Quantifying the Required Sample Size for Desired Confidence in Damage Detection of Structural Health Monitoring

SOREN MOLLER, THOMAS POTTHAST,
NIKLAS WINNEWISSER, CLEMENS JONSCHER,
MICHAEL BEER and RAIMUND ROLFES

Soren Moller, Clemens Jonscher, Raimund Rolfes: Leibniz Universita¨t Hannover, Institute of Structural Analysis, Appelstr. 9A. 30167 Hannover, Germany. Email: s.moeller@isd.uni-hannover.de
Thomas Potthast, Niklas Winnewisser: Leibniz Universitat Hannover, Institute for Risk and Reliability, Appelstr. 9A. 30167 Hannover, Germany
Michael Beer: Leibniz Universitat Hannover, Institute for Risk and Reliability, Appelstr. 9A, 30167 Hannover, Germany; Department of Civil and Environmental Engineering, University of Liverpool, Liverpool, L69 3GH, UK; International Joint Research Center for Engineering Reliability and Stochastic Mechanics and International Joint Research Center for Resilient Infrastructure, Tongji University, Shanghai, China

ABSTRACT

In the Structural Health Monitoring (SHM) context, the most critical question is often: Is the system damaged or not? Methods to detect damage rely on identifying novelties in the system response, where significant deviations from training data indicate structural damage. However, this process depends on a threshold, typically defined as a single crisp value based on data distribution or representativeness assumptions. If natural variability, i.e. aleatory uncertainty, is not fully captured, false positives may occur. Probabilistic approaches address aleatory uncertainty, while epistemic uncertainty – arising from lack of knowledge and imprecision – is often modelled using set theory, including interval, fuzzy sets, and other imprecise probability frameworks. Recent research suggests that integrating both types of uncertainty enhances confidence in structural integrity assessments, and fuzzy probabilities offer one feasible approach. This study develops a methodology using fuzzy probabilities to determine the required sample size for a predefined confidence level in damage detection. The fuzzy membership function aggregates plausible intervals into a unified framework, including quantifying how many samples are needed to achieve a certain confidence level. A two-degree-of-freedom system validates the approach, demonstrating its effectiveness in detecting damage compared to a crisp value-based threshold.

INTRODUCTION

Structural Health Monitoring (SHM) employs various strategies to detect and identify structural damage, following a hierarchical five-level approach [1]. The first two levels focus on reliably determining whether damage has occurred and locating it, while the higher levels address damage characterisation and remaining service life estimation. For most civil infrastructures, only healthy state data are available, leading to an unsupervised SHM problem where continuously extracted damage-sensitive features (DSFs) are compared against the healthy state of the structure. Any significant deviation is taken as an indicator of damage [2]. This procedure involves comparing the system response to a threshold to determine whether the structure is healthy. Consequently, the definition of this threshold is critical for enabling damage detection. The SHM literature offers a wide range of methods for both damage detection and threshold determination [3]. Some techniques, such as outlier analysis [4], assume that the DSFs follow a normal distribution, while others, like extreme value statistics [5], do not require this assumption. In practice, however, a decision maker must address various uncertainties present in healthy data – such as measurement noise, the number of available samples, and physical variability – that can compromise decision quality [6]. Although established probabilistic methods exist to handle both aleatory and epistemic uncertainty [7–9], and many authors have successfully investigated the management of uncertain DSFs in SHM [10–12], the threshold value is still typically assumed to be fixed. This leads to decisions based solely on a single crisp value. Since the threshold depends heavily on the available amount of data and thus trustworthiness, it is beneficial to define confidence intervals that account for both aleatory and epistemic uncertainties. This perspective aligns with recent advances in imprecise probability theory that advocate for combining uncertainty types to improve the confidence of structural integrity assessments. Therefore, this study introduces a fuzzy probability-based methodology [13] to determine the minimum number of samples required to achieve a predefined confidence level for damage detection using the Mahalanobis distance. By aggregating plausible intervals into a unified fuzzy membership framework, decision-makers can simultaneously evaluate multiple confidence levels. Figure 1 illustrates the transition zone between healthy and damaged states arising from the fundamental ambiguity in assessing the structure‘s state near the threshold. This inherent uncertainty underscores the need for a probabilistic framework that accommodates gradual transitions rather than enforcing crisp thresholds, ultimately enabling more robust decision-making.

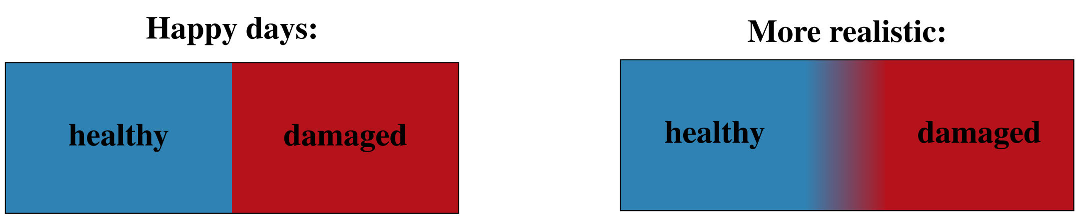

Figure 1. Illustration of the challenge that arises in the transition between healthy and damaged conditions. Happy days: a binary transition defined through a crisp threshold indicating between the healthy and damaged state. More realistic: a fuzzy transition between both states, where no reasonable decision can be made whether the structure is damaged or not.

The proposed approach also offers significant advantages for machine learning-based damage detection in SHM, where training time and computational cost are susceptible to sample size (e.g., in Gaussian process regression). By rigorously establishing the minimum number of samples required to attain a predetermined confidence level, the proposed method not only accelerates model training but also quantitatively assesses data sufficiency for robust performance. The proposed method is validated on a two-degree-of-freedom benchmark system, demonstrating its capacity to improve SHM by enabling more accurate and data-efficient decisions.

FUZZY SET THEORY

The fuzzy set theory enables a gradual assessment of element membership in a set by employing a membership function that relaxes the need for precise values or bounds. This approach allows for a smooth transition between elements that belong to a set and those that do not. A fuzzy number $\tilde{x}$ defined on $X = \mathbf{R}^n$ is represented as

$$\tilde{x} = \{(x, \mu(x))|x \in X\}, \quad \mu(x) \geq 0 \quad \forall x \in X. \tag{1}$$

Here, the membership function $\mu(x)$ assigns each $x \in X$ a real number in the interval $[0, 1]$, indicating the degree of membership of x in $\tilde{x}$. Various functional forms exist for $\mu(x)$; one common example is the triangular membership function:

$$\mu(x) = \begin{cases} \frac{x-a_1}{a_2-a_1}, & \text{if } a_1 \leq x \leq a_2 \\ \frac{a_3-x}{a_3-a_2}, & \text{if } a_2 \leq x \leq a_3 \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The corresponding crisp sets

$$x_\alpha = \{x \in X|\mu(x) \geq \alpha\} \tag{3}$$

are referred to as α-level sets for $\alpha \in (0, 1]$. These sets form a nested sequence with the property that

$$x_{\alpha_k} \subseteq x_{\alpha_i}, \ \forall \alpha_i, \text{ while } \quad \alpha_k \in (0,1] \quad \text{with} \quad \alpha_i \leq \alpha_k. \tag{4}$$

Each α-level set represents an interval defined by minimum and maximum values for the fuzzy input parameter. Figure 2 illustrates a triangular fuzzy number for the parameter x. The fuzzy set containing all elements with a membership degree of at least $\alpha \in (0, 1]$ is called the α-cut of the membership function. Given an α-level, the support of x_α reflects the level of confidence in the corresponding set of parameters. The interval for a triangular fuzzy number at a given α-level is characterised by

$$x_\alpha = [a_1 + \alpha(a_2 - a_1), a_3 - \alpha(a_3 - a_2)], \quad \forall \alpha \in (0, 1], \, a_1 \leq a_2 \leq a_3. \tag{5}$$

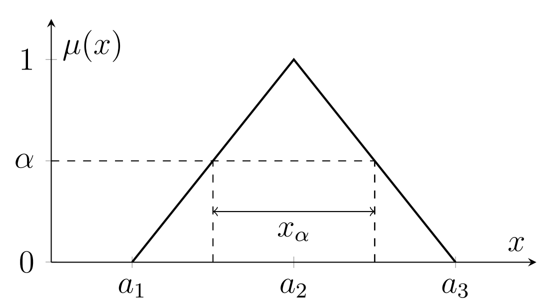

Figure 2. α-level for the triangular membership function $\mu(x)$ with parameters a_1, a_2, a_3.

CASE STUDY

The proposed method is demonstrated and verified using a simple two-degree-of-freedom (2DOF) system, depicted in Figure 3.

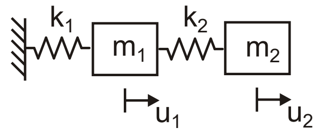

Figure 3. Mechanical model of the 2DOF system.

Analytically, the 2DOF system can be expressed using the equation of motion

$$\boldsymbol{M}\ddot{u} + \boldsymbol{K}u = \boldsymbol{0} \tag{6}$$

with $\boldsymbol{M}$ and $\boldsymbol{K}$ as mass and stiffness matrices. For the system shown in Figure 3, the matrices result in:

$$\boldsymbol{M} = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \quad \boldsymbol{K} = \begin{pmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{pmatrix}. \tag{7}$$

Given the corresponding values for mass and stiffness, the eigenvalues λ_1 and λ_2 of the 2DOF system result in:

$$\lambda_{1,2} = \frac{m_1k_2 + m_2k_1 + m_2k_2}{2m_1m_2} \pm \sqrt{\left(\frac{m_1k_2 + m_2k_1 + m_2k_2}{2m_1m_2}\right)^2 - \frac{k_1k_2}{m_1m_2}}. \tag{8}$$

For this study, the masses are set to $m_1 = m_2 = 1$ kg, and the stiffnesses are set to $k_1 = k_2 = 10$ N/m. To account for uncertainty in modelling, measurement, and identification of the model parameters, the linear spring stiffnesses k_1 and k_2, as well as the eigenvalues λ_1 and λ_2, are corrupted with Gaussian distributed noise. This study considers two distinct sources of uncertainty:

- **Input uncertainty**: Variability in the stiffness parameters k_1 and k_2 of the 2DOF system, modeled by a $\mathcal{N}(0, 0.05)$ distribution.
- **Model/measurement uncertainty**: Noise introduced to the eigenvalues λ_1 and λ_2, following a $\mathcal{N}(0, 0.01)$ distribution.

To enable multivariate outlier detection, the Mahalanobis distance (MD) is employed as a conventional method for multivariate outlier analysis in SHM [14]. The MD is defined as:

$$\mathrm{MD}_i = \sqrt{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{S}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})}. \tag{9}$$

Here, $\boldsymbol{x}_i$ represents the current data point, $\bar{\boldsymbol{x}}$ denotes the mean vector of the training data, and $\boldsymbol{S}$ is the covariance matrix. In this study, the corresponding values of λ_1 and λ_2 are utilised to define the vector $\boldsymbol{x}_i$.

To determine whether the system is in a healthy or damaged state, a threshold is typically established as follows [14]: The Mahalanobis distance (MD) is initially computed using a training dataset, and the threshold is set by identifying the 95% quantile through the Monte Carlo method using 3000 samples. For damage detection, the MD of each new data point is calculated. If the MD exceeds the threshold, the data point is considered novel compared to the training data. For the remainder of this study, a threshold determined in this manner is referred to as the *standard threshold*.

FUZZY THRESHOLD

The primary objective of this study is to determine the amount of data required to achieve a predefined confidence level for the threshold value, thereby improving the confidence of damage detection based on available training data. From an engineering perspective, establishing a confidence level for the threshold and quantifying the necessary data collection is highly beneficial. Following [15], a decision-maker can define a confidence metric as the range between the upper and lower quantiles of a given width for the threshold T, based on n samples:

$$\delta_T^{(\mathrm{n})} = Q_{\mathrm{upper}}(T^{(\mathrm{n})}) - Q_{\mathrm{lower}}(T^{(\mathrm{n})}). \tag{10}$$

By plotting these bounds relative to the data size, one can easily interpret how much data is required to achieve a given confidence level. To dynamically adjust the upper and lower quantiles, this study employs the framework of fuzzy probabilities [16]. The membership function, using α-cuts, provides a unified approach to defining different plausible intervals. Here, the threshold is modelled as a triangular fuzzy number, where parameters a_1, a_2, and a_3 are derived from the mean and the 0- and 1-quantiles of a given sample set n. This approach enables the examination of various quantile ranges via α-cuts without repeatedly computing statistical quantities. Figure 4 illustrates the required sample size to achieve specific quantile ranges for different α-cuts.

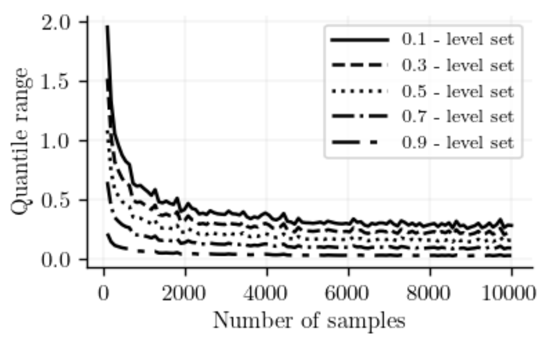

Figure 4. Convergence of the quantile range for the threshold for different α-cuts.

From Figure 4, for an α value of 0.1, a sample size of 1360 is required to reach a quantile range of 0.5. For the same sample size, an α of 0.9 results in a quantile range of 0.0312, while $\alpha = 0.5$ yields a quantile range of 0.158. As expected, lower α values correspond to lower confidence in the threshold, leading to wider confidence intervals.

DISCUSSION AND COMPARISON

For this study, an α-cut of 0.5 is assumed. Given a quantile range of $\delta_T^{(\mathrm{n})} := 0.25$, a total of 1900 samples are required, as indicated by Equation 10. The standard threshold and the MD shown in Figure 6 were determined as described earlier using 1900 samples, with the threshold computed on a set disjoint from the MD training set. As mentioned above, the standard threshold relies on the Monte Carlo method, which introduces inherent scattering. This behaviour is illustrated in Figure 5, where different realisations of the standard threshold were obtained by repeating the procedure 3000 times. It can be observed that the standard threshold follows a normal distribution, indicating that it spans an interval rather than existing as a crisp value. This interval is effectively captured by the fuzzy threshold, underscoring the importance of considering a range of thresholds rather than relying on a single crisp value for damage detection.

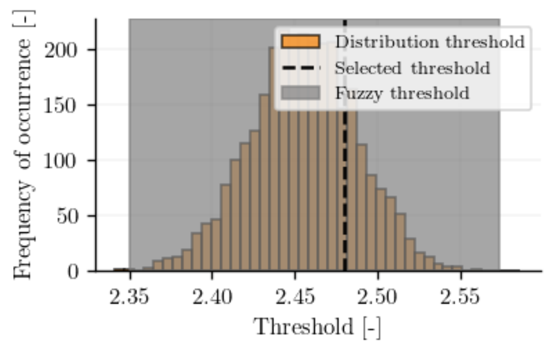

Figure 5. comparison of threshold realisations across various sample draws

Figure 6 shows the MD of 500 unknown samples from the healthy state as well as from the damage state, induced by a reduced stiffness k_2 (reduced from 10 N/m to 8.3 N/m), alongside both the standard threshold and the fuzzy threshold.

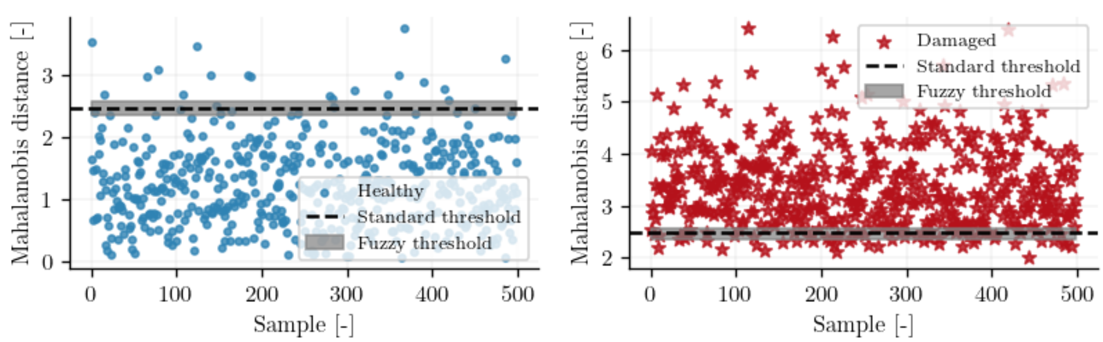

Figure 6. Healthy state of the 2DOF system (left) and damaged state (right).

The true positive rate (TPR) and false positive rate (FPR) are used in this study to judge the accuracy of damage detection. Here, the FPR is solely calculated based on the healthy data, while the true positive rate TPR is exclusively determined from the data of the damaged state. Moreover, a *fuzzy ratio* is introduced, which is defined as the number of samples within the α-cut divided by the overall number of samples of the current

TABLE I. Comparison of TPR, FPR, and fuzzy ratio for standard and fuzzy threshold.

	TPR	**FPR**	**Fuzzy ratio (healthy)**	**Fuzzy ratio (damaged)**
Standard	0.894	0.042	-	-
Fuzzy ($\alpha = 0.5$)	0.940	0.039	0.030	0.094
Fuzzy ($\alpha = 0$)	0.975	0.029	0.046	0.198

state of the 2DOF system. Table I lists results for the healthy and damaged states and demonstrates that the standard threshold performs worse in terms of damage detection compared to the approach based on the fuzzy threshold. Since the fuzzy threshold excludes samples that fall within its confidence interval, this method yields more accurate results for damage detection than the standard threshold. Furthermore, selecting a lower α (e.g., $\alpha = 0$) results in more samples being excluded, ensuring that only the most certain ones are considered for damage detection. As a result, the accuracy of damage detection improves, albeit at the cost of reducing the number of available samples.

CONCLUDING REMARKS

In this work, a 2DOF system was used to demonstrate a methodology based on fuzzy probabilities to determine the required number of samples for a given confidence level. Confidence intervals for the threshold were derived using α-cuts, allowing for better representation of statistical fluctuations and more accurate decisions near the threshold. The presented approach effectively determined the necessary sample size while accounting for the threshold's inherent stochastic variability. Compared to the standard method, the fuzzy-based approach provided a more robust distinction between healthy and damaged states using the MD. This study thus laid the foundation for more reasonable decision-making in the presence of uncertainties, which is particularly important when engineers are challenged a priori to determine the sample size to guarantee a certain confidence in the threshold used for damage detection within SHM. However, this study represents an initial investigation. Limitations include the use of only a triangular fuzzy number for α-cuts and the exclusive consideration of the MD for outlier detection. Future work should explore additional outlier detection methods, alternative fuzzy numbers for threshold definition, and different metrics for confidence evaluation. Additionally, incorporating targeted sampling strategies could further reduce uncertainties.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - subprojects C01 and C02, grant number 434502799, SFB 1463 and subproject D01, grant number 2388 501624329, Priority Program SPP 100+.

REFERENCES

1. Worden, K. 2010. *Structural Health Monitoring using Pattern Recognition*, Springer Vienna, Vienna, ISBN 978-3-7091-0399-9, pp. 183–246, doi:10.1007/978-3-7091-0399-9_5.
2. Farrar, C. R. and K. Worden. 2007. "An introduction to structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):303–315, doi:10.1098/rsta.2006.1928.
3. Falcetelli, F., N. Yue, R. D. Sante, and D. Zarouchas. 2022. "Probability of detection, localization, and sizing: The evolution of reliability metrics in Structural Health Monitoring," *Structural Health Monitoring*, 21(6):2990–3017, doi:10.1177/14759217211060780.
4. Worden, K., H. Sohn, and C. Farrar. 2002. "Novelty detection in a changing enviroment: regression and interpolation approaches," *Journal of Sound and Vibration*, 258(4):741–761, doi:10.1006/jsvi.2002.5148.
5. Roberts, S. J. 1999. "Novelty detection using extreme value statistics," *IEE Proceedings - Vision, Image and Signal Processing*, 146:124–129, doi:10.1049/ip-vis:19990428.
6. Jonscher, C., S. Möller, L. Liesecke, D. Schuster, B. Hofmeister, T. Grießmann, and R. Rolfes. 2024. "Identification Uncertainties of Bending Modes of an Onshore Wind Turbine for Vibration-Based Monitoring," *Structural Control and Health Monitoring*, 2024:3280697, doi:10.1155/2024/3280697.
7. Beer, M., S. Ferson, and V. Kreinovich. 2013. "Imprecise probabilities in engineering analyses," *Mechanical Systems and Signal Processing*, 37(1-2):4–29, doi:10.1016/j.ymssp.2013.01.024.
8. Valdebenito, M. A., M. Beer, H. A. Jensen, J. Chen, and P. Wei. 2020. "Fuzzy failure probability estimation applying intervening variables," *Structural Safety*, 83:101909, doi:10.1016/j.strusafe.2019.101909.
9. Faes, M. G., M. Daub, S. Marelli, E. Patelli, and M. Beer. 2021. "Engineering analysis with probability boxes: A review on computational methods," *Structural Safety*, 93:102092, doi:10.1016/j.strusafe.2021.102092.
10. Jonscher, C., S. Möller, L. Liesecke, B. Hofmeister, T. Grießmann, and R. Rolfes. 2024. "Heteroscedastic Gaussian processes for data normalisation in probabilistic novelty detection of a wind turbine," *Engineering Structures*, 305:117786, doi:10.1016/j.engstruct.2024.117786.
11. Qiu, Z., Y. Ma, and X. Wang. 2004. "Comparison between non–probabilistic interval analysis method and probabilistic approach in static response problem of structures with uncertain–but–bounded parameters," *Communications in Numerical Methods in Engineering*, 20(4):279–290, doi:10.1002/cnm.668.
12. Liu, G. and Z. Mao. 2017. "Structural damage diagnosis with uncertainties quantified using interval analysis," *Structural Control and Health Monitoring*, 24(10):e1989, doi:10.1002/stc.1989.
13. Kwakernaak, H. 1978. "Fuzzy random variables—I. definitions and theorems," *Information Sciences*, 15(1):1–29, doi:10.1016/0020-0255(78)90019-1.
14. Worden, K., G. Manson, and N. Fieller. 2000. "Damage detection using outlier analysis," *Journal of Sound and Vibration*, 229(3):647–667, doi:10.1006/jsvi.1999.2514.
15. Zhang, J. and M. D. Shields. 2018. "On the quantification and efficient propagation of imprecise probabilities resulting from small datasets," *Mechanical Systems and Signal Processing*, 98:465–483, doi:10.1016/j.ymssp.2017.04.042.
16. Hall, J. W. and J. Lawry. 2004. "Generation, combination and extension of random set approximations to coherent lower and upper probabilities," *Reliability Engineering and System Safety*, 85(1):89–101, doi:10.1016/j.ress.2004.03.005.