean is often considered the best estimate

$$\hat{x} = \overline{x} - b = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - b \tag{3}$$

of a directly measurable quantity, provided that the systematic errors b have been estimated and corrected. To this best estimate, a standard uncertainty $u(\hat{x})$ has to be attributed, that can be interpreted as standard deviation of the mean. It describes, how well $\hat{x}$ estimates the expected value [7], and equals the sample standard deviation divided by the square root of observed values $\sqrt{n}$.

$$u(\hat{x}) = \frac{S}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{4}$$

It should be noted that the simple description of observations may not lead to appropriate stochastic models. On the one hand, the frequently used standard deviation σ_X describes the dispersion of observations and allows for interpretations in a way that, e.g., 68 out of 100 single values observed in the future are expected to be included in an interval $(\overline{x} - \sigma_X;\ \overline{x} + \sigma_X)$. On the other, the standard measurement uncertainty, cf. Equation **??**, quantifies the scattering behavior of the directly measurable input quantity, that is, the characteristic of interest. This measure appears suitable in measured data-based modelling of basic variables since it is usually aimed to characterize (physical) characteristics, but not to predict single future observations. With regard to the distribution type, a convenient justification for choosing a normal distribution in type A evaluation (and also for the measurand Y) can be found in the central limit theorem for many practical cases.

***Example**: Consider T in Equation 2 as a set of random variables which not only consists of the observed TOFs T_A but also of further input quantities affecting the outcome of the measurement. These quantities $\sum T_i$ among others include unknown processes within the equipment ("black box"), the limited measuring scale resolution, the spacing between transmitting antenna, receiving antenna, and concrete surface, as well as frequency- and travel path dependent changes in pulse shape, and uncertainties in signal processing.*

$$T = T_A - T_V - \sum T_i \tag{5}$$

All identified and individually relevant input quantities such as the lead time T_V are usually modelled as random variables. The black box phenomena, for example, can be comparatively easily quantified. In this specific case, 200 TOFs have been derived from echoes originating from the backwall of an isotropic-homogeneous specimen (polyamide) at the same position to ensure repeatability condition approximately. All the TOFs equal $4,617ns$. Unknown processes possibly occurring in the equipment therefore do not lead to significant random errors. The related input quantity can thus be individually neglected.

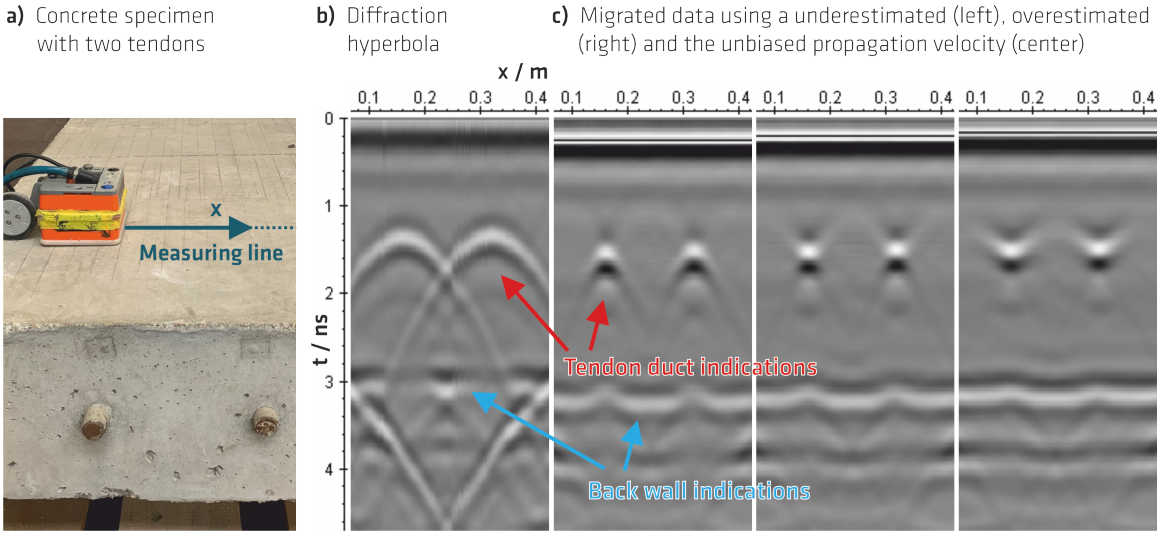

Figure 1. a) 2 GHz GPR antenna, specimen, and measuring line; b) Radargram with both tendon indications (diffraction hyperbola); c) Radargrams of Kirchhoff migrated data gathered on the same line using different propagation velocities. An overestimation leads to V-shaped amplitudes in upper direction (on the right). If underestimated, the diffraction hyperbola remains visible (third radargram from the right).

*Considering that metal layers with thicknesses less than $\lambda/20$ can be usually detected in GPR testing, that wavelengths $\lambda = 5 \ldots 12\ cm$ are common in concrete, and that the electromagnetic pulses are totally reflected at metallic objects, the measuring series $t_A = (t_{A,1}, \ldots, t_{A,n})^T$ contains $i = 1, \ldots, n$time stamps t_{Ai} describing the time it takes the pulse to travel through the concrete to the tendon duct and back to the antenna. Due to the requirement of i.i.d. observations, such a measuring series is only valid to describe the mounting depth of a tendon at one position in tendon length axis (sampling point) and therewith contains the values observed in a certain small area within the measuring plane. The sampling points can then be combined to interpolate the curve of the tendon. The time stamps gathered to describe the tendons mounting depth in the cross-section to be assessed in section 3 are: $2,81ns, 2,81ns, 2,81ns, 2,79ns, 2,79ns$, $2,79\ ns$, $2,79\ ns$. Thus, Equation 4 yields the best estimate $\hat{t}_A = 2,80\ ns$, and Equation **??** the standard uncertainty $u(\hat{t}_A) = 4 \times 10^{-3}\ ns$.*

The aim of TOF measurement is to determine the time span required for a pulse to travel a certain distance within a component. Picked time stamps $t_{A,i}$ additionally contain (at least partly) the time span necessary to generate, transmit, and sample the signal. This lead time or offset T_V is a systematic error which must be estimated and corrected to define an unbiased time zero. Different approaches have been proposed for this purpose (see, e.g., [14]). One option is the lead time estimation using the intercept of a regression line, where the -values describe a successively varied distance between the antenna and a metal plate. The -values are the respective measured TOFs. A numerical discussion is delimited. However, the -intercept at $x = 0\ mm$ yields the best estimate and the variance of the intercept the (squared) standard uncertainty.

In type B evaluation, the standard measurement uncertainty is derived from a distribution function defined using non-statistical methods. The standard uncertainties to be attributed to the best estimates of the input quantities depend on the distribution families chosen.

Example: *Consider the individual propagation velocity of the electromagnetic wave inside the concrete V, which (acc. to Equation 2) needs to be known to derive mounting depths and depends primarily on the relative permittivity. Provided the personnel is appropriately qualified, it can be sufficient to estimate the value of the velocity by analysing the shape of the diffraction hyperbola or focussing level of migrated indications of bar-shaped reflectors with round cross sections. Experts judged that a deviation from the physically reasonable shape of the respective indications could be identified in this individual case when the velocity is over- or underestimated by about $\pm 0,75\ cm/ns$. The effects are shown qualitatively in Figure 1. The subjectively highest focusing level of the tendon indications in the specific case outlined in sect. 3 could be achieved with a set velocity of $12\ cm/ns$.. If solely the limit values of an interval are known, in which the random variable realizes arbitrarily, then it can be derived from the principle of maximum entropy that the quantity is uniformly distributed. From this it follows, that $V \approx U$ with a best estimate $\hat{v} = 12cm/ns$ and a standard uncertainty $u(\hat{v}) = (b-a)/(2\sqrt{3}) = (12,75 - 11,25)/(2\sqrt{3}) = 0,433\ cm/ns$ with a, b being the maximum and minimum values. If it turns out that such "rough" modelling is not sufficient to achieve the desired measurement precision, models with large uncertainty contributions can be refined.*

Another input quantity T_Z that arises frequently in measurements describes the limited resolution of the measuring scale, in this case time axis, and can be

Abbr.	Description	Distr. Type	Best estimate $\hat{x}_i$	Standard uncertainty $u(\hat{x}_i)$
V	Propagation velocity	Uniform	12 cm/ns	0,433 ns
T_A	"Displayed" TOFs	Normal	2,8 ns	4×10^{-3} ns
T_V	Offset / delay	Normal	(0,9 ns)	$8,6 \times 10^{-2}$ ns
T_M	"Black box" NDT-system	const.	0 ns	0 ns
T_Z	Time axis resolution	Uniform	0 ns	7×10^{-3} ns
T_{AB}	Spacing antenna / surface	Uniform	(0,035 ns)	0,02 ns

TABLE 1. Individual models of the input quantities used to compute the measurement result. The biases regarding offset $\hat{t}_V$, antenna-surface-spacing $\hat{t}_{AB}$, and transmitter-receiver-spacing were corrected during data migration.

evaluated precisely via type B. The spacing Δt between two sample values equals the inverse of the sampling rate f_s. The quantity contributes to uncertainty in that an observed amplitude falls arbitrarily into an interval spanned symmetrically around a sample value. Thus, again, the limit values of the random variable can be specified, whereas no information about a weighted distribution of the probabilities within the interval is available. From this, $f_s = 42,7\,GHz$, and thus $\Delta t = 0,023\,ns$ it follows a uniform distribution with $b = \pm\Delta t/2$ and $u(\hat{v}_Z) \approx 7 \times 10^{-3}\,ns$.

Modeling should not rely too much on the consideration of perceived or physically reasoned correlations since it is the statistical relationship between (two) random variables and not the dependencies between the associated physical quantities that need to be estimated. It may therefore be sufficient in practice, to solely appreciate the pairwise correlations between type A evaluated input quantities, since empirical covariances could then be easily computed (among other parameters). Nevertheless, correlations with at least on type B evaluated input quantity would be disregarded, which is only permissible if the considered input quantities are not significantly correlated [7] or respective information is neither available nor appropriately obtainable.

If the input quantities have been identified individually, if suitable models have been defined at least for the relevant quantities, and if the input quantities X_i have been brought into a functional relationship, the main issue in measurement uncertainty calculation has been solved. Based on this model of the measurement, the calculation formulae provided in the GUM can be straightforwardly applied. Inserting the estimated values $\hat{x}_i$ of the input quantities (in case of type A evaluation acc. to Equations 4) into the model equation, cf. Equations 1, 2 and 5, yields the estimated value of the output quantity $\hat{y}$, which needs to be corrected for systematic errors that have not yet been taken into account:

$$\hat{y} = f(\hat{x_1}, \ldots, \hat{x}_n) \tag{6}$$

The uncertainty to be covered by the output quantity Y is referred to as combined standard measurement uncertainty $u(\hat{v})$, which expresses the measurement uncertainty as an estimated standard deviation of the measured quantity value $\hat{y}$ and is calculated by applying the error propagation law to the model equation:

$$u(\hat{y}) = \sqrt{\sum_{i=1}^{n} c_i^2\, u^2(\hat{x}_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} c_i\, c_j\, u(\hat{x}_i,\, \hat{x}_j)} \tag{7}$$

The empirical covariance is denoted by $u(\hat{x}_i,\, \hat{x}_j)$, and the sensitivity coefficient of the input quantity X_i by c_i. It should be mentioned that, in contrast to $FORM$, the model equation is partially derived with respect to the input quantities at the coordinates of the best estimates $\hat{x}_i$ in order to compute the sensitivity coefficients.

***Example**: The model equation used to calculate the measurement result is given in Equations 2 and 5. The stochastic models of the input quantities can be found in Table 1. A measured value of the mounting depth $\hat{y} = \hat{d}_{sp}$is calculated in each case for one tendon at one specific point in the longitudinal tendon axis by inserting the best estimates $\hat{x}_i$ into the model equation. From this, it follows that: $\hat{d}_{sp} = 16.8\,cm$.*

The combined standard measurement uncertainty $u(\hat{y}) = u(\hat{d}_{sp})$ is computed with Equations 7. First, the partial derivatives of the model equation with respect to the single input quantities at the best estimates yield the sensitivity coefficients . Second, the uncertainty contributions $c\,u(\hat{x}_i)$ can be easily computed. The squared contributions divided each by the sum of squares indicates the percentage distribution of the combined measurement uncertainty among the input quantities. The sum of $[c\,u(\hat{x}_i)]^2$, in turn, yields the squared combined standard measurement uncertainty in the case of uncorrelated input quantities. With $c_v = \partial d/\partial v \approx 1,4$, the uncertainty contribution of the propagation velocity, e.g., equals 0,61 cm. The influence of the velocity is thus $[c_v\,u(\hat{v})]^2 \,/\, \sum [c_i\,u(\hat{x}_i)]^2 = 0,61^2/0,66 \approx 56\,\%$, which appears consistent considering the comparatively "rough" modeling process of V. Third, the combined standard measurement uncertainty is the square root of $\sum [c\,u(\hat{x}_i)]^2$ since correlations do not affect the outcome in this particular case. The measurement result can now be expressed as follows: $d_{Sp} \sim N;$

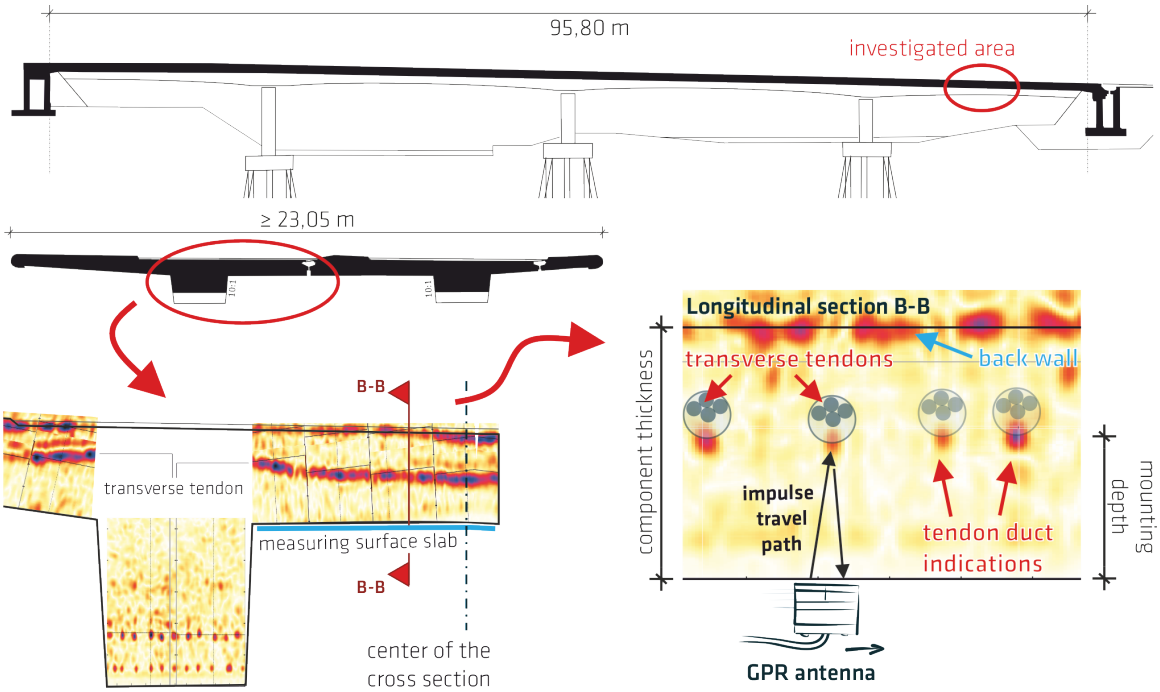

Figure 2. Longitudinal view and cross-section of the investigated bridge (top); excerpt of the decisive cross-section with imaged migrated data (bottom left) and perpendicular scan with sketched GPR echo arrangement (bottom right).

$d_{Sp} = 16,8\,cm$; $u(d_{Sp}) = 0,8\,cm$. *Comparative simulations result in slight differences of* $\Delta < \pm 0,01\,cm$ *each. The choice of the normal distribution is thus individually suitable, although the propagation velocity, which has been modelled uniformly distributed, can be considered as a dominant uncertainty component - as shown above. The interval* $16,8\,cm \pm 0,8\,cm$ *covers the set of true measured values with a probability of approx. 68 %. If this coverage probability is considered individually too low, the interval can be easily extended.*

A suitable way to unambiguously express a measurement result acc. to the GUM is the specification of the measured quantity value $\hat{y}$, the combined standard measurement uncertainty $u(\hat{y})$ and the distribution type of the measurand Y. It may be helpful to provide information on the considered correlations. Another option is to compute the expanded measurement uncertainty in order to derive coverage intervals. Further details can be found within the GUM framework and schematically, e.g., in [15].

When developing the GUM, it was aimed to provide a general method and to enable the further use of the results [7]. The comparability of the results allows for successive calculations of measurement uncertainties in relation to usual boundary conditions, which may be used as orientation in the future. Compared to the traditional approaches, the application of the GUM yields rather realistic than disproportionately large values for the measurement uncertainty [7]. One point of criticism is that the type A input quantities imply the frequentist interpretation of probability and those evaluated acc. to type B the subjective interpretation [17]. Nevertheless, the classical estimators could be interpreted as an approximation of the estimators according to type B evaluation, so that the equal treatment of all input quantities is also justified in terms of probability theory [17]. Overall, the GUM rules yield exact results for linear model equations and normally distributed input quantities. However, they are in most cases sufficiently accurate for practical applications. Simulation techniques can be additionally used to verify the choice of a distribution family for the measured quantity.

3. Demonstration - Measurement uncertainty in structural reliability assessment

The investigated structure is a four-span longitudinally and transversely prestressed concrete bridge with a total length of approx. 96 m and a slab-and-beam cross section with a width of more than 23 m (Figure 2). The bridge has been assessed in serviceability limit state (SLS) decompression in transverse direction. Prior finite element analyses revealed the center of a cross-section within the right span highlighted at the top of Figure 2 to be decisive. The investigations thus refer to this specific component. The stress analysis was performed using a representative one-meter strip in the longitudinal bridge direction ($b = 1$ m) at the upper extreme fiber of the cross section. The limit state function contains the normal forces N and bending moments M as well as the dimensions A, h, b of the cross section and an inner lever arm z_p describing the eccentricity of the transverse strands

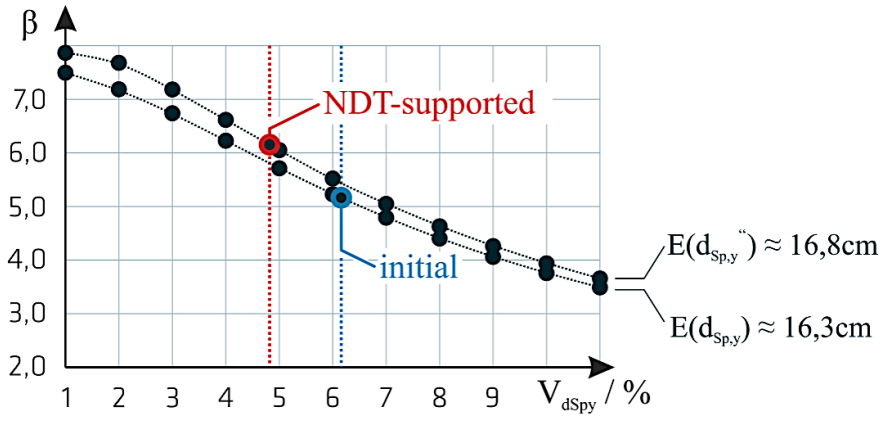

FIGURE 3. Individual effects of including the non-destructively measured vertical position of the transverse tendons on reliability using FORM in SLS decompression; $\beta_{HL,init} = 5,18$, $\beta_{HL,NDT} = 6,18$; $d_{Sp,init} \sim N(16,3\,cm,\ 1,0\,cm)$, cf. [8]; $d_{Sp,NTD} \sim N(16,8\,cm,\ 0,8\,cm)$; extracted from [16].

related to the vertical center of the cross-section. This inner lever arm z_p can be written as a function of the spacing between the tendon ducts and the undersurface of the slab d_{Sp}. This distance d_{Sp} can be efficiently measured with GPR and has been found to be decisive ($\alpha_{r,\,d_{Sp}} = 0,74$) during pre-investigations. The initially assumed standard deviation of 1 cm (Appendix A) is based on the JCSS Probabilistic Model Code [8]. The initial mean $\mu = 16,3$ cm assumed prior to onsite testing corresponds to the available drawings. The sensitivity of the structural behavior to conceivable geometrical imperfections, cf. [8], motivated NDT on-site. The stochastic models can be found in Appendix A. The limit state function is:

$$g(\sigma_c) = 0 - \left(\frac{N}{A} + \frac{M}{W}\right) = 0 - \frac{\Theta_{E,N} \sum N_i}{h\,b} - \frac{\Theta_{E,M}\,(N_p\,z_p + \sum M_i)}{h^3\,b/12}\,\frac{h}{2}; \quad (8)$$

$$z_p = -\frac{h}{2} + d_{Sp,\,y} + \epsilon$$

Chosen measured data are illustrated in Figure 2. The measurement uncertainty calculation in section 2 yields the measurement result $d_{Sp} \sim N$ with $d_{Sp} = 16,8$ cm and $u(d_{Sp}) = 0,8$ cm. This result describes the vertical tendon position in relation to the measuring plane (the slab undersurface) in the center of the cross section, which is shown at the bottom left of Figure 2. Thus, the set of population, cf. [8], equals the specific component to be assessed in this particular reliability analysis, i.e. decompression proof. The spatial variability would have to be considered additionally, for instance, in system reliability assessments, and temporal changes (which appear rather unlikely with respect to geometrical dimensions) for time-variant quantities.

The combined measurement uncertainty $u(\hat{y})$ expresses the inherent variability of the observed quantity as well as the measurement uncertainty itself as standard deviation. The measured quantity value $\hat{y}$ can be used as expected value of the basic variable especially in the case of a justified choice of a normal distribution. Both values serve as starting point for the NDT-based modeling of the basic variable d_{Sp}. It can be necessary to cover additional uncertainties such as model and statistical uncertainties. In view of the fact that modeling in particular may significantly influence the computational results, the application of the GUM concept seems to be beneficial as the comparability of measuring data-supported modeling processes can be increased. Competing models and the basic challenges in stochastic modeling of basic variables mentioned in section 2 should also be taken into account. Models can be regarded as competing if, on the basis of the available information, it cannot be decided without arbitrariness which is individually more suitable. However, the measurement result equals the stochastic model of the NDT-supported basic variable in this particular case since the tail-sensitivity problem does not affect the update of this geometrical quantity (the initial model is also normally distributed), the uncertainty of the measurement model remains insignificant (which is not unusual in metrology) and other model uncertainties have been implicitly covered, as appropriate, by entering competing models sequentially and observing the change in reliability. The statistical uncertainty was estimated using the standard deviation of the experimental standard deviation of the mean of type A evaluated input quantities [7]. The additional appreciation of this statistical uncertainty in type A evaluated input quantities yields a difference in the combined measurement uncertainty of individually less than 0,1 mm. Finally, the incorporation of prior knowledge would have been necessary, e.g., if the measured data could not describe the characteristic of interest comprehensively. Regarding the localization of tendons, this scenario may occur, if the center of a tendon bundle is to be measured, but not *all* single tendons could be detected. Then, the

quality of the measurement result also depends on the detection frequency. Probability of Detection analyses may be used to objectively evaluate whether certain objects can be reliably detected.

The effects of incorporating the non-destructively measured inner geometrical dimension, that is, the vertical position of the transverse tendons in the cross-section center, on the Hasofer-Lind-reliability is summarized in Figure 3. It can be seen that the slight shift compared to the initially assumed position and the reduction of uncertainty in relation to the probabilistic modeling recommendations [8] yield an increase in reliability of approx. + 19 %. Although the value of the reliability index calculated based on the knowledge available prior to any testing has been found to be already comparatively high, it can be deduced that a larger deviation between the as-planned and as-built position of the tendons would affect reliability considerably stronger. In other cases where inspection results are incorporated into reliability assessment, a noticeable reduction in numerical reliability may also be observed - certainly in favor of a more realistic structural assessment and to support the engineer in making reasonable decisions. Further information about the case-study outlined above can be found in [11]. A second study dealing with ultimate limit states was published in [18, 19].

4. Discussion and conclusions

It is vitally important to be aware of the condition, i.e., the relevance, trueness, and precision, of information to be used in reliability assessment. The uncertainty associated with information acquisition and processing should generally be considered; especially if measured data are intended to be incorporated into assessment. An advantage of the GUM approach is that the results are verifiable and thus objectively deniable. The rules are internationally accepted and broadly applicable. Furthermore, inferences from the calculated distribution to realizations that have not been observed can be more likely drawn by *synthesizing* the various uncertainty components, cf. [1]. In addition, measurement uncertainty can be taken as a comparable measure expressing the capability of testing methods (or measurement procedures, respectively) and may therewith be used for comparison with the results of pre-posterior-analyses, e.g., to initiate useful inspections. The measurement uncertainty can and probably should also be included when updating information based on, e.g., long-term monitoring data via Bayes' theorem (see, for example, [20]) and in monitoring-based condition assessment particularly when recorded values are close to certain thresholds.

A practical limitation is both the relatively large effort and detailed knowledge of the measurement processes required to model a measurement appropriately. For this reason, the authors aim to develop a systematic repository of models and quantified input quantities for specific boundary conditions that are considered common in structural engineering, which can then serve as orientation for future measurement uncertainty calculations performed by the engineers working in practice. In addition, structure-specific partial safety factors are intended to be derived as a function of the type and extent of different on-site measurements in order to increase the practicability of the outlined approach.

References

[1] P. Thoft-Christensen., M. J. Baker. *Structural reliability-theory and its application*s. Berlin: Springer, 1982.

[2] D. Diamantidis. *Probabilistic assessment of existing structures: A publication of the Joint Committee on Structural Safety (JCSS)*. Cachan: RILEM Publications, 2001..

[3] K. Bergmeister. *Monitoring and safety evaluation of existing concrete structures: State-of-art report prepared by Task Group 5.1*. Lausanne: fib, 2003.

[4] D. M. Frangopol, A. Strauss, S. Kim. Bridge Reliability Assessment Based on Monitoring. *Journal of Bridge Engineering* **13**(3):258-70, 2008. https://doi.org/10.1061/(asce)1084-0702(2008)13:3(258).

[5] D. M. Frangopol, A. Strauss, S. Kim. Use of monitoring extreme data for the performance prediction of structures: General approach. *Engineering Structures* **30**(12):3644-53, 2008. https://doi.org/10.1016/j.engstruct.2008.06.010.

[6] M. H. Faber. Reliability based assessment of existing structures. *Progress in Structural Engineering and Materials* **2**(2):247-53, 2000. https://doi.org/10.1002/1528-2716(200004/06)2:2<247::Aid-pse31>3.0.Co;2-h.

[7] Joint Committee for Guides in Metrology: JCGM 100:2008. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. *JCGM*, pp. 120, 2008. https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6.

[8] JCSS. JCSS Probabilistic Model Code Part 1-3. Zurich: Joint Committee on Structural Safety, 2001-2002. https://www.jcss-lc.org/jcss-probabilistic-model-code/.

[9] A. D. Kiureghian, O. Ditlevsen. Aleatory or epistemic? Does it matter? *Structural Safety* **31**(2):105-12, 2009. https://doi.org/10.1016/j.strusafe.2008.06.020.

[10] Joint Committee for Guides in Metrology: JCGM 200:2012. International vocabulary of metrology - Basic and general concepts and associated terms (VIM): 3rd edition, pp. 91, 2012. https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf/f0e1ad45-d337-bbeb-53a6-15fe649d0ff1.

[11] S. Küttenbaum, T. Braml, A. Taffe, et al. Reliability assessment of existing structures using results of nondestructive testing. *Structural Concrete* **22**(5):2895-915, 2021. https://doi.org/10.1002/suco.202100226.

[12] Joint Committee for Guides in Metrology: JCGM 101:2008. Evaluation of measurement data - Supplement 1 to the Guide to the expression of uncertainty in measurement - Propagation of distributions using a Monte Carlo method, pp. 82, 2008. https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c.

[13] Joint Committee for Guides in Metrology: JCGM 102:2011. Evaluation of measurement data - Supplement 2 to the Guide to the expression of uncertainty in measurement - Extension to any number of output quantities, pp. 72, 2011. https://www.bipm.org/documents/20126/2071204/JCGM_102_2011_E.pdf/6a3281aa-1397-d703-d7a1-a8d58c9bf2a5.

[14] R. Yelf, D. Yelf. Where is True Time Zero? *Electromagnetic Phenomena* **18**(1)158-163, 2006.

[15] Joint Committee for Guides in Metrology: JCGM 104:2009. Evaluation of measurement data - An introduction to the Guide to the expression of uncertainty in measurement and related documents, pp. 20, 2009. https://www.bipm.org/documents/20126/41373460/JCGM_104_2009.pdf/19e0a96c-6cf3-a056-4634-4465c576e513.

[16] S. Küttenbaum, T. Braml, A. Taffe, et al. From Uncertainty in Measurement to Certainty in Bridge Reassessment. *Proceedings of the 1st Conference of the European Association on Quality Control of Bridges and Structures. Lecture Notes in Civil Engineering*, pp. 509-17, 2022. https://doi.org/10.1007/978-3-030-91877-4_58.

[17] R. Kacker, A. Jones. On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent. *Metrologia* **40**(5):235-48, 2003. https://doi.org/10.1088/0026-1394/40/5/305.

[18] S. Küttenbaum S. et al. Towards NDT-supported decisions on the reliability of existing bridges. *ICOSSAR 2021-2022,13th International Conference on Structural Safety & Reliability: Conference postponed*. Contribution accepted. (Editors: J. Li et al.), 20-24 June 2022, 2022.

[19] S. Küttenbaum, S. Maack, T. Braml, et al. Bewertung von Bestandsbauwerken mit gemessenen Daten, Teil 2. *Beton- und Stahlbetonbau* **116**(3):183-99, 2020. https://doi.org/10.1002/best.202000087.

[20] S. Thöns. Monitoring based condition assessment of offshore wind turbine support structures: Dissertation. ETH Zurich, 2011.

[21] T. Braml. Zur Beurteilung der Zuverlässigkeit von Massivbrücken auf der Grundlage der Ergebnisse von Überprüfungen am Bauwerk. Dissertation. München, 2010.

[22] J. Thierling. Vergleich zwischen semiprobabilistischer und probabilistischer Nachweisführung am Beispiel des Dekompressionsnachweises in Querrichtung einer Bestandsbrücke. Masterarbeit. Berlin, 2020.

[23] H. Bachmann et al. Tragwerkszuverlässigkeit. Einwirkungen [8]. *Der Ingenieurbau (Editor: G. Mehlhorn)*. Berlin: Ernst, 1996.

[24] CEN/TC 250. *Eurocode 1: Actions on structures - Part 2: Traffic loads on bridges(EN 1991-2)*. Comitée Européen de Normalisation, 2003-09-01.

[25] E. M. Eichinger. Beurteilung der Zuverlässigkeit bestehender Massivbrücken mit Hilfe probabilistischer Methoden (German). Dissertation. Wien, 2003.

[26] A. Strauss. Stochastische Modellierung und Zuverlässigkeit von Betonkonstruktionen (German). Dissertation. Wien, 2003.

A. Appendix A

Individual stochastic models of the basic variables used to analyze SLS decompression with FORM; cf. [11].

Abbr.	Description	Distr. Type	Mean	Std. dev.	Dim.
$\Theta_{E,N}$	Model uncertainty (normal forces) (values based on [21])	N (Normal)	1,0	$\sigma = 0,05$ CoV = 5,0 %	–
N_G	Normal force due to dead loads	N	-1,1 [22]	$\sigma \approx 0,07$ CoV = 6,0 % [23]	kN/m
$N_{Q,TS}$	Normal force due to traffic loads (TS, load model 1 acc. to EN 1991-2 [24])	GUMBEL	-1,23	$\sigma \approx 0,18$ CoV = 15,0 % [21]	kN/m
$N_{Q,UDL}$	Normal force due to traffic loads (UDL, LM 1 acc. to EN 1991-2)	GUMBEL	1,01	$\sigma \approx 0,15$ CoV = 15,0 % [21]	kN/m
N_P	Normal force due to prestressing	N	-2036 [22]	$\sigma = 203,6$ CoV = 10,0 % [25, 26]	kN/m
N_{K+S}	Normal force due to creep and shrinkage	N	270 [22]	$\sigma = 40,5$ CoV = 15,0 % [25, 26]	kN/m
N_{SE}	Normal force due to load case: subsidence	const.	-0,63 [22]	–	kN/m
N_T	Normal force due to load case: temperature	const.	12,80	–	kN/m
$h_{y=0}$	Height of the cross-section	N	0,327	$\sigma = 0,01$ [8] CoV $\approx$ 3,1 % [25, 26]	m
$\theta_{E,M}$	Model uncertainty (moments) [21]	N	1,0	$\sigma = 0,10$ CoV $\approx$ 10,0 %	–
M_G	Bending moment due to dead loads	N	25,98 [22]	$\sigma \approx 1,56$ CoV = 6,0 % [23]	kNm/m
$M_{Q,TS}$	Bending moment due to traffic loads (TS, LM 1 acc. to EN 1991-2)	GUMBEL	0,98	$\sigma \approx 0,15$ CoV = 15,0 % [21]	kNm/m
$M_{Q,UDL}$	Bending moment due to traffic loads (UDL, LM 1 acc. to EN 1991-2)	GUMBEL	4,87	$\sigma = 0,73$ CoV = 15,0 % [21]	kNm/m
M_{K+S}	Bending moment due to creep and shrinkage	N	-12,0 [22]	$\sigma = 1,8$ CoV = 15,0 % [25, 26]	kNm/m
M_{SE}	Bending moment due to subsidence	const.	0,37 [22]	–	kNm/m
M_T	Bending moment due to temperature	const.	$-0,58$	–	kNm/m
$d_{Sp,init}$	Initial distance between bottom of the slab and the bottom of the tendon duct	N	0,163 [22]	$\sigma = 0,01$ [8] CoV $\approx$ 6,1 %	M
ϵ	Eccentricity of the strands inside the tendon duct	N	0,034 [22]	$\sigma = 0,0068$ [8] CoV = 20,0 %	M
$d_{Sp,GPR}$	Measured distance between bottom of the slab and the bottom of the tendon duct	N	0,168	$\sigma = 0,008$ CoV $\approx$ 4,8 %	M