UNCECOMP 2023
5th ECCOMAS Thematic Conference on
Uncertainty Quantification in
Computational Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Athens, Greece, 12-14 June 2023

UNCERTAINTIES IN THE SYNTHETIC DATA GENERATION FOR THE CREATION OF BRIDGE DIGITAL TWINS

Alejandro Jiménez Rios[1], Vagelis Plevris[2] and Maria Nogal[3]

[1] Structural Engineering Research Group (SERG), Department of Built Environment (DBE), Faculty of Technology, Art, and Design (TKD), Oslo Metropolitan University (OsloMet)
Rebel Building, Universitetsgata 2, 0164 Oslo, Norway
e-mail: alejand@oslomet.no

[2] Department of Civil and Architectural Engineering, Qatar University
Doha P.O. Box 2713, Qatar
e-mail: vplevris@qu.edu.qa

[2] Materials, Mechanics, Management & Design Department, Delft University of Technology
2628 CN Delft, The Netherlands
e-mail: m.nogal@tudelft.nl

Abstract

Digital twins (DTs) are virtual replicas of physical assets that can be used to monitor and manage their performance. To date, the DT concept has been effectively implemented in various industries, including aeronautics, manufacturing, medicine, and more recently, in the architecture, engineering, and construction sector. In the latter, these assets can be related to buildings, bridges, or other important infrastructures of the built environment. Although the creation of synthetic benchmark datasets for the validation of novel damage detection approaches has been attempted in the past, such alternatives are not easily findable or accessible. Thus, a new synthetic data generation framework is proposed within the DT paradigm context, that can produce FAIR benchmark databases that are characterized by Findability, Accessibility, Interoperability, and Reuse. This paper aims at exploring the uncertainty types, sources, and quantification approaches involved in the synthetic data generation methodologies and tools of the intended framework which could be used as a faster and cheaper alternative to real monitoring, for the creation and development of DT prototypes of bridges for both industry and research-oriented purposes. This work also highlights the benefits and drawbacks of implementing synthetic data for these purposes and points out tentative future improvements in the field.

Keywords: Digital Twins, Bridges, Synthetic FAIR Data, Prototyping, Uncertainties.

1 INTRODUCTION

The Architecture, Engineering, and Construction (AEC) industry is rapidly developing and adopting the concepts of the Digital Twin (DT) paradigm. Although there are still numerous challenges and obstacles to overcome before the complete deployment and widespread implementation of DTs, there is a shared vision and consensus among the scientific community and bridge practitioners regarding the potential of the technology for revolutionizing bridge design, management, and operation procedures [1]. However, to establish a proper application framework that can be adopted on a large scale by the industry, further research is required.

As the AEC industry delves deeper into the development and adoption of the DT paradigm, the need for a reliable benchmark database for testing and validating newly designed technologies and algorithms have arisen. Both development and validation processes are better performed during a prototyping stage. Prototypes play a critical role in bridging the gap between ideation and implementation, enabling a smooth transition from design to production and implementation [2]. In addition, when it comes to implementing new technologies and materials on bridges with cultural heritage value [3-6], comprehensive validation is advised before such interventions can be deemed adequate, in compliance with the guidelines set forth by the ICOMOS International Scientific Committee on the Analysis and Restoration of Structures of Architectural Heritage (ISCARSAH) [7]. The Committee plays an important role in ensuring the preservation and continued use of historic buildings and structures around the world, and in promoting the recognition and protection of cultural heritage sites as vital components of our global heritage.

Although the creation of similar synthetic benchmark datasets has been proposed in the literature in the past [8], such alternatives are hard to find and even harder to access. Thus, newly created synthetic databases should be made available, while being “Findable, Accessible, Interoperable, and Reusable”. In other words, such data should adhere to the so-called “FAIR” principles [9]. These principles were developed to address the challenges of managing and sharing data in an increasingly complex and interconnected digital landscape. With this idea in mind, a novel synthetic FAIR data generation framework within the context of the DT paradigm has been recently proposed [10] (see Figure 1).

In this framework, a series of benchmark databases containing meaningful data including different damage scenarios will be created. It will account for the creation of multi-metric data, namely, vibration, strain, visual and mixed synthetic data under both undamaged and damaged scenarios. Both environmental and operational conditions can be fine-tuned and included in the data generation process. A synchronizing module will ensure that all data can be correctly tracked over time.

There are several uncertainties that need to be considered in a synthetic data generation process, that have to do with data quality, model accuracy, variability, data coverage, uncertainty propagation, among others. These uncertainties can have significant impacts on the accuracy and reliability of the DT. Thus, they should be carefully considered and managed to ensure that the digital twin accurately represents the behavior of the bridge under a wide range of real-world conditions. The proposed framework takes this into account by considering both epistemic and aleatory uncertainties for the adequate generation of realistic scenarios [11, 12], which are key features to be accounted for in the validation process of any newly developed technology. The data generated would be suitable for use in the development and validation of model-based, data-driven, and physics-informed components for damage detection, localization, description, and prognosis of the bridge DT, thus reducing the time and money required for the creation of novel prototypes. Within this context, the synthetically generated data would mock the physical asset components of the DT.

In this paper, the uncertainties component of the framework is presented, explored, and discussed in detail. The rest of the manuscript is organized as follows: in Section 2 of the paper the source, description, and advised mitigation measurements of the considered uncertainties within the proposed synthetic data generation framework are presented. In Section 3 the strategy adopted for the inclusion of the main types of uncertainty within the generated data is exposed and discussed, and finally, in Section 4 conclusions are drawn and further work is proposed.

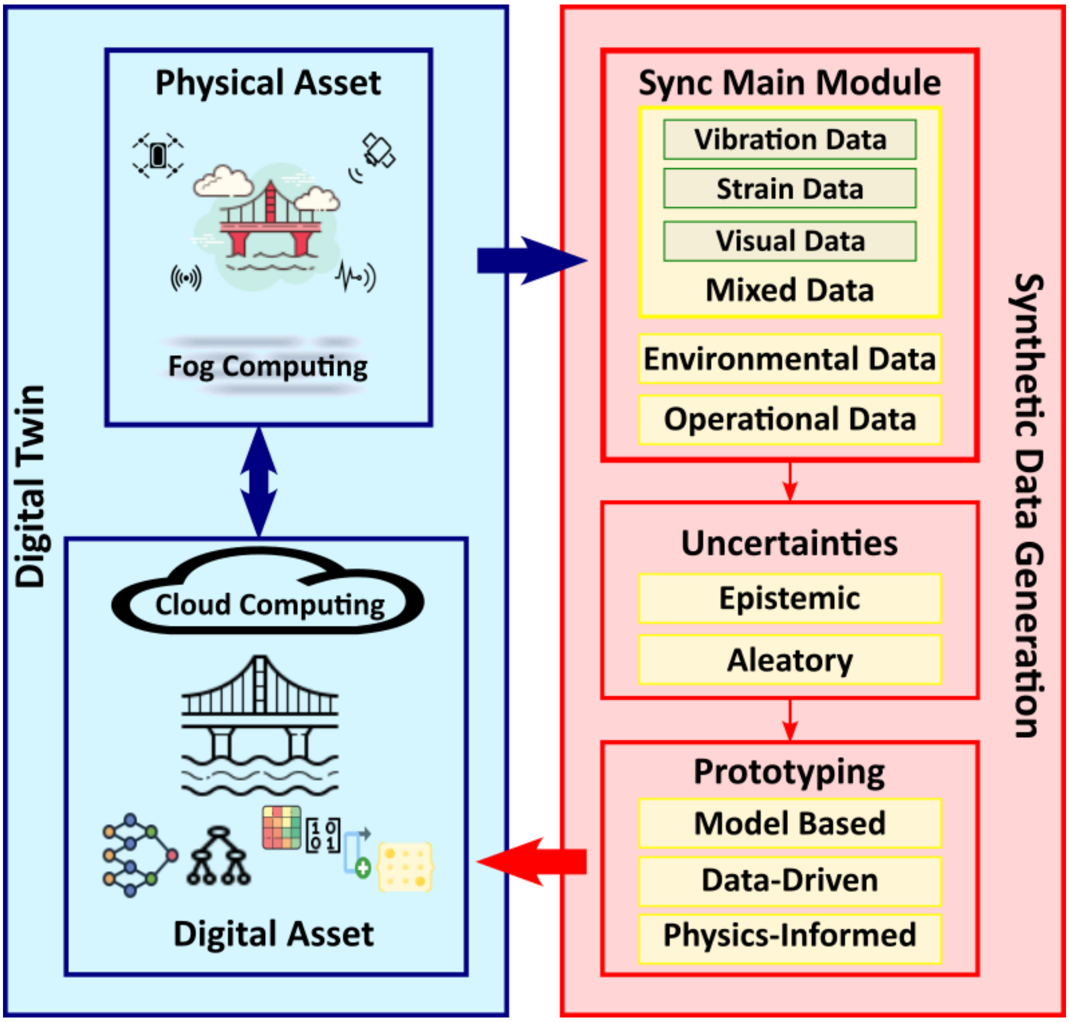

Figure 1. The proposed framework for synthetic data generation, which accounts for uncertainties and could be used in the prototyping of DTs.

2 DATA UNCERTAINTIES

Aleatory and epistemic uncertainties are two types of uncertainties commonly encountered in engineering and scientific fields [13]. Aleatory uncertainty arises from the inherent variability or randomness in the system being studied. It is also known as natural variability or irreducible uncertainty. Such uncertainties cannot be easily reduced or eliminated by simply acquiring more data or improving the employed measurement techniques. Examples include natural disasters, random fluctuations in material properties, and variability in environmental conditions. Table 1 presents the main sources of aleatory uncertainty that could affect the data quality collected on a real bridge DT project.

Source	Description	Mitigation
Measurement errors	Sensor misalignment, installation errors, or sensor damage can affect both vibration-based [14] and strain-based [15] damage detection reliability.	Using high-quality, carefully calibrated, properly installed/maintained sensors, and adequate filtering techniques.
Image Quality	Camera resolution, focus, and exposure settings produce low-quality images that preclude visual-based [16] approaches.	Using high-quality, properly calibrated, and well-maintained cameras.

Table 1. Aleatory uncertainties: Sources, description, and corresponding mitigation actions.

Source	Description	Mitigation
Environmental	Temperature variations influence the dynamic response of bridges as well as the strain distribution on them, which can affect vibration-based [17] and strain-based [18] damage detection results, respectively. Visual-based methods are highly susceptible to changes in visibility-related environmental effects such as the presence of rain, mist, or fog [19].	Measuring environmental conditions at the time of data collection and accounting for their effects on the dynamic response of the bridge. Collecting visual data under consistent lighting and viewing conditions.
Modeling	Simplified finite element analysis or other modeling techniques introduce errors that reduce the accuracy of vibration-based [20] damage detection results.	Using high-quality models that accurately reproduce the bridge's geometry, material, and loading conditions.
Baseline	The accuracy of both vibration-based [21] and strain-based [22] damage detection results depends on the accuracy of the baseline information, which is collected under undamaged scenarios.	Improving data collection and collecting data over a longer period, as well as implementing statistical analysis to detect and discard outliers.
Damage location	Vibration-based [23, 24] and strain-based [25] damage detection techniques are typically most effective at identifying damage at mid-span or near supports locations, whereas visual-based [26] damage detection techniques are typically most effective at identifying damage on the bridge surface.	Increasing the number and quality of sensors and/or implementing sensor placement optimization techniques.
Material variability	The material properties of the bridge components, such as the modulus of elasticity and Poisson's ratio of the bridge materials, may vary due to manufacturing variability or aging. This can introduce uncertainties in the strain-based [27] damage detection results.	Increasing the number of sensors and probe specimens along the bridge.
Image Interpretation	The interpretation of visual data can be subjective and dependent on the experience and expertise of the analyst. This can introduce uncertainties in the visual-based damage detection results [28].	Establishing clear guidelines and criteria for image interpretation and providing training to the analysts.

Table 2. Epistemic uncertainties: Sources, description, and corresponding mitigation actions.

Epistemic uncertainty, on the other hand, arises from incomplete knowledge or understanding of the system being studied. It is also known as reducible uncertainty because it can be reduced or eliminated by acquiring more data or improving the understanding of the system. Examples include measurement errors, modeling assumptions, and incomplete knowledge of the system being studied. Table 2 contains the sources of epistemic uncertainty considered in the proposed synthetic data generation framework within the DT paradigm.

Overall, it is important to carefully consider and quantify these uncertainties to ensure that regardless of the damage detection, localization, quantification, and/or prognosis approach used (i.e., model-based, data-driven, or physics-informed), results will be accurate and reliable. The Uncertainty Quantification (UQ) can be done using several methodologies and techniques [29, 30], such as:

- Fuzzy methods, which map responses to structural parameters using fuzzy clustering. As a drawback, these methods rely on membership functions, also called grades, which are difficult to define accurately, thus inducing new sources of uncertainty.
- Probabilistic methods, which rely on the use of Probability Density Functions (PDFs), a likelihood function, and weighting coefficients. They are highly reliant on the chosen stochastic method and are relatively expensive in terms of computational cost.
- Interval methods, which do not require assumptions made in terms of PDFs, discrete sampling techniques nor modal measure-of-fit but can explore the entire space within the feasible parameters' range.

Mixed damage detection approaches on bridges, which combine multiple types of data such as vibration, strain, and visual data, can help to reduce the uncertainties associated with each individual data type. However, further uncertainties associated with mixed damage detection methods could arise leading to some undesirable scenarios, as explained below:

- Integration of data [31]: Integrating multiple types of data can be challenging and requires careful consideration of the uncertainties associated with each type of data and their possible interactions.
- Calibration and synchronization [32]: The calibration and synchronization of the different types of sensors used for data collection can affect the accuracy of the mixed damage detection results.
- Uncertainty Propagation [33]: Uncertainties in the individual data types can propagate through the data integration process, which can affect the accuracy of the mixed damage detection results.

Furthermore, the implementation of mixed damage detection approaches can also be hindered by their increased cost and complexity, in comparison to using a single type of data. Finally, it is worth noting that regardless of the damage detection approach employed, the damage detection threshold adopted will be of paramount importance and can vary depending on the type and severity of the damage under study. Small or subtle damage may be difficult to detect, which can introduce uncertainties and lead to undesirable results with limited reliability.

3 UNCERTAINTIES MODULE AND DISCUSSION

The epistemic uncertainties commonly present in real-world study cases will not represent a concern within the context of the synthetic databases generated by the proposed framework as the undamaged scenario data would correspond to the so-called "*ground truth*", in other words, an accurate and objective baseline model. Nevertheless, environmental uncertainties will be included in the creation of environmental data, and those related to modeling and material variability will be represented by the generation of different bridge models. Finally, damage location and image interpretation would be accounted for by the generation of several damaged scenarios.

On the other hand, the proposed framework will generate data with different levels of aleatory uncertainty so that the prototypes could be validated in face of different feasible study case scenarios. Measurement errors can be synthetically generated by the inclusion of artificial noise in the data, with different magnitudes [34]. Uncertainties related to the quality of digital images and the effect of real-world noise could be properly simulated by reducing the image resolution and adding artificial noise, i.e., by introducing random variations in the pixel values of the digital image. In general, adding noise is a common technique used in various fields, such as computer vision, image processing, and data science, to simulate real-world scenarios or to enhance the robustness of algorithms.

4 CONCLUSIONS AND FURTHER WORK

A generalized shift towards a Digital Twin paradigm by members of the engineering, architecture, and construction industry is evident. As this new approach is currently in its early stages of development and adoption, new proposals to optimize its application need to be developed. Such attempts require the development of prototypes that need to be validated against benchmark data. Thus, a framework for the creation of a synthetic data generation tool has been proposed capable of producing high-quality FAIR data that allows novel developed prototypes to be validated and consequently be implemented on further stages of the Digital Twin creation for real infrastructure assets.

This paper explored the uncertainty types, sources, and quantification approaches involved in the synthetic data generation methodologies and tools of the intended framework. Further work needs to be done, mainly within three directions: (i) operationalization of the proposed framework; (ii) self-validation of the generated synthetic data; and (iii) continuous maintenance/support. The results of these attempts will be presented by the authors in future publications.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 101066739.

GLOSSARY

AEC	Architecture, Engineering, and Construction
DT	Digital Twin
FAIR	Findability, Accessibility, Interoperability, and Reuse
ICOMOS	International Council on Monuments and Sites
ISCARSAH	International Scientific Committee on the Analysis and Restoration of Structures of Architectural Heritage
PDF	Probability Density Function
UQ	Uncertainty Quantification

REFERENCES

[1] Jiménez Rios, A., V. Plevris, and M. Nogal, *Bridge management through digital twin-based anomaly detection systems: A systematic review.* Frontiers in Built Environment, 2023. **9**(1176621) DOI: https://doi.org/10.3389/fbuil.2023.1176621.

[2] Deserti, A., et al., *Co-design for society in innovation*, in *Atlas of Social innovation 2nd volume: A world of new practice*. 2019, Oekom. p. 91-96.

[3] Jiménez Rios, A. and D. O'Dwyer. *External Post-tensioning System for the Strengthening of Historical Stone Masonry Bridges*. in *Structural Analysis of Historical*

Constructions. 2019. Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-99441-3_168.

[4] Jurina, L. and E.O. Radaelli. *"Reinforced Arch Method" as Retrofitting Technique for Masonry Arches. Experimental Tests and Numerical Modelling*. in *Proceedings of ARCH 2019*. 2020. Cham: Springer International Publishing.

[5] Kioumarsi, M., V. Plevris, and A. Shabani, *Vulnerability assessment of cultural heritage structures*, in *8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2022)*. 2022: Oslo, Norway. DOI: https://doi.org/10.23967/eccomas.2022.294.

[6] Shabani, A., et al., *3D simulation models for developing digital twins of heritage structures: challenges and strategies.* Procedia Structural Integrity, 2022. **37**: p. 314-320 DOI: https://doi.org/10.1016/j.prostr.2022.01.090.

[7] Roca, P., L. Pelà, and C. Molins. *The Iscarsah Guidelines on the Analysis, Conservation and Structural Restoration of Architectural Heritage*. in *12th International Conference on Structural Analysis of Historical Constructions (SAHC)*. 2021.

[8] Johnson, E.A., et al., *Phase I IASC-ASCE Structural Health Monitoring Benchmark Problem Using Simulated Data.* Journal of Engineering Mechanics, 2004. **130**(1): p. 3-15 DOI: https://doi.org/10.1061/(ASCE)0733-9399(2004)130:1(3).

[9] Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship.* Scientific Data, 2016. **3**(1): p. 160018 DOI: https://doi.org/10.1038/sdata.2016.18.

[10] Jiménez Rios, A., V. Plevris, and M. Nogal, *Synthetic Data Generation for the Creation of Bridge Digital Twins What-if Scenarios*, in *9th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2023)*. 2023, ECCOMAS: Athens, Greece.

[11] Peng, T., et al., *Planning low-error SHM strategy by constrained observability method.* Automation in Construction, 2021. **127**: p. 103707 DOI: https://doi.org/10.1016/j.autcon.2021.103707.

[12] Peng, T., et al. *Role of Sensors in Error Propagation with the Dynamic Constrained Observability Method*. Sensors, 2021. **21**, DOI: https://doi.org/10.3390/s21092918.

[13] Plevris, V., *Innovative computational techniques for the optimum structural design considering uncertainties*. 2009, National Technical University of Athens: Athens, Greece. p. 312.

[14] Kernicky, T., M. Whelan, and E. Al-Shaer, *Vibration-based damage detection with uncertainty quantification by structural identification using nonlinear constraint satisfaction with interval arithmetic.* Structural Health Monitoring, 2018. **18**(5-6): p. 1569-1589 DOI: https://doi.org/10.1177/1475921718806476.

[15] Whelan, M.J. and M.V. Gangone, *Effect of measurement uncertainties on strain-based damage diagnostics for highway bridges.* Journal of Civil Structural Health Monitoring, 2015. **5**(3): p. 321-335 DOI: https://doi.org/10.1007/s13349-015-0110-2.

[16] Ri, S., et al., *Dynamic Deformation Measurement by the Sampling Moiré Method from Video Recording and its Application to Bridge Engineering.* Experimental Techniques, 2020. **44**(3): p. 313-327 DOI: https://doi.org/10.1007/s40799-019-00358-4.

[17] Avendaño-Valencia, L.D. and E.N. Chatzi, *Sensitivity driven robust vibration-based damage diagnosis under uncertainty through hierarchical Bayes time-series representations.* Procedia Engineering, 2017. **199**: p. 1852-1857 DOI: https://doi.org/10.1016/j.proeng.2017.09.111.

[18] Shi, Q., et al., *Uncertain identification method of structural damage for beam-like structures based on strain modes with noises.* Applied Mathematics and Computation, 2021. **390**: p. 125682 DOI: https://doi.org/10.1016/j.amc.2020.125682.

[19] Spencer, B.F., V. Hoskere, and Y. Narazaki, *Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring.* Engineering, 2019. **5**(2): p. 199-222 DOI: https://doi.org/10.1016/j.eng.2018.11.030.

[20] Ghiasi, R., et al. *Uncertainty Handling in Structural Damage Detection via Non-Probabilistic Meta-Models and Interval Mathematics, a Data-Analytics Approach*. Applied Sciences, 2021. **11**, DOI: https://doi.org/10.3390/app11020770.

[21] Huang, Q., P. Gardoni, and S. Hurlebaus, *A probabilistic damage detection approach using vibration-based nondestructive testing.* Structural Safety, 2012. **38**: p. 11-21 DOI: https://doi.org/10.1016/j.strusafe.2012.01.004.

[22] Alavi, A.H., et al., *An intelligent structural damage detection approach based on self-powered wireless sensor data.* Automation in Construction, 2016. **62**: p. 24-44 DOI: https://doi.org/10.1016/j.autcon.2015.10.001.

[23] Limongelli, M.G. and A. Fathi, *The interpolation method for vibration based damage localization: influence of feature uncertainties*, in *Maintenance, safety, risk, management and life-cycle performance of bridge*. 2018. p. 953-960.

[24] Georgioudakis, M. and V. Plevris, *Investigation of the performance of various modal correlation criteria in structural damage identification*, in *ECCOMAS Congress 2016 - Proceedings of the 7th European Congress on Computational Methods in Applied Sciences and Engineering*, M. Papadrakakis, et al., Editors. 2016: Crete Island, Greece. p. 5626-5645. DOI: https://doi.org/10.7712/100016.2207.11846.

[25] Fang, S.-E. and J.-Y. Huang *Statics-Based Model-Free Damage Detection under Uncertainties Using Modal Interval Analysis*. Materials, 2020. **13**, DOI: https://doi.org/10.3390/ma13071567.

[26] Kang, D., et al., *Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning.* Automation in Construction, 2020. **118**: p. 103291 DOI: https://doi.org/10.1016/j.autcon.2020.103291.

[27] Liu, G., Z. Mao, and J. Luo, *Damage detection with interval analysis for uncertainties quantification.* 2015.

[28] Rinaldi, C., J. Ciambella, and V. Gattulli, *Image-based operational modal analysis and damage detection validated in an instrumented small-scale steel frame structure.* Mechanical Systems and Signal Processing, 2022. **168**: p. 108640 DOI: https://doi.org/10.1016/j.ymssp.2021.108640.

[29] Simoen, E., G. De Roeck, and G. Lombaert, *Dealing with uncertainty in model updating for damage assessment: A review.* Mechanical Systems and Signal Processing, 2015. **56-57**: p. 123-149 DOI: https://doi.org/10.1016/j.ymssp.2014.11.001.

[30] Mo, J., L. Wang, and K. Gu, *A two-step interval structural damage identification approach based on model updating and set-membership technique.* Measurement, 2021. **182**: p. 109464 DOI: https://doi.org/10.1016/j.measurement.2021.109464.

[31] Soman, R., et al., *Numerical evaluation of multi-metric data fusion based structural health monitoring of long span bridge structures.* Structure and Infrastructure Engineering, 2018. **14**(6): p. 673-684 DOI: https://doi.org/10.1080/15732479.2017.1350984.

[32] Bao, Y., et al., *Data Fusion-Based Structural Damage Detection Uder Varying Temperature Conditions.* International Journal of Structural Stability and Dynamics, 2012. **12**(06): p. 1250052 DOI: https://doi.org/10.1142/S0219455412500526.

[33] Datteo, A., et al., *On the use of AR models for SHM: A global sensitivity and uncertainty analysis framework.* Reliability Engineering & System Safety, 2018. **170**: p. 99-115 DOI: https://doi.org/10.1016/j.ress.2017.10.017.

[34] Georgioudakis, M. and V. Plevris, *A Combined Modal Correlation Criterion for Structural Damage Identification with Noisy Modal Data.* Advances in Civil Engineering, 2018. **2018**(3183067): p. 20 DOI: https://doi.org/10.1155/2018/3183067.