Available online at www.eccomasproceedia.org
Eccomas Proceedia COMPDYN (2023) 4801-4809

ECCOMAS
Proceedia

COMPDYN 2023
9th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Athens, Greece, 12-14 June 2023

SYNTHETIC DATA GENERATION FOR THE CREATION OF BRIDGE DIGITAL TWINS WHAT-IF SCENARIOS

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

The Digital Twin (DT) concept, as understood nowadays, appeared in the early 2000s as an attempt to create virtual replicas of physical assets, such as bridges, that can be used to examine, monitor and manage their performance. Up to this day, it has been successfully applied in the fields of aeronautics, manufacturing, medicine, and more recently, in the architecture, engineering, and construction industry. The DT of a bridge requires the creation of a virtual replica of the real-life asset, along with the connection and feedback of information channel between the two of them. This connection is currently achieved through the generation of real-time data by the placement of sensors in the real bridge and the application of structural health monitoring techniques to analyze such data. This connection could result in a complex, time-consuming, and expensive process which would hinder the creation of DT prototypes for development purposes in the bridge engineering field. This paper aims at exploring the currently available synthetic data generation methodologies and tools, which could be used as a faster and a more economically feasible alternative to real monitoring, for the creation and development of DT prototypes of bridges, for both industry and research-oriented purposes. A synthetic data generation framework is proposed that can produce FAIR benchmark databases that are based on Findability, Accessibility, Interoperability, and Reuse, which could be used in the prototyping of bridge DTs. Finally, tentative future improvements in this topic are discussed.

Keywords: Digital Twins, Bridges, Synthetic FAIR Data, Prototyping, Damage Detection.

ISSN:2623-3347 © 2023 The Authors. Published by Eccomas Proceedia.
Peer-review under responsibility of the organizing committee of COMPDYN 2023.
doi: 10.7712/120123.10760.21262

1 INTRODUCTION

Digital twins (DTs) are virtual replicas of physical assets, such as bridges, that can be used to monitor and manage their performance. By creating a digital twin of a built asset, the Architecture, Engineering, and Construction (AEC) professionals can monitor its performance in real-time and identify potential issues before they become serious problems. This can lead to improved safety and reduced maintenance costs, especially for the case of vulnerable cultural heritage structures [1, 2] where there are additional challenges because of the special nature and requirements for these valuable assets [3, 4].

The DT paradigm is nowadays in its early stages of development/adoption within the AEC industry. Nevertheless, and despite all the challenges and constraints that need to be overcome for its full deployment and broad implementation, there seems to be a unified vision and consensus toward the future adoption of DT for bridge design, management, and operation among the scientific community and bridge practitioners [5]. Thus, further research is required before a proper application framework is well established and widely adopted by the industry.

Among the several frameworks proposed for the materialization of DTs for bridges, one that can account also for the cultural heritage of existing bridges was proposed as the integration of fully inter-operable data, geometry, finite element, and data-driven modules within an As-Is [6] Historical Bridge Information Model (AI-HBrIM). The interconnection between the real and digital assets within this AI-HBrIM (and also within other frameworks) is achieved through the continuous real-time collection of multi-metric [7] structural, environmental, and operational data. These data are normally collected by a series of sensors organized within a Bridge Health Monitoring (BHM) system. To optimize the data collection, processing, and storage, such a framework would implement optimized sampling methodologies along with fog and cloud computing services.

The main advantages of implementing a DT paradigm are that the data-driven surrogate models allow real-time damage detection and early warning alerts, whereas the detailed finite element models of the asset are used for damage prognosis purposes and the simulation of what-if scenarios that, in combination with probabilistic and reliability analysis, aid asset managers to take informed decisions about the optimal maintenance, retrofitting and repairing of the physical asset. This new paradigm results in extended life, operation cost reduction, as well as increased resilience and sustainability of the built environment.

Unfortunately, BHM demands a lot of resources, both economic ones and in terms of time. Other than a few well-known benchmark bridges, i.e., Z24 [8], Dona [9], I-40 [10], data are not available for exploitation on open-source databases by researchers working in the field. Moreover, to develop tools capable of damage detection [11], the collection of meaningful data (an actual damage scenario) may not be available within the time frame of most research and development projects (for example, an MSc or Ph.D. thesis, a Postdoctoral research project, etc). Furthermore, the data collected through BHM can be affected by both epistemic and aleatory uncertainties [12] and it is necessary to implement adequate strategies to reduce the estimation error of prediction models using such data [13].

Therefore, the need for a reliable database of benchmark data arises so that new technologies and algorithms can be properly tested and validated within the prototyping stage for bridge DTs. Prototypes are essential means to move from design to production and implementation, thus allowing to overcome the classic gap between ideation and implementation [14]. Furthermore, for the implementation of any new technologies and/or materials [15, 16] on bridges with cultural heritage value, such interventions need to be thoroughly validated before they could be considered adequate in accordance with the guidelines of the ICOMOS International Scientific Committee on the Analysis and Restoration of Structures of Architectural Heritage (ISCARSAH Guidelines) [17].

In this paper, we propose a new synthetic data generation framework within the context of the DT paradigm. The framework will be used to generate a series of benchmark databases containing meaningful data including different damage scenarios which could be used in the development and validation of DT components thus reducing the time and money required for the creation of novel prototypes. Within that context, the synthetically generated data would mock the physical asset components of the DT. This data will be particularly suitable for the prototyping of model-based, data-driven, and/or physics-informed components used for damage detection, localization, description, and prognosis of bridge DT. In Section 2 of this paper, the required data is described. In Section 3 the proposed framework is exposed and discussed. Finally, in Section 4 conclusions are drawn, and further work is proposed.

2 REQUIRED DATA

The benchmark data required for prototyping validation depends on the type of anomaly detection algorithm being tested. Anomaly detection algorithms can be classified into four general categories: (i) vibration-based; (ii) strain-based; (iii) visual-based; and (iv) mixed, as shown schematically in Figure 1.

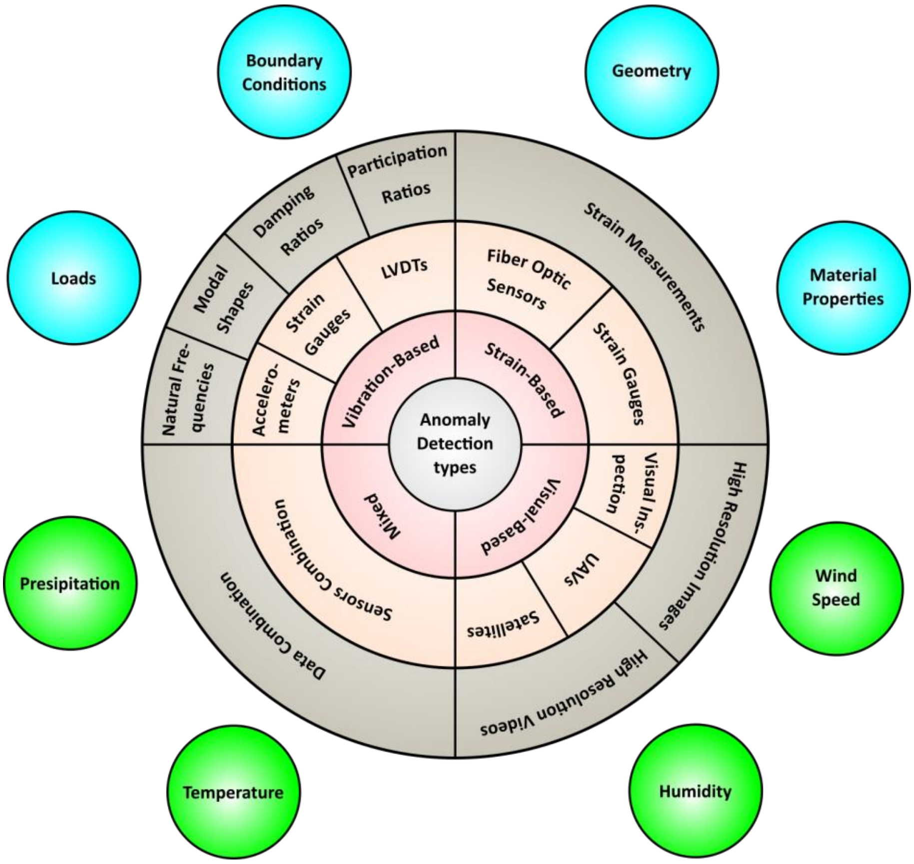

Figure 1. Anomaly detection types are classified based on the data type required, how data is collected, the type of data collected along with environmental and operational conditions.

Vibration-based damage identification on bridges typically requires data on the dynamic response of the bridge [18], which can be obtained through various types of sensors, such as accelerometers, strain gauges, and displacement transducers (Linear Variable Differential

Transformers, LVDTs) [19]. This data is used in Structural Health Monitoring (SHM) to determine the natural frequencies, mode shapes, damping ratios, and modal participation factors of the bridge before and after damage occurs [20].

On the other hand, strain-based damage detection on bridges is based on strain distribution data along the bridge, which can be obtained through strain gauges or fiber optic sensors [21]. The data should include strain measurements at multiple locations on the bridge, both under undamaged and damaged scenarios [22].

Another alternative approach is the so-called visual-based damage detection [23]. The different methodologies implementing this approach, exploit high-resolution images or videos of the bridge surface and its components, which can be obtained through visual inspection, Unmanned Aerial Vehicles (UAVs), or other imaging technologies [24].

Finally, Mixed damage detection on bridges typically involves the combination of multiple types of data to enhance the accuracy and reliability of the damage detection results [25]. The required data can vary depending on the specific combination of techniques used but typically involves a combination of vibration-based, strain-based, and visual-based data [26].

In addition to the collection of the required data for each anomaly detection type, it is also necessary to collect accurate information on the bridge's geometry, material properties, and loading conditions. This can include details such as the bridge span, cross-sectional area, modulus of elasticity, and the types and weights of vehicles that typically cross the bridge. Furthermore, it is important to have data on environmental conditions, such as temperature, wind speed, and precipitation, as these factors can affect the dynamic response of the bridge. Finally, it is essential to have a baseline dataset of the measurements on the undamaged bridge, so that changes in data patterns can be identified and interpreted as damage. Overall, the more comprehensive the data set is and the lower the data uncertainties are, the more accurate and reliable the damage detection results will be.

3 PROPOSED FRAMEWORK AND DISCUSSION

The proposed novel framework for the creation of synthetic data that could be used to form benchmark study cases for prototyping and validation of DT components is shown schematically in Figure 2.

The data generated by a BHM system come in the form of time series, i.e., data collected over time at certain intervals, also known as sampling rates. Therefore, the first component of the proposed synthetic data generation framework is the sync main module. This element ensures the correct alignment in time of the different time series data generated, i.e., synthetic monitoring data, environmental data, and operational conditions data. To achieve the synchronization of the different synthetic time series generated within the proposed framework, a Dynamic Time Warping (DTW) function is suggested [27].

The synthetic data artificially generated by the proposed framework, namely, vibration, strain, and visual data, will provide information that can be used in place of real historic data. For such purposes, the generated data will have to be programmable. In other words, the statistical features of the data could be rebalanced, imputed, or have a stricter or looser adherence to the original distributions and correlations [28]. This will allow improving DT bridge model prototyping performance by enabling the simulation of what-if scenarios and the generation of test data with an improved ability to test and validate prototypes. Several Python libraries readily available have the required features to effectively generate programmable data such as PyOD [29], which is a specialized, comprehensive and scalable library for the detection/generation of outlying objects; and ctgan [30], which is a collection of high fidelity synthetic data generators based on deep learning algorithms. Both libraries' capabilities will be integrated and combined within the proposed synthetic data generation framework.

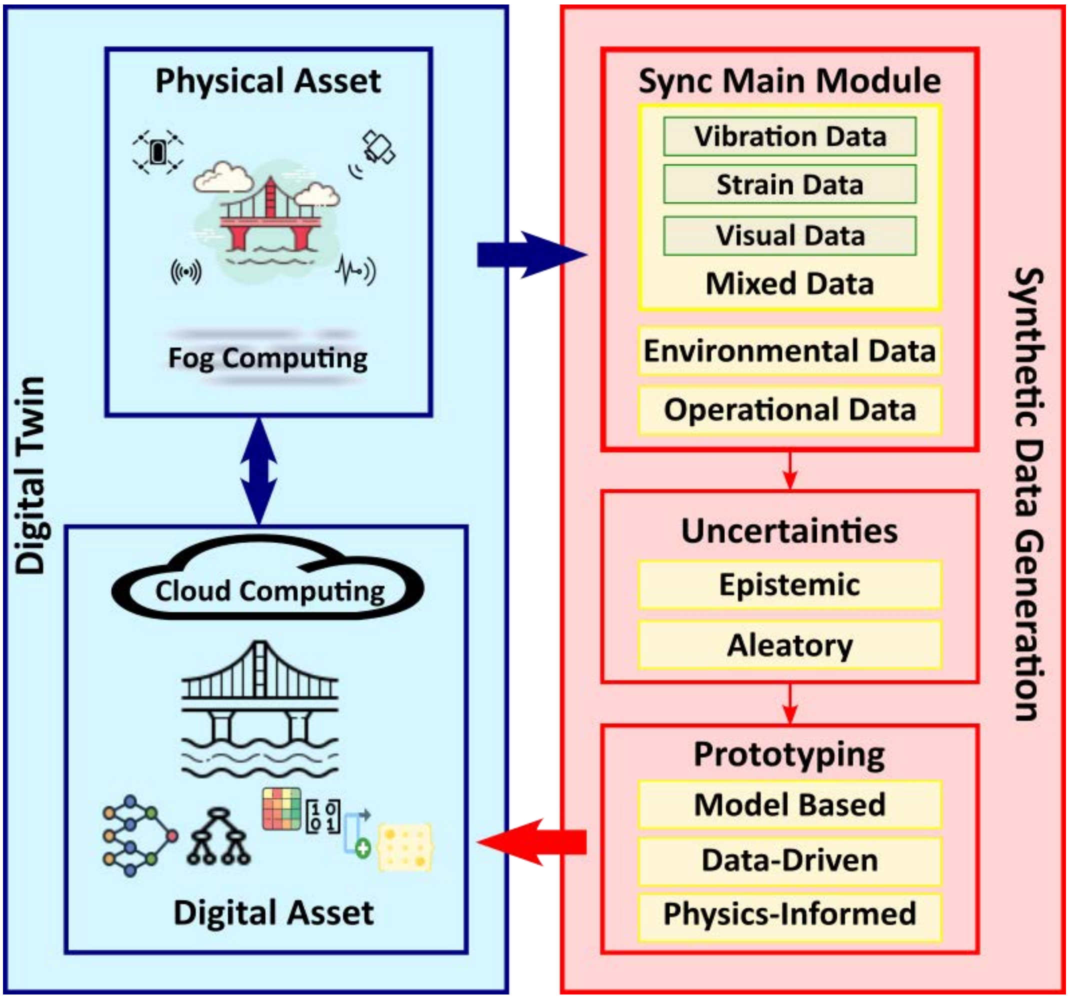

Figure 2. The proposed framework for synthetic data generation, which accounts for uncertainties and could be used in the prototyping of DTs.

The generation of vibration, strain, and visual data will be done within a mixed data component. This sub-module will be developed to generate benchmark databases of multi-metric data which could be used by mixed anomaly detection algorithms prototyping requirements.

The operational conditions data can be used to easily differentiate between damaged and undamaged scenarios, as anomaly conditions often reflect on the properties of the bridge structural components, such as supports, stiffness of members, extraordinary loads leading to permanent deformations, etc. Accounting for environmental data is also of paramount importance within the context of synthetic data generation, as it has been proved by several authors [31, 32] that environmental conditions can modify some properties of the bridge, which does not necessarily mean the asset has suffered some damage.

The framework considers both epistemic and aleatory uncertainties, which are key features to be accounted for in the validation process of any newly developed technology. The uncertainties component of the framework is presented and discussed in more detail in another study of the authors [33].

It is worth mentioning that the data generated by the proposed framework could be used for the prototyping and validation of either model-based or data-driven algorithms as well as by the more complex physics-informed ones. It is important to highlight that the proposed framework aims to produce realistic data capable of depicting general bridge management and operation scenarios, not to accurately replicate the observable data of a specific bridge study case. Thus, the data produced, should serve as a benchmark, so that novel prototypes can be tested and validated, consequently preparing them for their real-world deployment.

Although the creation of similar synthetic benchmark datasets has been proposed in the past [34], such alternatives are hard to find and even harder to access. Therefore, it is of paramount importance that the data generated comply with the FAIR principles having to do with the "Findability, Accessibility, Interoperability, and Reuse" of digital assets [35], i.e. the data needs to be findable, accessible, interoperable, and usable. To ensure these requirements are fully met, the proposed framework will follow the steps highlighted in the Three-point FAIRification Framework [36].

4 CONCLUSIONS AND FURTHER WORK

In this article, the shift towards a Digital Twin paradigm by all stakeholders in the engineering, architecture, and construction industry is recognized. As this new approach is currently in its early stages of development and adoption, new proposals to optimize its application need to be explored and developed. Such attempts require the development of prototypes that need to be validated against benchmark data.

Overall, synthetic data used for the creation of "what-if" scenarios for bridge digital twins can provide valuable insights into the bridge's performance under different conditions, allowing AEC professionals to identify potential issues and make informed decisions about maintenance and design challenges.

In this paper, a framework for the creation of a synthetic data generation tool has been proposed. Such framework produces high-quality FAIR data that allows novel developed prototypes to be validated and consequently be implemented in further stages of the Digital Twin creation for real infrastructure assets. The main characteristics of the proposed framework are the following:

- It accounts 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.
- A synchronizing module ensures that all data can be correctly tracked over time.
- It considers both epistemic and aleatory uncertainties for the adequate generation of real-world-like scenarios.
- The data generated is suitable for its use in the development and validation of model-based, data-driven, and physics-informed components of a digital asset.

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
AI-HBrIM	As-Is Historical Bridge Information Model
BHM	Bridge Health Monitoring
DT	Digital Twin
DTW	Dynamic Time Warping
FAIR	Findability, Accessibility, Interoperability, and Reuse
HBrIM	Historical Bridge Information Model
ICOMOS	International Council on Monuments and Sites
ISCARSAH	International Scientific Committee on the Analysis and Restoration of Structures of Architectural Heritage
SHM	Structural Health Monitoring
UAV	Unmanned Aerial Vehicle

REFERENCES

[1] 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.

[2] Shabani, A., et al., *Seismic Vulnerability Assessment and Strengthening of Heritage Timber Buildings: A Review.* Buildings, 2021. **11**(12) DOI: https://doi.org/10.3390/buildings11120661.

[3] 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.

[4] Asteris, P. and V. Plevris, eds. *Handbook of Research on Seismic Assessment and Rehabilitation of Historic Structures*. 2015, IGI Global. 1130. DOI: https://doi.org/10.4018/978-1-4666-8286-3.

[5] 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.

[6] Hosamo, H.H. and M.H. Hosamo, *Digital Twin Technology for Bridge Maintenance using 3D Laser Scanning: A Review.* Advances in Civil Engineering, 2022. **2022** DOI: https://doi.org/10.1155/2022/2194949.

[7] Lawal, O., et al. *Development and Validation of a Framework for Smart Wireless Strain and Acceleration Sensing*. Sensors, 2022. **22**, DOI: https://doi.org/10.3390/s22051998.

[8] Maeck, J. and G. De Roeck, *Description of Z24 Benchmark.* Mechanical Systems and Signal Processing, 2003. **17**(1): p. 127-131 DOI: https://doi.org/10.1006/mssp.2002.1548.

[9] Dilena, M. and A. Morassi, *Dynamic testing of a damaged bridge.* Mechanical Systems and Signal Processing, 2011. **25**(5): p. 1485-1507 DOI: https://doi.org/10.1016/j.ymssp.2010.12.017.

[10] Farrar, C.R., et al., *Dynamic characterization and damage detection in the I-40 bridge over the Rio Grande.* 1994 DOI: https://doi.org/10.2172/10158042.

[11] 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.

[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] 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.
[14] 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.
[15] 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.
[16] 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.
[17] 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.
[18] 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.
[19] Lu, Q., et al., *Developing a Digital Twin at Building and City Levels: Case Study of West Cambridge Campus.* Journal of Management in Engineering, 2020. **36**(3) DOI: https://doi.org/10.1061/(ASCE)ME.1943-5479.0000763.
[20] Neves, A.C., et al., *Structural health monitoring of bridges: a model-free ANN-based approach to damage detection.* Journal of Civil Structural Health Monitoring, 2017. **7**(5): p. 689-702 DOI: https://doi.org/10.1007/s13349-017-0252-5.
[21] Febrianto, E., et al., *Digital twinning of self-sensing structures using the statistical finite element method.* Data-Centric Engineering, 2022. **3**(3) DOI: https://doi.org/10.1017/dce.2022.28.
[22] Weinstein, J.C., M. Sanayei, and B.R. Brenner, *Bridge Damage Identification Using Artificial Neural Networks.* Journal of Bridge Engineering, 2018. **23**(11) DOI: https://doi.org/10.1061/(ASCE)BE.1943-5592.0001302.
[23] Shim, C.S., et al., *Development of a bridge maintenance system for prestressed concrete bridges using 3D digital twin model.* Structure and Infrastructure Engineering, 2019. **15**(10): p. 1319-1332 DOI: https://doi.org/10.1080/15732479.2019.1620789.
[24] Perry, B.J., et al., *Streamlined bridge inspection system utilizing unmanned aerial vehicles (UAVs) and machine learning.* Measurement: Journal of the International Measurement Confederation, 2020. **164** DOI: https://doi.org/10.1016/j.measurement.2020.108048.
[25] Bao, Y., et al., *Computer vision and deep learning–based data anomaly detection method for structural health monitoring.* Structural Health Monitoring, 2019. **18**(2): p. 401-421 DOI: https://doi.org/10.1177/1475921718757405.
[26] 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.

[27] Parisi, F., et al., *Automated location of steel truss bridge damage using machine learning and raw strain sensor data.* Automation in Construction, 2022. **138**: p. 104249 DOI: https://doi.org/10.1016/j.autcon.2022.104249.

[28] Mostly AI, *Synthetic Data Generator Documentation*. 2022.

[29] Zhao, Y., Z. Nasrullah, and Z. Li, *PyOD: A Python Toolbox for Scalable Outlier Detection.* Journal of Machine Learning Research, 2019. **20**(96): p. 1-7.

[30] Xu, L., et al. *Modeling Tabular data using Conditional GAN*. in *Advances in Neural Information Processing Systems*. 2019.

[31] Zhu, Y.C., W. Xiong, and X.D. Song, *Structural performance assessment considering both observed and latent environmental and operational conditions: A Gaussian process and probability principal component analysis method.* Structural Health Monitoring, 2022. **21**(6): p. 2531-2546 DOI: https://doi.org/10.1177/14759217211062099.

[32] Filizadeh, R., E. Hernandez, and D. Rosowsky, *Time-Dependent Damage Evolution of Reinforced Concrete Bridge Piers: Implications for Multihazard Analysis.* ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 2023. **9**(1) DOI: https://doi.org/10.1061/AJRUA6.RUENG-953.

[33] Jiménez Rios, A., V. Plevris, and M. Nogal, *Uncertainties in the Synthetic Data Generation for the Creation of Bridge Digital Twins*, in *5th International Conference on Uncertainty Quantification in Computational Science and Engineering (UNCECOMP 2023)*. 2023, ECCOMAS: Athens, Greece.

[34] 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).

[35] 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.

[36] Schultes, E., *Official GO FAIR Foundation icons for the Three-Point FAIRification Framework.* 2021 DOI: https://doi.org/10.5281/zenodo.4678333.