Roland Platz
Garrison Flynn
Kyle Neal
Scott Ouellette   *Editors*

Model Validation and Uncertainty Quantification, Volume 3

Proceedings of the 41st IMAC, A Conference and Exposition on Structural Dynamics 2023

SEM

Springer

Conference Proceedings of the Society for Experimental Mechanics Series

Series Editor

Kristin B. Zimmerman
Society for Experimental Mechanics, Inc.,
Bethel, CT, USA

The Conference Proceedings of the Society for Experimental Mechanics Series presents early findings and case studies from a wide range of fundamental and applied work across the broad range of fields that comprise Experimental Mechanics. Series volumes follow the principle tracks or focus topics featured in each of the Society's two annual conferences: IMAC, A Conference and Exposition on Structural Dynamics, and the Society's Annual Conference & Exposition and will address critical areas of interest to researchers and design engineers working in all areas of Structural Dynamics, Solid Mechanics and Materials Research.

Roland Platz • Garrison Flynn • Kyle Neal • Scott Ouellette
Editors

Model Validation and Uncertainty Quantification, Volume 3

Proceedings of the 41st IMAC, A Conference and Exposition on Structural Dynamics 2023

🐎 Springer

Editors

Roland Platz
Mechanical Engineering and Mechatronics
TH Deggendorf
Weißenburg i. Bay., Germany

Kyle Neal
Sandia National Laboratories
Albuquerque, NM, USA

Garrison Flynn
Los Alamos National Laboratory
Santa Fe, NM, USA

Scott Ouellette
Los Alamos National Laboratory
Santa Fe, NM, USA

Preface

Model Validation and Uncertainty Quantification represents one of ten volumes of technical papers presented at the 41st IMAC, A Conference and Exposition on Structural Dynamics, organized by the Society for Experimental Mechanics, held February 13–16, 2023. The full proceedings also include volumes on Nonlinear Structures and Systems; Dynamics of Civil Structures; Dynamic Substructures; Special Topics in Structural Dynamics and Experimental Techniques; Computer Vision and Laser Vibrometry; Dynamic Environments Testing; Sensors and Instrumentation and Aircraft/Aerospace Testing Techniques; Topics in Modal Analysis and Parameter Identification; and Data Science in Engineering.

Each collection presents early findings from experimental and computational investigations on an important area within Structural Dynamics. Model Validation and Uncertainty Quantification (MVUQ) is one of these areas.

Modeling and simulation are routinely implemented to predict the behavior of complex dynamical systems. These tools powerfully unite theoretical foundations, numerical models, and experimental data which include associated uncertainties and errors. The field of MVUQ research entails the development of methods and metrics to test model prediction accuracy and robustness while considering all relevant sources of uncertainties and errors through systematic comparisons against experimental observations.

The MVUQ Technical Division at the Society for Experimental Mechanics would like to thank the authors, presenters, session organizers, and session chairs for their participation and effort in this track.

Weißenburg, Bavaria, Germany   Roland Platz
Santa Fe, NM, USA   Garrison Flynn
Albuquerque, NM, USA   Kyle Neal
Santa Fe, NM, USA   Scott Ouellette

Contents

Chapter 1
Introducing a Round-Robin Challenge to Quantify Model Form Uncertainty in Passive and Active Vibration Isolation

Roland Platz, Xinyue Xu, and Sez Atamturktur

Abstract The aim is to quantify model form uncertainty in a passive and active vibration isolation system example during a round-robin challenge among IMAC's Model Validation and Uncertainty Quantification (MVUQ) technical division. In this context, passive means that the vibration isolation only depends on preset inertia, damping, and stiffness properties. Active means that additional controlled forces enhance the vibration isolation. The focus is on studying multiple mathematical models of the same one-mass oscillator system to predict its structural dynamic behavior against a consistent set of experimental data to ensure direct comparability. The models differ in their scope and complexity; the experimental data will be offered to different research groups that are yet to be constituted during this IMAC. The participants are welcome to join the research group and discuss their results in an exclusive IMAC session reserved for round-robin results in the following years.

Keywords Model form uncertainty · Vibration isolation · Round-robin challenge

1.1 Introduction

Model form uncertainty expresses the imperfection of mathematical models that aim to predict real-world phenomena. It results from unknown, incomplete, inadequate, or unreasonable functional relations between the model input and output, as well as between model parameters and state variables when compared to observations from the real experimental test [1]. An engineer's dilemma in early-stage design – before calibration, verification, and validation processes start – implicates the extent of uncertainty for different mathematical model options. The engineer is not sure if a simple model with minor efforts in modeling, but less precision may be still adequate with tolerable uncertainty when compared to a more complex model with major efforts in modeling, high precision, and less uncertainty. There is a considerable risk of under or overfitting the model to become too general or too specific, resulting in high or low adaptation capabilities of the model when changing the system, for example, by further development and advancement over time. Mathematical models are usually a combination of axiomatic and/or empiric functional relations, thus, adequate weighting between the laws of physics and phenomenological experiences is challenging and often leads to the remaining model form uncertainty. Functional relations may also describe linear or nonlinear as well as time-variant or time-invariant dynamic behavior.

The works [2, 3] introduce a general relation between an actual observation from experiments and a mathematical model. In structural dynamic systems, an observation reflects the measured physical outcome, mostly as states like forces, displacement, accelerations, etc. The model's prediction capability depends on data and their functional relations in the selected model form. Generally, stochastic models explicitly consider the uncertainty associated with the capacity of the selected model form to predict the observed system behavior. The difference between the outcome of the observation and the simulation model is also a culmination of noise and bias in the measurements; bias primarily stems from sensor calibration

R. Platz (✉)
Department of Mechanical Engineering and Mechatronics, Technology Center in Weißenburg, Deggendorf Institute of Technology DIT, Weißenburg i. Bay, Germany
e-mail: roland.platz@th-deg.de

X. Xu
Department of Architectural Engineering, The Pennsylvania State University, University Park, PA, USA

S. Atamturktur
Harry and Arlene Schell Professor and Department of Architectural Engineering, The Pennsylvania State University, University Park, PA, USA

errors. Differences that cannot be directly related to data variations, functional relations, and deviations from measurements will be represented by a remaining discrepancy function that is completely non-axiomatic and non-empiric. The round-robin challenge seeks to find a metric to adequately quantify model form uncertainty derived from multiple mathematical models of the same vibration isolation system and by consistently validating them against the same experimental data.

1.2   Experimental Test Environment

The investigated one-mass oscillator was introduced in [4] at IMAC in 2022. It is equipped with a velocity feedback controller that realizes passive and active damping. Considering this structural dynamic example, Platz et al. so far investigated the influence of data uncertainty on the vibrational behavior in the frequency domain by numerical simulations in [5–7]. Lenz et al. [8] conducted experimental investigations regarding the data uncertainty of the system as shown in [1, 9]. Data includes model parameters and state variables; the model determines the functional relation, [10].

The simplest analytical model form of the one-mass oscillator is based on pure axiomatic relations in a differential equation of motion, [1]. An analytical two-mass oscillator is the next and more complex alternative and will be modeled in the round-robin challenge. It is the basis for the experimental simulation with an additional frame mass to realize the base point excitation as shown in Fig. 1.1, right. Figure 1.2, left and center, shows the real test setup and its representation as a finite element (FE) model in Fig. 1.2, right, [11]. The FE model is considered one of another possible alternative and more complex models to the simplest analytical one- and two-mass oscillator model. The rigid frame model in Fig. 1.1, right, with mass m_f is fixed by idealized gliding support assumed to have no friction perpendicular to the z-direction. The support permits a frame movement only in z-direction. The frame is constrained by an idealized damper with the damping coefficient b_f and spring with the stiffness k_f in z-direction. The frame suspends from a rigid mount via elastic straps vertical to the z-direction, allowing low-frequency pendulum motion of the frame in z-direction (Fig. 1.2). This motion is the translational absolute excitation displacement $w(t)$ in z-direction when the frame is excited by a hammer impulse. The frame in Figs. 1.1 and 1.2 retains two supports that fix a leaf spring at its ends at A and C, with the effective bending length l on sides A–B and B–C, and with the rigid mass m in the center position at B in Fig. 1.1, right. The leaf spring is the practical realization of the spring elements in Fig. 1.1, left. Its cross-section area is $d \cdot h$, with the cross-section width d and height h; its stiffness k is a function of the bending stiffness EI. E is the elastic or Young's modulus of the leaf spring made from carbon fiber reinforced polymer (CFRP), I is the area moment of inertia. The two supports at A and C are adjustable along l to tune the
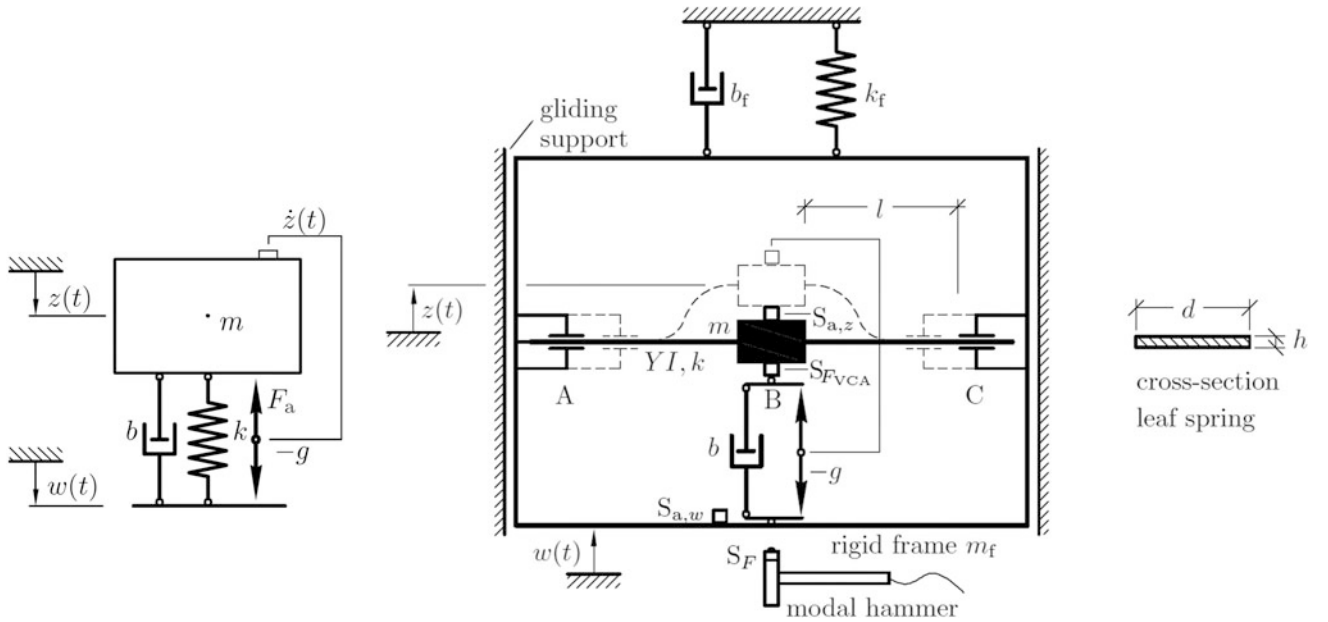
Fig. 1.1 *Left*: one-mass oscillator model with damping b, stiffness k, and mass m, position excitation $w(t)$ of a massless base point, and active vibration isolation by active velocity feedback control force F_a; *right*: schematic diagram of a real test set up with an additional rigid frame mass m_f excited by a modal hammer, as well as sensors $S_{a,z}$, $S_{a,w}$ $S_{F VCA}$, S_F, and $S_{F VCA}$ to measure the mass and frame acceleration a_z and a_w, as well as the electromotive and hammer force F_{VCA} and F, [4]
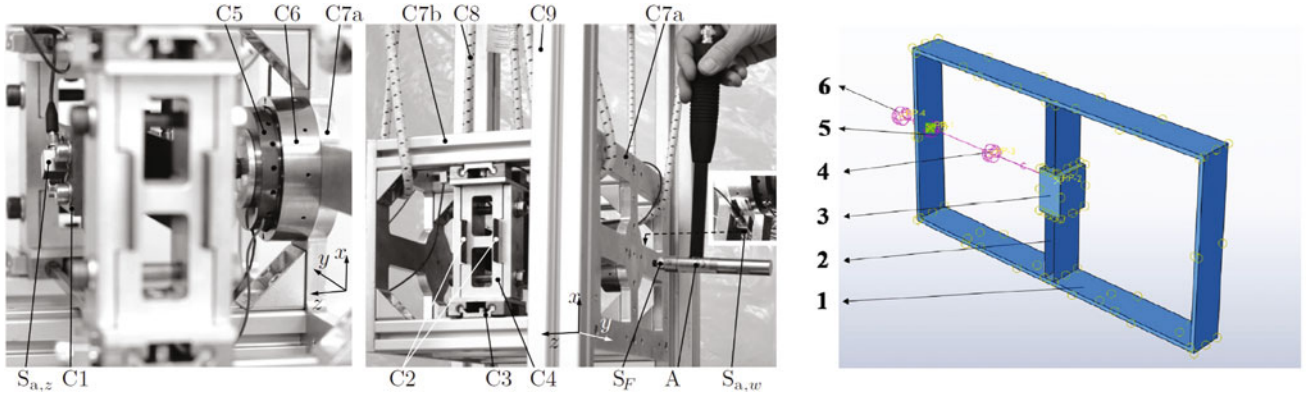
Fig. 1.2 Physical test setup; *left*: assembly of leaf spring and VCA; *center*: hammer impulse on frame; the components are: acceleration sensor $S_{a, z}$ attached to the oscillating mass C1, two leaf springs C2 with partial stiffness $k/2$ on each side of C1, glide support C3, fixed leaf spring support C4, VCA coil support/holder C5, VCA stator, magnet outer ring C6, front/side structure of rigid frame C7a/b with total mass m_f, elastic strap C8, mount C9 to suspend the frame with elastic straps, acceleration sensor $S_{a, w}$ (hidden behind the frame) on the frame mass m_f, and force sensor S_F measuring the impulse force from the model hammer A; *right*: representation as a FE-model with the rigid frame 1, leaf springs 2, mass oscillator 3, the damper 4, active force 5, and spring 6, [11]

leaf spring's bending deflection, and eventually its effective stiffness k. A voice coil actuator (VCA) provides the passive and active damping forces F_b and $F_a = -g z$ with the gain factor g; they combine into the electromotive force F_{VCA}, measured by the force sensor $S_{F\,VCA}$, [4].

1.3   Test and Analysis

The authors conducted the first deterministic and probabilistic analysis of the model from uncertainty. In the deterministic approach in [1], the relevant vibration isolation outcomes are the excitation hammer force, frame, and mass accelerations, as well as phase and amplitudes from experiments and models after calibration in the time domain. The outputs were measured by experiments and calculated from the simple analytic one-mass oscillator model. For example, the objective function

$$\min_{\theta \epsilon \mathbb{R}} \frac{1}{N} \sum_{n}^{N} \sum_{p}^{P} \frac{\left\{ y_p\left(X_n\right) - v_p\left(X_n, \theta\right) \right\}^2}{\max \left|y_p\left(X_n\right)\right|^2} \tag{1.1}$$

for deterministic model calibration or, respectively, model updating is a least squares minimization (LSM). It uses P observation outcomes $y_p(X_n)$ and P predicted model outcomes $v_p(X_n, \theta)$ with n control parameter $X_n = [t_n, f_n]$ as discrete time and frequency elements, and calibration parameters $\theta = [k, b, g]$. In [1], the outcomes are in the time and/or in the frequency domain, leading to $N = 4096$ time samples and/or $N = 2048$ frequency samples. For example, the control parameter X_n is frequency. The number of outcomes is $P = 2$, and only the amplitude and phase angle were considered. The LSM is conducted by particle swarm optimization (PSO). The investigation in [1] showed that with higher passive damping, the prediction of the dynamic outcome via the calibrated analytical model becomes less adequate. In the case of data, resp. measurement uncertainty, the deviation of calibrated damping is up to 9% for the highest applied damping and up to 30% for stiffness. The stiffness prediction becomes more adequate in cases of active damping with only up to 8% at the highest active damping. In case of model form uncertainty, the active damping cases lead to poor prediction quality of the calibrated passive damping coefficient and the active gain, up to 50%.

In the probabilistic approach in [11], the first goal was to find a posterior distribution of unknown or imperfectly known parameters by calibrating the one-mass oscillator model using the experimental observations. The second goal was to find settings for the controllable system inputs such that the resulting system output is optimized with respect to the design requirements. This Bayesian framework places a prior on the calibration parameters θ, pairing it with a Gaussian process (GP) metamodel of the computer model of interest and a GP prior on the model discrepancy, and using the available observations y_p of the real system to find the posterior distribution. It established the ability to quantify the uncertainty remaining in

the true value of the calibration parameter, the optimal settings for the design input, and the resulting model output for the vibration isolation example in [11].

1.4 Conclusion

This current state of investigation is the basis for successive research on evaluating the model form uncertainty for the vibration isolation example within the round-robin challenge. The existing test environment allows reproducible, consistent, and comparable validation of different models of always the very same oscillator system.

References

1. Platz, R.: Approach to assess basic deterministic data and model form uncertainty in passive and active vibration isolation. In: Pelz, P.F., Groche, P. (eds.) Uncertainty in Mechanical Engineering, pp. 208–223. Springer International Publishing, Cham (2021)
2. Higdon, D., Gattiker, J., Williams, B., Rightley, M.: Computer model calibration using high-dimensional output. J. Am. Stat. Assoc. **103**(482), 570–583 (2008). https://doi.org/10.1198/016214507000000888
3. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. J. R. Stat. Soc. Series B Stat. Methodology. **63**(3), 425–464 (2001). https://doi.org/10.1111/1467-9868.00294
4. Platz, R.: Comprehensive testing environment to evaluate approaches in uncertainty quantification for passive and active vibration isolation. In: Mao, Z. (ed.) Model Validation and Uncertainty Quantification Conference Proceedings of the Society for Experimental Mechanics Series, vol. 3. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-04090-0_11
5. Platz, R., Ondoua, S., Enss, G.C., Melz, T.: Approach to evaluate uncertainty in passive and active vibration reduction. In: IMAC–XXXII A Conference and Exposition on Structural Dynamics, Orlando, FL, USA, pp. 345–352, 3–6 Feb 2014
6. Platz, R., Enß, G.C.: Comparison of uncertainty in passive and active vibration isolation. In: IMAC–XXXIII A Conference and Exposition on Structural Dynamics, Orlando, FL, USA, pp. 15–25, 2–5 Feb 2015
7. Platz, R., Melzer, C.: Uncertainty quantification for decision making in early design phase for passive and active vibration isolation. In: ISMA 2016 including USD 2016 International Conference on Uncertainty in Structural Dynamics, Leuven, Belgium, pp. 4501–4513, 19–21 Sept 2016
8. Lenz, J., Platz, R.: Quantification and evaluation of parameter and model uncertainty for passive and active vibration isolation. In: IMAC–XXXVII A Conference and Exposition on Structural Dynamics, Orlando, FL, USA, pp. 135–147, 28–31 Jan 2019 (2020)
9. Perfetto, S., Rohlfing, J., Infante, F., Mayer, D., Herold, S.: Test rig with active damping control for the simultaneous evaluation of vibration control and energy harvesting via piezoelectric transducers. J. Phys. Conf. Ser. **744**, 012010 (2016)
10. Melzer, C.M., Platz, R., Melz, T.: Comparison of methodical approaches to describe and evaluate uncertainty in the load-bearing capacity of a truss structure. In: Fourth International Conference on Soft Computing Technology in Civil Engineering, Prague, Czech Republic, 1–4 Sept 2015
11. Ehrett, C., Brown, D.A., Kitchens, C., Xu, X., Platz, R., Atamturktur, S.: Simultaneous Bayesian calibration and engineering design with an application to a vibration isolation system. ASME. J. Verif. Valid. Uncert. **6**(1), 011007 (2021)

Chapter 2
An Uncertainty-Aware Measure of Model Calibration Flexibility

Xinyue Xu, Yang Yu, Roland Platz, and Sez Atamturktur

Abstract Physics-based models of structural dynamic systems are needed for various engineering applications, including structural controls and condition monitoring. These models often need to be calibrated against experimental measurements to mitigate uncertainties in poorly known model parameters and account for systemic model errors. In such a calibration campaign, under- or overfitting of a model to measured data may impede obtaining generalizable predictions. The underfitted calibration campaign fails to fully capture the underlying patterns, misses out on opportunities to learn from the measured data, and leads to an inferior predictive capability; while the overfitted calibration campaign may yield a satisfactory goodness-of-fit, it degrades generalizability, and in turn the usefulness of the calibrated model. There is a well-known trade-off between goodness-of-fit to measured data and generalizability in unmeasured settings. In this context, the generalizability of a model calibration campaign denotes the ability of the model to fit alternative datasets. For a given set of available experiments, determining the optimal flexibility of a model calibration campaign is necessary to achieve the maximum possible generalizability. This work presents a generally applicable metric to quantify the flexibility of model calibration that effectively takes these factors into account. We present the computational framework for the metric and demonstrate its application on a polynomial problem.

Keywords Predictive modeling · Model selection · Uncertainty quantification · Bayesian inference · Uncertainty awareness

2.1 Introduction

Physics-based computer models in engineering define a relationship between input parameters that control a system and the system response of interest. Control variables determine the domain of applicability in which experimentalists have full control. The model also includes poorly known input parameters that define the system's characteristics and must be inferred against experimental measurements, which are regarded as calibration parameters. The simulation models predict unknown output responses within a predefined domain of applicability based on the values of control variables and calibration parameters [1].

In this context, model calibration has been widely employed to provide the best-estimated solutions for input parameters given the observed data in structural engineering fields [2–4]. However, the design of model calibration heavily depends on the availability of experimental data, the quality of experimental data, and the modeler's knowledge of the underlying system [5–9]. Given a set of possible model calibration scenarios that can be applied to represent the underlying system, the choice of the most suitable model calibration scenarios plays a pivotal role. Attempting to select the most suitable model calibration scenarios based on the best-estimated solutions may lead to an inferior predictive model. Because there is a well-known trade-off between model fitting and model generalizability.

This work introduces a quantitative measure, known as Indicator for Calibration Flexibility (*ICaF*), which evaluates the calibration flexibility of different models for a particular set of measured data and calibration campaign. *ICaF* assesses both

X. Xu (✉) · Y. Yu · S. Atamturktur
Department of Architectural Engineering, The Pennsylvania State University, University Park, PA, USA
e-mail: xkx5055@psu.edu

R. Platz
Department of Mechanical Engineering and Mechatronics, Technology Transfer Center in Weißenburg, Deggendorf Institute of Technology DIT, Weißenburg in Bayern, Germany
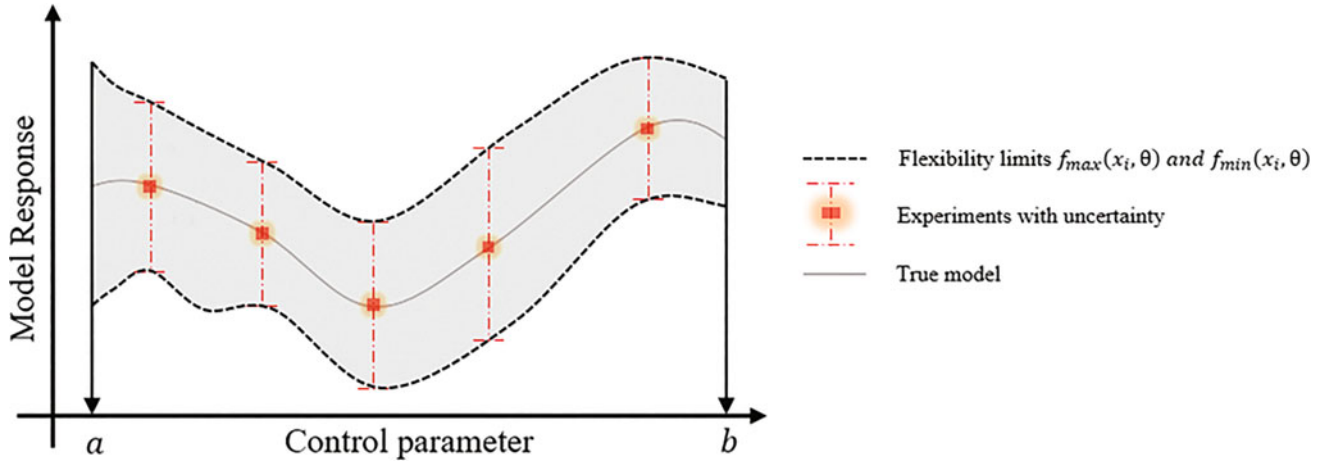
Fig. 2.1 A paradigm of the second term S (i.e., the overall predictive uncertainty). The dashed line represents predictive flexibility defined by posterior predictive distributions at discretized points

the goodness-of-fit and generalizability of the model through Bayesian inference. The purpose of model selection using *ICaF* is to identify the model that provides a good fit and excellent predictive performance simultaneously.

2.2 Methodology

Here we define the term S to quantify the generalizability of a model calibration campaign in the Bayesian model calibration framework. To illustrate this concept, imagine an elastic band that is anchored at multiple points along its length (corresponding to the domain of applicability). Our algorithm calculates how far this virtual band can be stretched if it is pulled upwards (towards the highest extreme values) or downwards (towards the lowest extreme values) at various points using Markov Chain Monte Carlo [10], which is expressed as shown in Fig. 2.1.

2.3 Case Study I: Polynomial Regressions

Here, we assume noninformative priors for the regression coefficients, and that the analyst is uncertain about the exact order of the true polynomial model. Under this context, suppose there are nine candidate models indexed by M_k, for $K = 1, 2, \ldots,$ 9, and the polynomial order increases in ascending order.

The true model that generates synthetic data is:

$$y = -x^3 + 2x^2 - 10x - 10 \tag{2.1}$$

After conducting model calibration to the model calibration campaign, we calculate the predictive flexibility S and the goodness-of-fit from M_1 to M_9 (Fig. 2.2).

The overall normalized *ICaF* is shown in Fig. 2.3. We can see that M_4 is observed to be the most appropriate model since it has the minimal *ICaF* value in the calibration campaign. Moreover, M_4 has the same model form as the true model that generated the synthetic data. In this case, *ICaF* shows its effectiveness in model selection considering the trade-off between the goodness-of-fit and model generalizability.
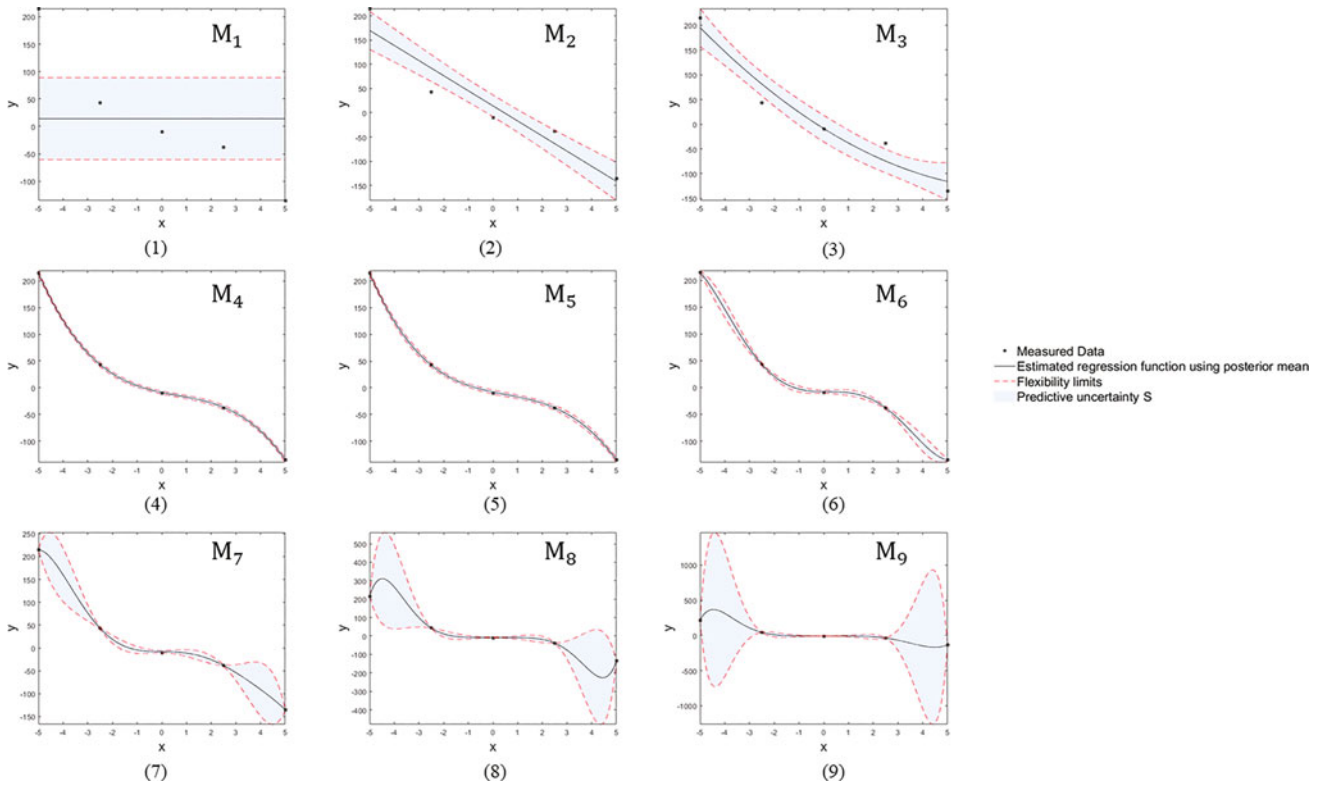
Fig. 2.2 The predictive flexibility S performance in a calibration campaign of regression models
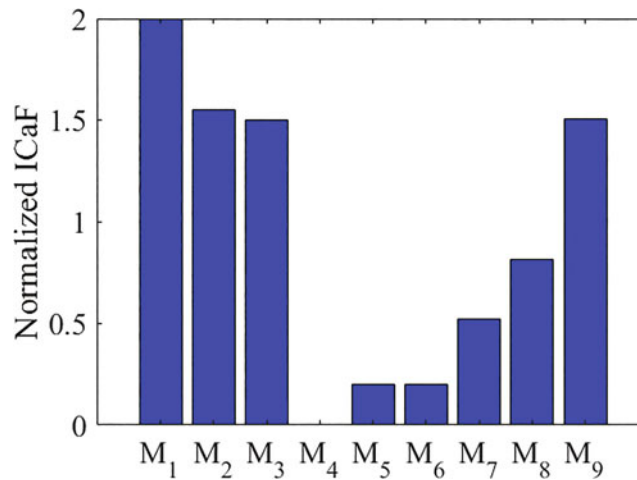
Fig. 2.3 *ICaF* performance on competing models with an increasing polynomial order

2.4   Conclusion

In this work, we proposed an uncertainty-aware metric for model selection that addresses the trade-off between goodness-of-fit and model generalizability given a set of alternative models and noisy data. We applied this novel metric to determine the most appropriate candidate model in a given calibration campaign for a regression problem.

References

1. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B Stat. Methodol. **63**(3), 425–464 (2001). https://doi.org/10.1111/1467-9868.00294
2. Higdon, D., Gattiker, J., Williams, B., Rightley, M.: Computer model calibration using high-dimensional output. J. Am. Stat. Assoc. **103**(482), 570–583 (Jun. 2008). https://doi.org/10.1198/016214507000000888
3. Farajpour, I., Atamturktur, S.: Error and uncertainty analysis of inexact and imprecise computer models. J. Comput. Civ. Eng. **27**(4), 407–418 (2013). https://doi.org/10.1061/(asce)cp.1943-5487.0000233
4. Bayarri, M.J., et al.: A framework for validation of computer models. Technometrics. **49**(2), 138–154 (2007). https://doi.org/10.1198/004017007000000092
5. Ehrett, C., Brown, D.A., Kitchens, C., Xu, X., Platz, R., Atamturktur, S.: Simultaneous Bayesian calibration and engineering design with an application to a vibration isolation system. J. Verif. Valid. Uncert. Quantif. **6**(1) (2021). https://doi.org/10.1115/1.4050075
6. Brown, D.A., Atamturktur, S.: Nonparametric functional calibration of computer models. Stat. Sin. **28**(2), 721–742 (2018). https://doi.org/10.48550/arXiv.1602.06202
7. Veksler, V.D., Myers, C.W., Gluck, K.A.: Model flexibility analysis. Psychol. Rev. **122**(4), 755–769 (2015). https://doi.org/10.1037/a0039657
8. Bard, Y.: Nonlinear Parameter Estimation. Academic Press (1974)
9. Letham, B., Karrer, B., Ottoni, G., Bakshy, E.: Constrained Bayesian optimization with noisy experiments. Bayesian Anal. **14**(2), 495–519 (2019)
10. Roberts, G.O., Smith, A.F.: Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. Stoch. Process. Appl. **49**(2), 207–216 (1994)

Chapter 3
Quantifying Model Form Uncertainty in Spring-Mass-Damper Systems

Rileigh Bandy and Rebecca Morrison

Abstract Models built from coupled ordinary differential equations are common in mechanics, chemical kinetics, electrodynamics, and many other fields. A canonical example, in both theory and experiments, is a system of linked spring-mass-dampers. Modeling all interactions between these objects often becomes intractable either due to computational expense or incomplete knowledge of the system. Common reduced models may involve only interactions between a small subset of the spring-mass-dampers. But these simplifications can lead to high model error, rendering the model useless for prediction. In this work, we explore decreasing model error through interpretable model correction: an inadequacy operator augments the reduced model to form an enriched model. We calibrate the enriched model with hierarchical Bayesian inference and validate it with posterior predictive assessments. Physical theory informs the inadequacy operator, which contains terms to capture the effect of the omitted objects on the reduced model. Several analytical and numerical examples are given. Results show that most of the model error can be recovered with a simple time-varying inadequacy operator.

Keywords Model error · Model discrepancy · Hierarchical Bayesian calibration · Model validation · Mechanical oscillations

3.1 Introduction

Examples of interacting dynamical systems abound in nearly all areas of science and engineering. In chemical kinetics, simple reactions involve hundreds of intermediate compounds [1, 2]; in ecology, the population of one species depends on many other species and environmental variables in the ecosystem [3]; and in mechanics, the motion of a particle is influenced by neighboring particles and other forces in the system [4, 5]. In all of these examples, detailed models that track every object and their interactions exist, but they frequently become intractable to solve due to computational expense and/or incomplete knowledge of every interaction in the system. For example, a spring system with hundreds of masses linked together can approximate the dynamics of a string [6], but it would require solving hundreds of coupled ordinary differential equations (ODEs), which quickly gets expensive when the dynamics are modeled for longer than a few seconds [7]. In this case, a reduced model involving a subset of the masses from the detailed model can be formed, but model error arises from unmodeled forces on the remaining masses. These inaccuracies lead to discrepancies between the reduced model and the detailed system.

When the model error is too large, model corrections can be constructed to decrease the discrepancy. A common approach is to add a nonintrusive, stochastic term to the model output [8, 9]. The stochastic term can interpolate within the calibration regime but cannot extrapolate beyond it. Another approach is to embed an inadequacy operator into the reduced model to form an enriched model [1, 10, 11]. The inadequacy operator is theory-informed and specialized to the particular modeling scenario, which allows for extrapolative predictions. In the string example, the theory of mechanics and knowledge about the source of the model error can be leveraged for proper treatment of the uncertainty.

In previous work, an embedded linear inadequacy operator captured the discrepancy caused by modeling a large system of interacting species with a system of only one or two variables of interest [12]. In this study, we will extend the embedded approach to systems with more challenging transient behavior—systems that oscillate several times before reaching stable equilibrium.

R. Bandy (✉) · R. Morrison
Department of Computer Science, The University of Colorado, Boulder, CO, USA
e-mail: rileigh.bandy@colorado.edu; rebeccam@colorado.edu

3.2 Spring-Mass-Damper Models

A chain of springs, illustrated by the Fermi–Pasta–Ulam–Tsingou problem [13], is a widely studied system that exhibits mechanical oscillations. Newton's Second Law of Motion gives the reliable theory $F = \text{diag}(m)a$, where F is a vector of the forces acting on each mass in the system, $\text{diag}(m)$ is a matrix with the masses on the diagonal and zeros on the off-diagonals, and a is a vector of their accelerations. In the standard spring-mass-damper system, the only forces acting on the system are spring forces $f_s(t)$ and damping forces $f_d(t)$, while all other forces are assumed negligible. This yields

$$\text{diag}(m)a = f_d(t) + f_s(t), \tag{3.1}$$

where the forces on each object are

$$
\begin{aligned}
f_{d,i}(t) &= -c_i v_i(t) \\
f_{s,i}(t) &= [-k_i(x_i(t) - x_{i-1}(t)) + k_{i+1}(x_{i+1}(t) - x_i(t))] (1 + \alpha_i (x_{i+1}(t) - x_{i-1}(t))).
\end{aligned}
\tag{3.2}
$$

A mass's displacement from its equilibrium position at time t is denoted by $x_i(t)$, $v_i(t)$ is a mass's instantaneous velocity, c_i is the damping coefficient, k_i is the spring constant, and α_i is the strength of the nonlinear interactions on mass i.

Detailed Models Given the above, a complete model links together N masses, where $N \geq 2$. The coupled ODEs of the N positions and velocities are

$$
\begin{aligned}
\frac{dx_i(t)}{dt} &= v_i(t) \\
\frac{dv_i(t)}{dt} &= \frac{1}{m_i} \{ f_{d,i}(t) + f_{s,i}(t) \} \\
&= \frac{1}{m_i} \{ -c_i v_i(t) + [-k_i(x_i(t) - x_{i-1}(t)) + k_{i+1}(x_{i+1}(t) - x_i(t))] (1 + \alpha_i(x_{i+1}(t) - x_{i-1}(t))) \},
\end{aligned}
\tag{3.3}
$$

where $i = 1, \ldots, N$, $x_0(t) = 0$, $x_{N+1}(t) = 0$ and $k_{N+1} = 0$. In this study, we constrain $x_i(t = 0) > 0$, $v_i(t = 0) = 0$, $0 \leq \alpha_i < 1$ and $k_i \geq m_i > c_i \ \forall i = (1, \ldots, N)$ to guarantee that the system be underdamped and oscillatory for some time before reaching its stable equilibrium.

Reduced Models There are numerous techniques for building reduced models such as projection-based model reduction [14, 15] and model reduction for slow-fast stochastic systems [16, 17]. How to form the best reduced model is application specific and an active research area. For this study, we assume the reduced model is a subsystem of the first M masses from the detailed model, where $1 \leq M < N$. A graphical representation of the spring-mass-damper system is shown in Fig. 3.1. All of the gray and white blocks are represented in the detailed model, while only the gray blocks are modeled by the reduced model. The error in the reduced model is caused by the omission of the white blocks' forces affecting the gray blocks.

The ODEs of the M positions and $M - 1$ velocities for the reduced model are the same as in Eq. (3.3). However, the Mth mass no longer depends on the $M + 1$th mass, which results in the Mth velocity of

$$\frac{dv_M(t)}{dt} = \frac{1}{m_M} \{ -c_M v_M(t) + [-k_M(x_M(t) - x_{M-1}(t))] (1 + \alpha_M(-x_{M-1}(t))) \}, \tag{3.4}$$
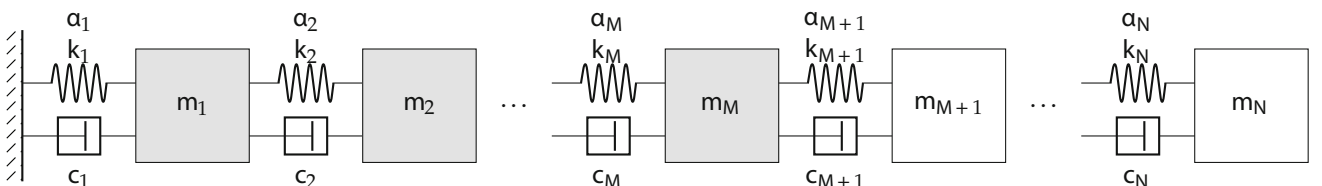
Fig. 3.1 Spring-mass-damper diagram for a system of N masses. The gray blocks represent the masses modeled by the reduced model, and the white blocks represent the additional masses modeled in the detailed model

where $x_0 = 0$. This reduction will propagate through to the first $M-1$ objects and alter their modeled positions and velocities. Therefore, we seek to augment the reduced model with an inadequacy operator to capture some of the important dynamics affecting the first M masses.

3.3 Enriched Model

If the detailed model involves N masses, but the reduced model only includes the first M masses, the error in the reduced model is caused by the omission of

$$\frac{1}{m_M} \left\{ k_{M+1}(x_{M+1}(t) - x_M(t)) + \alpha_M x_{M+1}(t)(-k_M(x_M(t) - x_{M-1}(t))) + \right.$$

$$\left. \alpha_M(x_{M+1}(t) - x_{M-1}(t))(k_{M+1}(x_{M+1}(t) - x_M(t))) \right\} \tag{3.5}$$

in $\frac{dv_M(t)}{dt}$. Inspection of Eq. (3.5) shows that the reduced model's error is more specifically caused by removing mass $M + 1$ and its corresponding spring constant k_{M+1}. We propose approximating the movement of mass $M + 1$ with a simple, damped wave that is dependent on time but independent from the rest of the system. Therefore, the enriched model adds a single time-dependent term to the derivative of the Mth mass's velocity, while $2(N - M)$ ODEs are omitted from the detailed model.

In the enriched model, the ODEs of the M positions and $M - 1$ velocities are the same as in the reduced model, and the Mth velocity is augmented with an inadequacy operator

$$\frac{dv_M(t)}{dt}$$

$$= \frac{1}{m_M} \left\{ -c_M v_M(t) + [-k_M(x_M(t) - x_{M-1}(t))] \left(1 + \alpha_M(-x_{M-1}(t))\right) \right\}$$

$$+ \frac{1}{m_M} \left\{ \delta\left(x_{M+1}^*(t) - x_M(t)\right) + \alpha_M x_{M+1}^*(t)(-k_M(x_M(t) - x_{M-1}(t))) + \alpha_M(x_{M+1}^*(t) - x_{M-1}(t))(\delta(x_{M+1}^*(t) - x_M(t))) \right\}$$

$$= \frac{1}{m_M} \left\{ -c_M v_M(t) + \left[-k_M(x_M(t) - x_{M-1}(t)) + \delta(x_{M+1}^*(t) - x_M(t)) \right] \left(1 + \alpha_M(x_{M+1}^* - x_{M-1}(t))\right) \right\}, \tag{3.6}$$

where $x_{M+1}^*(t) = a \exp(-\beta t) \cos(\omega t)$ defines a simple oscillator that approximates the position of mass $M + 1$. While $x_{M+1}^*(t)$ is only time-dependent, when its parameters $\boldsymbol{\theta} = (\delta, a, \beta, \omega)$ are calibrated, they are informed by observations from the detailed system making $x_{M+1}^*(t)$ indirectly dependent on the forces in the system. Furthermore, $\delta \approx k_{M+1}$, $a \approx x_{M+1}(t = 0)$, β captures the damping effect on $x_{M+1}^*(t)$, and ω captures the frequency of $x_{M+1}^*(t)$. Additional knowledge from the system constrains the inadequacy operator. Since the system is underdamped, we know $k_{M+1} > 0$ and thus $\delta > 0$. The initial position of mass $M + 1$ is always positive, so its initial amplitude is always positive, which makes $a > 0$. To guarantee that the system oscillates several times before dissipating to the stable equilibrium, $0 < \beta < \delta$. Finally, $0 \leq \omega \leq 2\pi$ captures the angle in radians of the simple oscillator.

3.4 Observations and Predictions

In numerical experiments, we simulate the detailed spring-mass-damper model for a given period of time and take discrete observations of the positions of the first M masses. Noisy observations from the detailed models are generated as

$$\boldsymbol{D}(X) = \boldsymbol{x}(t; X) + \boldsymbol{\epsilon}, \forall t = (0, tstep, \dots, T), \tag{3.7}$$

where $\boldsymbol{x}(t; X)$ is a vector of M positions at time t given the initial conditions X, T is the final time of the trajectory, $tstep$ is the timestep between discrete observations, and $\boldsymbol{\epsilon}$ is an M-vector of the measurement error added to the positions. The distribution of the measurement error on each position is

$$p_\epsilon = \mathcal{N}\left(0, \sigma_\epsilon^2\right). \tag{3.8}$$

Therefore, $\boldsymbol{D}(\boldsymbol{X})$ is a vector of length $M(\frac{T}{tstep} + 1)$. We denote a single observation from $\boldsymbol{D}(\boldsymbol{X})$ as $\boldsymbol{D}(\boldsymbol{X}; t, m)$, where $t \in (0, tstep, \ldots, T)$ is the instantaneous time and $m \in (1, \ldots, M)$ is a specific mass. We generate calibration data with one set of initial conditions, defined as $\boldsymbol{D}_c = \boldsymbol{D}(\boldsymbol{X}_c)$, and a different set of initial conditions to generate validation data, defined as $\boldsymbol{D}_v = \boldsymbol{D}(\boldsymbol{X}_v)$. We utilize black-box observations from the detailed model as a feasibility study, but experimental observations could be used because the calibration of the enriched model is agnostic to how observations are produced.

Predictions are generated by simulating the enriched model for a given period of time and sampling the positions of the M masses at discrete intervals in time. This results in the predicted outputs

$$\boldsymbol{Y}(\boldsymbol{\theta}, \boldsymbol{X}) = \boldsymbol{x}(t; \boldsymbol{\theta}, \boldsymbol{X}), \forall t = (0, tstep, \ldots, T), \tag{3.9}$$

where $\boldsymbol{\theta} = (\delta, a, \beta, \omega)$ are the inadequacy operator parameters. We denote a single prediction from $\boldsymbol{Y}(\boldsymbol{\theta}, \boldsymbol{X})$ as $\boldsymbol{Y}(\boldsymbol{\theta}, \boldsymbol{X}; t, m)$. Predictions for the calibration scenario are defined as $\boldsymbol{Y}_c(\boldsymbol{\theta}) = \boldsymbol{Y}(\boldsymbol{\theta}, \boldsymbol{X}_c)$, and predictions for the validation scenario are defined as $\boldsymbol{Y}_v(\boldsymbol{\theta}) = \boldsymbol{Y}(\boldsymbol{\theta}, \boldsymbol{X}_v)$.

3.5 Hierarchical Bayesian Calibration

Hierarchical Bayesian calibration samples the posterior distribution of the model parameters and hyperparameters given a set of initial conditions and noisy observations of the detailed model. We perform the calibration using the No-U-Turn (NUTS) sampler [18] implemented in the Turing software package [19]. Following [20], the joint distribution is decomposed into a probabilistically valid series of conditional models that reflects the causal mechanisms:

$$p(\boldsymbol{D}_c, \boldsymbol{\theta}, \boldsymbol{\phi}) = p(\boldsymbol{D}_c \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta} \mid \boldsymbol{\phi}) p(\boldsymbol{\phi}), \tag{3.10}$$

where $\boldsymbol{D}_c$ is the calibration data from the noisy observations of the detailed model, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4) = (\delta, a, \beta, \omega)$ are the model parameters of the enriched model, and $\boldsymbol{\phi} = (\mu_{\theta_1}, \sigma_{\theta_1}, \mu_{\theta_2}, \sigma_{\theta_2}, \mu_{\theta_3}, \sigma_{\theta_3}, \mu_{\theta_4}, \sigma_{\theta_4}) = (\mu_\delta, \sigma_\delta, \mu_a, \sigma_a, \mu_\beta, \sigma_\beta, \mu_\omega, \sigma_\omega)$ are the hyperparameters. Here, the hyperparameters are random variables with their own prior distributions, which accounts for uncertainty in our knowledge about the true form of the discrepancy.

The archetypal hierarchical model from [20] can be rewritten using Bayes rule to define the posterior of the model parameters and hyperparameters:

$$\pi_{post} = p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \boldsymbol{D}_c) \propto p(\boldsymbol{D}_c \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta} \mid \boldsymbol{\phi}) p(\boldsymbol{\phi}). \tag{3.11}$$

Prior The prior distribution is

$$\pi_{pri} = p(\boldsymbol{\theta} \mid \boldsymbol{\phi}) p(\boldsymbol{\phi}) = \prod_{i=1}^{4} p(\theta_i \mid \boldsymbol{\phi}_{\theta_i}) p(\boldsymbol{\phi}_{\theta_i}), \tag{3.12}$$

where

$$\begin{aligned}
p(\theta_i \mid \boldsymbol{\phi}_{\theta_i}) &= \mathcal{N}(\mu_{\theta_i}, \sigma_{\theta_i}^2) \\
p(\boldsymbol{\phi}_{\theta_i}) &= p(\mu_{\theta_i}) p(\sigma_{\theta_i}) \\
p(\sigma_{\theta_i}) &= \mathrm{Exp}(100) \\
p(\mu_\delta) = p(\mu_a) &= \mathcal{U}(0, 1000) \\
p(\mu_\beta) &= \mathcal{U}(0, 10) \\
p(\mu_\omega) &= \mathcal{U}(0, 2\pi).
\end{aligned} \tag{3.13}$$

The hyperparameter priors are maximum entropy priors. The expected value of the variances is 0.01. We know a priori that μ_δ and μ_a should be nonnegative, μ_β should also be nonnegative with a smaller magnitude to allow the system to oscillate several times before dissipating to the stable equilibrium, and μ_ω should be between zero and 2π radians to capture the frequency of mass $M + 1$.

Likelihood The likelihood distribution is

$$\pi_{like} = p(\boldsymbol{D}_c \mid \boldsymbol{\theta}, \boldsymbol{\phi}) = p(\boldsymbol{D}_c \mid \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{D}_c - \boldsymbol{Y}_c(\boldsymbol{\theta}) \mid \boldsymbol{0}, \Sigma_\epsilon), \tag{3.14}$$

which is defined by the measurement error on each observation in Eq. (3.8). The covariance matrix of the measurement error is $\Sigma_\epsilon = \mathrm{diag}(\sigma_\epsilon^2)$.

3.6   Validation

Before using the enriched model in place of the detailed system, the consistency between predictions from the enriched model and observations must be assessed. The posterior predictive assessment [21] calculates the probability that an observation $\boldsymbol{D}(X; t, m)$ is an outcome of the enriched model given the calibration data $\boldsymbol{D}_c$:

$$p(\boldsymbol{D}(X; t, m) \mid \boldsymbol{D}_c) = \int_{\boldsymbol{\theta}\boldsymbol{\phi}} \left(p(\boldsymbol{D}(X; t, m) \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(Y(\boldsymbol{\theta}, X; t, m) \mid \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \boldsymbol{D}_c) \right) d\boldsymbol{\theta}\boldsymbol{\phi}, \tag{3.15}$$

where $p(\boldsymbol{D}(X; t, m) \mid \boldsymbol{\theta}, \boldsymbol{\phi})$ is the likelihood that the observation is an outcome of the enriched model, which is defined by the measurement error in Eq. (3.14) as $\mathcal{N}\left(\boldsymbol{D}(X; t, m) - Y(\boldsymbol{\theta}, X; t, m) \mid 0, \sigma_\epsilon^2\right)$. Then, the probability from Eq. (3.15) is compared to the possible model predictions. In particular, we are interested in how much of the distribution corresponds to model predictions less likely than $\boldsymbol{D}(X; t, m)$, which is given by the γ-value:

$$\gamma = 1 - \int_S p(y \mid \boldsymbol{D}_c) dy, \tag{3.16}$$

where $S = \{y \in \mathbb{R} : p(y \mid \boldsymbol{D}_c) \geq p(\boldsymbol{D}(X; t, f) \mid \boldsymbol{D}_c)\}$ is the β-highest probability density credibility region [1]. A high γ-value indicates consistency between the observation and the enriched model, while a low γ-value indicates a potential discrepancy. A graphical depiction is shown in Fig. 3.2. Next, the γ-value is compared to a tolerance threshold $\tau = \alpha / \left(M \left(\frac{T}{tstep} + 1 \right) \right)$, where $\alpha = 0.05$ corresponds to a 95% confidence interval, and $M \left(\frac{T}{tstep} + 1 \right)$ is the Bonferroni
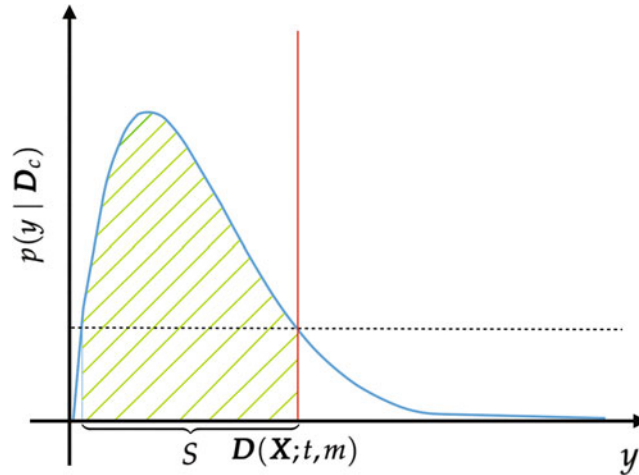
Fig. 3.2 The γ-value is the probability of the model (blue curve) predicting the observation (red line) or an event less likely, where the integral over S (green hashed region) contains the events more likely to be predicted by the model than the observation

correction. We iterate over all of the calibration observations $D(X; t, m) \in D_c$ and all of the validation observations $D(X; t, m) \in D_v$ and return the fraction of γ-values less than τ as $G(\gamma, D_c)$ for the calibration scenario and $G(\gamma, D_v)$ for the validation scenario.

3.7 Results

We investigate scenarios with varied (N, M) for both linear and nonlinear detailed models.

Linear Analytical Results When $\alpha_i = 0 \ \forall i \in (1, \ldots, N)$, the detailed model is linear, which yields an analytical solution. There are N complex-conjugate pairs of eigenvalues and eigenvectors because the constraint $k_i \geq m_i > c_i \ \forall i \in (1, \ldots, N)$ always results in an underdamped system. For the case $N = 2$, the linear detailed system is $\frac{du(t)}{dt} = Au(t)$, where $u(t) = (x_1(t), x_2(t), v_1(t), v_2(t))^T$. Let the parameters and initial conditions be defined as

$$
\begin{aligned}
m &= (5, 10)^T \\
c &= (1.5, 1.5)^T \\
k &= (40, 80)^T \\
u(t = 0) &= (5, 10, 0, 0)^T.
\end{aligned}
\tag{3.17}
$$

The real-valued coefficient matrix is

$$
A = \begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
-\frac{k_1 - k_2}{m_1} & \frac{k_2}{m_1} & -\frac{c_1}{m_1} & 0 \\
\frac{k_2}{m_2} & -\frac{k_2}{m_2} & 0 & -\frac{c_2}{m_2}
\end{bmatrix}.
\tag{3.18}
$$

The eigenvalues are

$$
\begin{aligned}
\lambda_1, \lambda_2 &\approx -0.1341 \mp 5.4621i \\
\lambda_3, \lambda_4 &\approx -0.0909 \mp 1.4614i,
\end{aligned}
\tag{3.19}
$$

the eigenvectors are

$$
\xi_1, \xi_2 \approx \begin{pmatrix} -0.0042 \pm 0.1691i \\ 0.0033 \mp 0.0618i \\ 0.9237 \\ -0.3381 \mp 0.0100i \end{pmatrix} \quad
\xi_3, \xi_4 \approx \begin{pmatrix} -0.0233 \pm 0.3324i \\ -0.0282 \pm 0.441i \\ 0.4879 \pm 0.0039i \\ 0.6662 \end{pmatrix},
\tag{3.20}
$$

and the corresponding general solution is

$$
u(t) = z_1 \xi_1 e^{\lambda_1 t} + z_2 \xi_2 e^{\lambda_2 t} + z_3 \xi_3 e^{\lambda_3 t} + z_4 \xi_4 e^{\lambda_4 t},
\tag{3.21}
$$

where $z_1, z_2 \approx 0.0000 \pm 5.4154i$ and $z_3, z_4 \approx -0.0813 \mp 10.2679i$.

This analytical solution indicated that all of the eigenvalues have roughly the same magnitude and none dominate control of the dynamics of the system. Therefore, a reduction that captures the slow dynamics is not trivial.

The analytical solution to the corresponding enriched model with $M = 1$ can be solved using variation of parameters. Let $\frac{du(t)}{dt} = Au(t) + g(t)$, where $Au(t)$ is the linear homogeneous part of the enriched model with

$$
A = \begin{bmatrix}
0 & 1 \\
-\frac{k_1 - \delta}{m_1} & -\frac{c_1}{m_1}
\end{bmatrix},
\tag{3.22}
$$

and $\boldsymbol{u}(t) = (x_1(t), v_1(t))^T$. The nonhomogeneous part of the enriched model is

$$\boldsymbol{g}(t) = \begin{pmatrix} 0 \\ \frac{1}{m_1} \delta a \exp(-\beta t) \cos(\omega t) \end{pmatrix}, \tag{3.23}$$

where $\boldsymbol{\theta} = (\delta, a, \beta, \omega) = (108.0728, 8.7808, 0.09210, 1.4549)$ is the maximum a posteriori estimate. The fundamental matrix of the homogeneous system is

$$\boldsymbol{\Psi}(t) = \left[e^{rt}(\boldsymbol{a}\cos(\mu t) - \boldsymbol{b}\sin(\mu t)) \; e^{rt}(\boldsymbol{a}\sin(\mu t) + \boldsymbol{b}\cos(\mu t)) \right], \tag{3.24}$$

where $\lambda_1, \lambda_2 = r \mp i\mu$, $r \approx -0.1500$, $\mu \approx 5.4399$ and $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 = \boldsymbol{a} \pm i\boldsymbol{b}$, $\boldsymbol{a} \approx \begin{pmatrix} -0.0050 \\ 0.9835 \end{pmatrix}$, and $\boldsymbol{b} \approx \begin{pmatrix} 0.1807 \\ 0 \end{pmatrix}$. Then the general solution of the enriched model is

$$\boldsymbol{u}(t) = \boldsymbol{\Psi}(t)\boldsymbol{z} + \boldsymbol{\Psi}(t) \int_{t_0=0}^{t} \boldsymbol{\Psi}^{-1}(s)\boldsymbol{g}(s)ds, \tag{3.25}$$

where $\boldsymbol{z} \approx \begin{pmatrix} 0 \\ 27.6757 \end{pmatrix}$. We numerically solve the integral in Eq. (3.25), and a comparison of the analytical solution for the linear detailed model, linear reduced model, and the enriched model is shown in Fig. 3.3. The reduced model's trajectory of the mass's positions quickly deviates from the detailed model's trajectory. In particular, the reduced model underestimates the amplitude, overestimates the frequency, and misses higher frequency behavior. The enriched model almost perfectly matches the trajectory of the linear detailed model, but there are slight discrepancies in the trajectories when their tangents are close to zero.

Linear Numerical Results Here, we present numerical experiments for a few combinations of (N, M) of the linear detailed model (i.e., when $\alpha_i = 0 \; \forall i \in (1, \ldots, N)$). Numerical results are shown in Figs. 3.4 and 3.5 (and captions give exact values of the initial conditions and parameters). The enriched model is calibrated on noisy observations from the detailed model. Then, for the validation scenario, the values of the masses in the detailed model are changed (both increased and decreased), and the calibrated enriched model is used without recalibrating. Overall, the enriched model captures the linear detailed model's amplitude and frequency. By "captures," we mean that the uncertainty in the parameters and hyperparameters propagates through to the model predictions, and the observations fall within the predictions' band of uncertainty. However, the enriched model is not able to replicate the total variations of the observations. Specifically, there were small discrepancies between observations and model predictions when the tangents of the trajectories approached zero.

Nonlinear Numerical Results As the final example, we present numerical experiments for a few combinations of (N, M) for the nonlinear detailed model (i.e., when $0 < \alpha_i < 1 \; \forall i \in (1, \ldots, N)$). Numerical results from the nonlinear systems are shown in Figs. 3.6 and 3.7. Overall, the enriched model usually captures the nonlinear detailed model's amplitude and
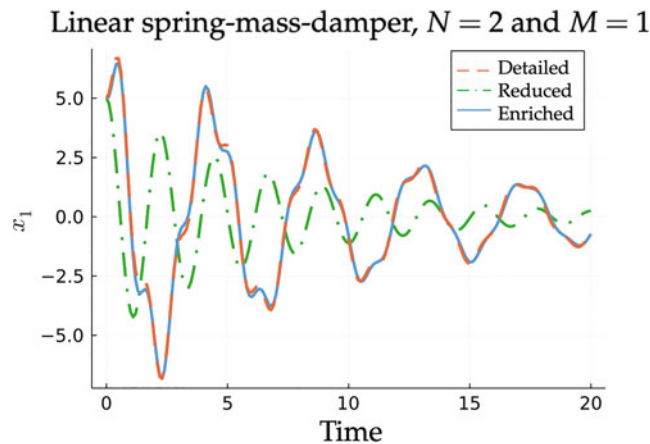
Fig. 3.3 Resulting trajectories from the analytical solutions to the linear detailed model (red dashed line), the linear reduced model (green dash-dotted line), and the enriched model (blue line)
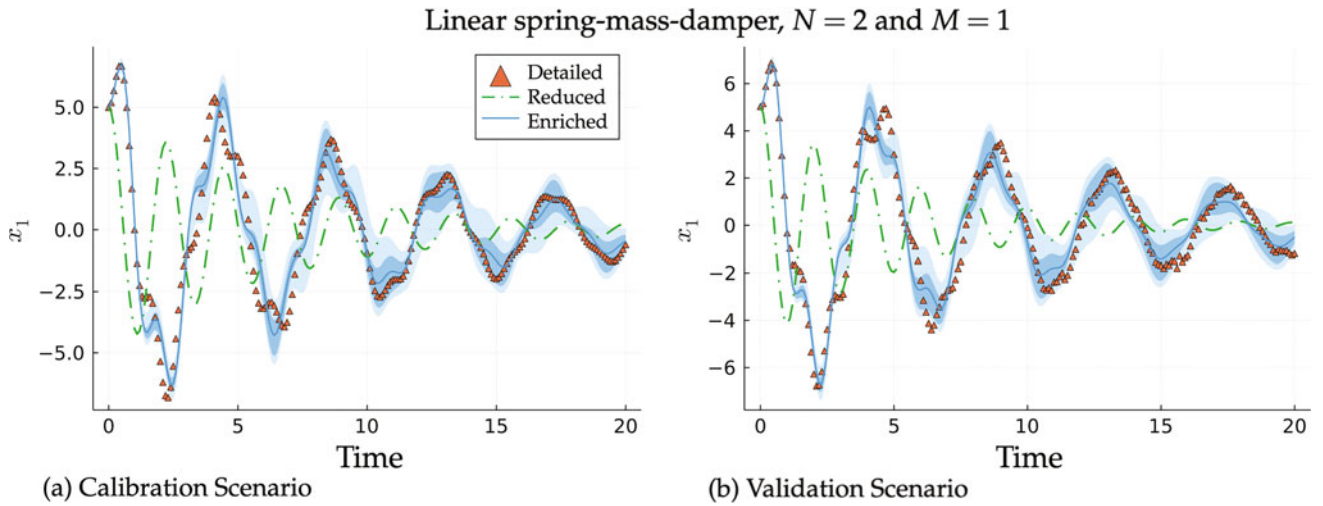
Fig. 3.4 Enriched model mean predictions (solid blue curve) plotted with 50% and 95% confidence intervals (shaded regions) compared to observations from the linear detailed model (red triangles) and output from the linear reduced model (dashed green curve) where $N = 2$ and $M = 1$. The initial conditions and parameters of the detailed model in (**a**) the calibration scenario are as follows: $x(t = 0) = (5, 10)^T$, $v(t = 0) = (0, 0)^T$, $m = (5, 10)^T$, $c = (1.5, 1.5)^T$, and $k = (40, 80)^T$. In (**b**), the validation scenario, $m = (4, 11)^T$, and the enriched model is not recalibrated
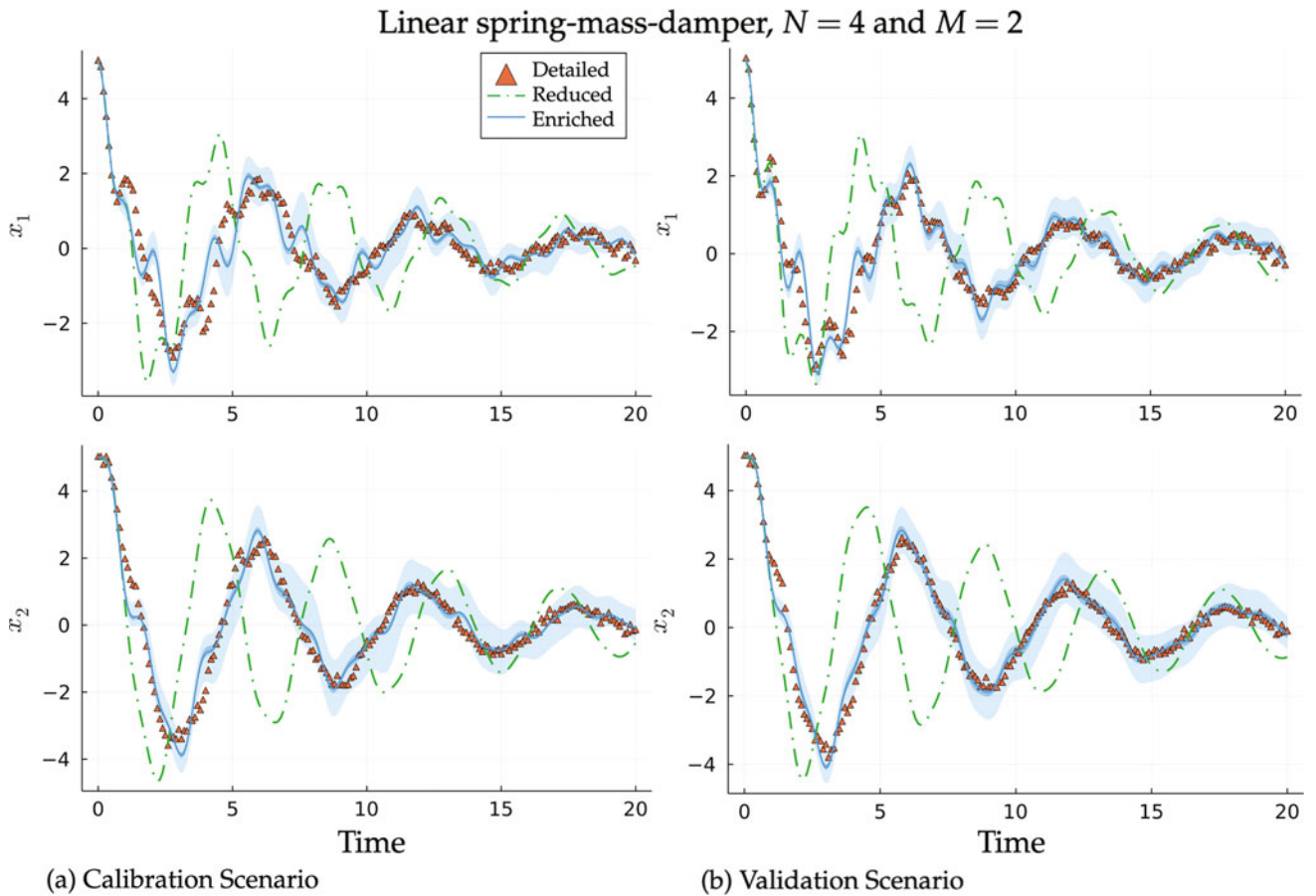
Fig. 3.5 Enriched model predictions (solid blue curve) plotted with 50% and 95% confidence intervals (shaded regions) compared to observations from the linear detailed model (red triangles) and output from the linear reduced model (dashed green curve) where $N = 4$ and $M = 2$. The initial conditions and parameters of the detailed model in (**a**) the calibration scenario are as follows: $x(t = 0) = (5, 5, 5, 5)^T$, $v(t = 0) = (0, 0, 0, 0)^T$, $m = (5, 10, 7, 3)^T$, $c = (1.5, 1.5, 1.5, 1.5)^T$, and $k = (40, 80, 120, 60)^T$. In (**b**), the validation scenario, $m = (4, 11, 8, 2)^T$, and the enriched model is not recalibrated
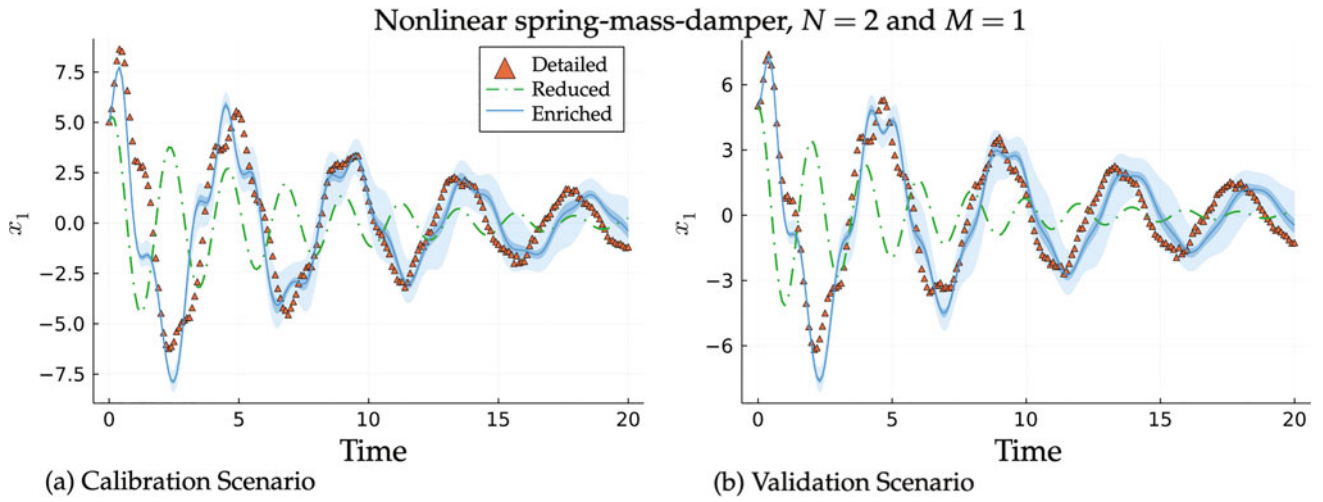
Fig. 3.6 Enriched model predictions (solid blue curve) plotted with 50% and 95% confidence intervals (shaded regions) compared to observations from the nonlinear detailed model (red triangles) and output from the nonlinear reduced model (dashed green curve) where $N = 2$ and $M = 1$. The initial conditions and parameters of the detailed model in (**a**) the calibration scenario are as follows: $x(t = 0) = (5, 10)^T$, $v(t = 0) = (0, 0)^T$, $m = (5, 10)^T$, $c = (1.5, 1.5)^T$, $k = (40, 80)^T$, and $\alpha = (0.05, 0.05)^T$. In (**b**), the validation scenario, $m = (4, 11)^T$, and the enriched model is not recalibrated
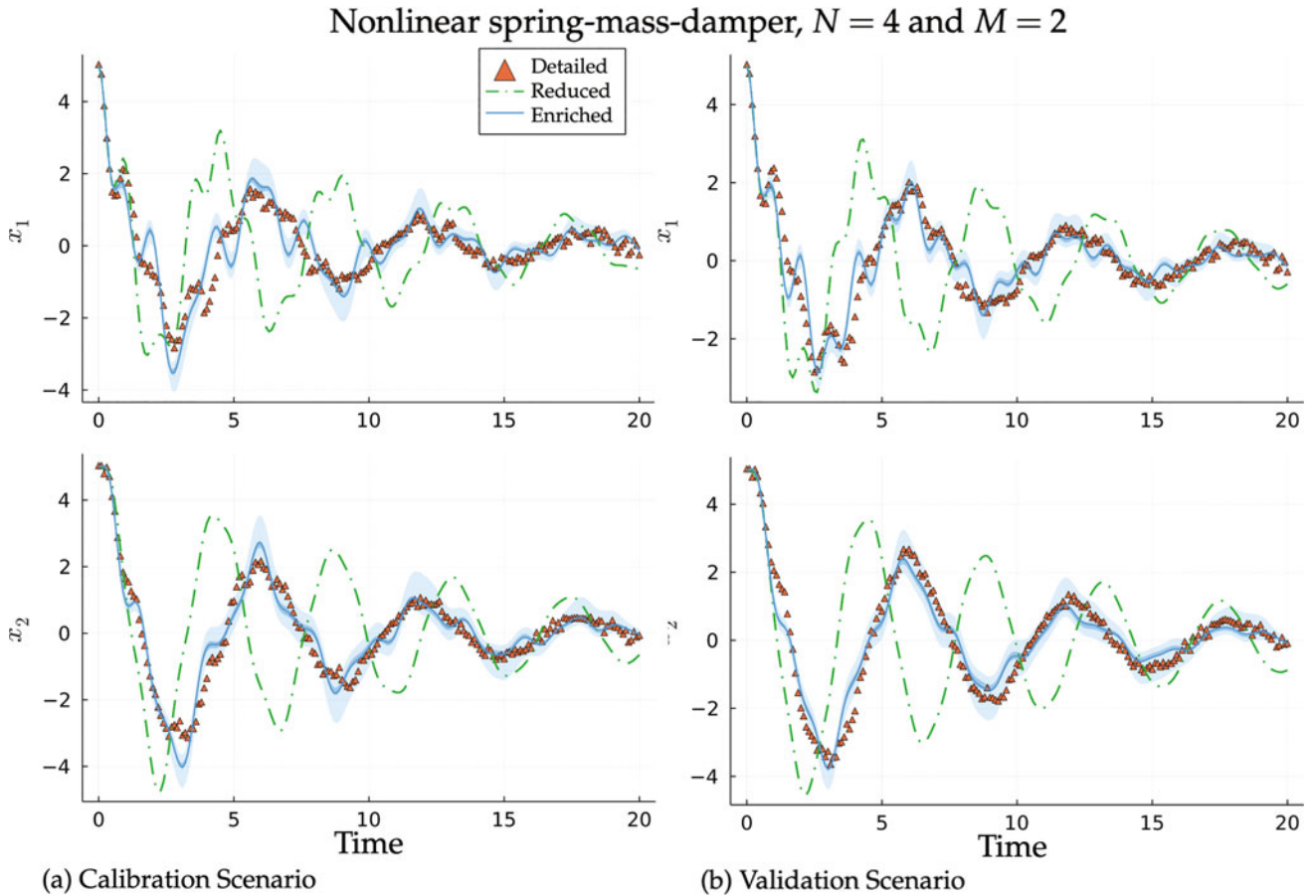
Fig. 3.7 Enriched model predictions (solid blue curve) plotted with 50% and 95% confidence intervals (shaded regions) compared to observations from the nonlinear detailed model (red triangles) and output from the nonlinear reduced model (dashed green curve) where $N = 4$ and $M = 2$. The initial conditions and parameters of the detailed model in (**a**) the calibration scenario are as follows: $x(t = 0) = (5, 5, 5, 5)^T$, $v(t = 0) = (0, 0, 0, 0)^T$, $m = (5, 10, 7, 3)^T$, $c = (1.5, 1.5, 1.5, 1.5)^T$, $k = (40, 80, 120, 60)^T$, and $\alpha = (0.1, 0.1, 0.1, 0.1)^T$. In (**b**) the validation scenario, $m = (4, 11, 8, 2)^T$, and the enriched model is not recalibrated

Table 3.1 Fraction of γ-values below the tolerance threshold for the calibration scenario ($\boldsymbol{D}_c$) and the validation scenario ($\boldsymbol{D}_v$) for each system

	Linear		Nonlinear	
System	$N = 2, M = 1$	$N = 4, M = 2$	$N = 2, M = 1$	$N = 4, M = 2$
$G(\boldsymbol{\gamma}, \boldsymbol{D}_c)$	0.04	0.01	0.11	0.06
$G(\boldsymbol{\gamma}, \boldsymbol{D}_v)$	0.00	0.02	0.11	0.04

frequency, but the enriched model systemically overestimates the trough around 2.5 s. In Fig. 3.7, the detailed models' trajectories are smoother than the enriched models' trajectories, and there are small discrepancies when the trajectories' tangents approach zero. We hypothesize that the unmodeled masses from the detailed model have a smoothing effect on the first M masses' trajectories, while the enriched model's trajectories are more volatile.

Validation Results Quantitative validation results are shown in Table 3.1, where the fraction of γ-values below the tolerance threshold is reported for each calibration and validation scenario for a few combinations of (N, M) for the linear system and the nonlinear system. These results show it is uncommon that the enriched model output does not capture an observation, indicating that the enriched model recovers most of the model error caused by omitting the dynamics of objects $M + 1$ through N from the ODEs of the first M objects.

3.8   Conclusion

This study seeks to represent model discrepancy in linear and nonlinear oscillatory systems by placing an inadequacy operator within an ODE of a reduced model to form an enriched model. The inadequacy operator introduces a simple damped oscillator into one velocity derivative from the reduced model to approximate the effect of the omitted masses on the modeled masses. Embedding an inadequacy operator into the reduced model decreased the discrepancy between observations and model predictions while retaining the interpretability of the reduced model. This embedding also allows the enriched model to make extrapolative predictions (i.e., the masses in the detailed system could be changed without needing to recalibrate the enriched model). Furthermore, this approach only relies on black-box observations from the detailed model, which could be replaced with experimental observations. While the nonlinear enriched model did not perfectly reproduce all of the observations, it could predict the overall amplitude and frequency of the masses' positions, and it significantly decreased the model error of the reduced model.

Generally, as the number of omitted masses $N - M$ increases, the model discrepancy increases, which intuitively makes sense because there are more omitted forces affecting the first M masses. In future work, we will explore how the number of omitted masses impacts model discrepancy. In particular, can the model error be further decreased by adding more simple damped oscillators to the inadequacy operator? We would like to investigate additional forms of model error including inaccuracies caused by simplified forces. For example, the detailed model could involve several nonlinear forces on N masses, while the reduced model still uses linear forces on M masses. Finally, we plan to investigate more complex dynamical properties of true interacting system, such as solitons, bifurcations, and chaos.

References

1. Morrison, R.E., Oliver, T.A., Moser, R.D.: Representing model inadequacy: a stochastic operator approach. SIAM/ASA J. Uncertain. Quantif. **6**(2), 457–496 (2018)
2. Hegedűs, L., Wittmann, M., Noszticzius, Z., Yan, S., Sirimungkala, A., Försterling, H.-D., Field, R.J.: HPLC analysis of complete BZ systems. evolution of the chemical composition in cerium and ferroin catalysed batch oscillators: experiments and model calculations. Faraday Discuss. **120**, 21–38 (2002)
3. Fort, H.: Making quantitative predictions on the yield of a species immersed in a multispecies community: the focal species method. Ecol. Model. **430**, 109108 (2020)
4. Noid, W.G.: Perspective: coarse-grained models for biomolecular systems. J. Chem. Phys. **139**(9), 09B201_1 (2013)
5. Farrell, K., Oden, J.T., Faghihi, D.: A Bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems. J. Comput. Phys. **295**, 189–208 (2015)

6. Tuck, J.L., Menzel, M.T.: The superperiod of the nonlinear weighted string (FPU) problem. Adv. Math. **9**(3), 399–407 (1972)

7. Porter, M.A., Zabusky, N.J., Hu, B., Campbell, D.K.: Fermi, Pasta, Ulam and the birth of experimental mathematics: a numerical experiment that Enrico Fermi, John Pasta, and Stanislaw Ulam reported 54 years ago continues to inspire discovery. Am. Sci. **97**(3), 214–221 (2009)

8. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B (Stat Methodol.) **63**(3), 425–464 (2001)

9. Higdon, D., Gattiker, J., Williams, B., Rightley, M.: Computer model calibration using high-dimensional output. J. Am. Stat. Assoc. **103**(482), 570–583 (2008)

10. Oliver, T.A., Terejanu, G., Simmons, C.S., Moser, R.D.: Validating predictions of unobserved quantities. Comput. Methods Appl. Mech. Eng. **283**, 1310–1335 (2015)

11. Portone, T., Moser, R.D.: Bayesian inference of an uncertain generalized diffusion operator. SIAM/ASA J. Uncertain. Quantif. **10**(1), 151–178 (2022)

12. Morrison, R.E.: Data-driven corrections of partial Lotka–Volterra models. Entropy **22**(11), 1313 (2020)

13. Berman, G.P., Izrailev, F.M.: The Fermi–Pasta–Ulam problem: fifty years of progress. Chaos: An Interdisciplinary Journal of Nonlinear Science **15**(1), 015104 (2005)

14. Abdi, H., Williams, L.J.: Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. **2**(4), 433–459 (2010)

15. Benner, P., Gugercin, S., Willcox, K.: A survey of projection-based model reduction methods for parametric dynamical systems. SIAM Rev. **57**(4), 483–531 (2015)

16. Bruna, M., Chapman, S.J., Smith, M.J.: Model reduction for slow–fast stochastic systems with metastable behaviour. J. Chem. Phys. **140**(17), 174107 (2014)

17. Ye, F.X.-F., Yang, S., Maggioni, M.: Nonlinear model reduction for slow-fast stochastic systems near manifolds (2021)

18. Hoffman, M.D., Gelman, A., et al.: The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res. **15**(1), 1593–1623 (2014)

19. Ge, H., Xu, K., Ghahramani, Z.: Turing: a language for flexible probabilistic inference. In: International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9–11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain, pp. 1682–1690 (2018)

20. Berliner, L.M.: Hierarchical Bayesian time series models. In: Hanson, K.M., Silver, R.N. (eds.) Maximum Entropy and Bayesian Methods, pp. 15–22. Springer, Netherlands (1996)

21. Gelman, A., Meng, X.-L., Stern, H.: Posterior predictive assessment of model fitness via realized discrepancies. Stat. Sin. **6**(4), 733–760 (1996). https://www.jstor.org/stable/i24306033

Chapter 4
Event Detection Using Floor Vibrations with a Probabilistic Framework

Check for updates

Yohanna MejiaCruz, Juan M. Caicedo, Zhaoshuo Jiang, and Jean M. Franco

Abstract Using floor vibrations has shown potential in human detection for security and human health applications. A key aspect of these methodologies is identifying the event location on the floor. The excitation can be due to a step, a fall, or another type of activity. Wave propagation methodologies used for this purpose face challenges due to wave dispersion, multipath fading, and unknown energy dissipation mechanisms. A new model is proposed using a Bayesian probabilistic framework to identify the location of the excitation to enable human tracking. In the proposed model, combining information from multiple sensors, the amplitude of the acceleration is a function of the distance from the event location to the sensor's locations, and the unattenuated amplitude, the localization of the event, and the decay rate are represented by random variables. Preliminary results of the probabilistic model were obtained from a ten-impact test bed. The results showed that the model could establish the most likely area of the event location while providing a measure of the uncertainty in the estimation. However, from a probabilistic perspective, the decisions about the localization must be withheld, considering the significant uncertainty in the predicted quantities.

Keywords Floor vibrations · Event detection · Wave dispersion · Bayesian inference · Uncertainty quantification

4.1 Introduction

Event detection technologies are currently used for various applications, including human health monitoring [1, 2], facilities security, or intelligent building operations by using detection technologies to control building resources [3]. Floor vibrations can be used to track human activity for different applications by improving some of the most critical aspects of human tracking technologies, which is privacy. Floor vibration technology for event detection has emerged with the development of different approaches, including time of arrival methods (ToA) and their derivations, which rely on wave propagation and sensors location to derive spatial relationships to locate the event [4, 5]. Other methodologies include the force estimation methods [6] that rely on the dynamic properties of the system characterized by input–output relationships. Wave propagation methods, such as ToA, face challenges associated with wave dispersion across the structural system and low signal-to-noise ratio (SNR) [7]. The location estimation is linked to the sensor's placement and the wave propagating to the sensors' locations, creating multiple-path fading and phase changes. In addition to having a time difference between the time of arrival of the wave to the sensors, the amplitude of the signal also changes. The authors believe that this attenuation variability could be used to estimate the location of the event. This is accomplished using a Bayesian probabilistic framework that idealizes the wave propagation and dispersion from the event location through the location of the sensors. A probabilistic formulation could help with multipath fading by modeling the wave's attenuation rate and distance as random variables.

Y. MejiaCruz (✉) · J. M. Caicedo · J. M. Franco
Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC, USA
e-mail: mejiacru@email.sc.edu; caicedo@cec.sc.edu; francolj@email.sc.edu

Z. Jiang
School of Engineering, San Francisco State University, San Francisco, CA, USA
e-mail: zsjiang@sfsu.edu
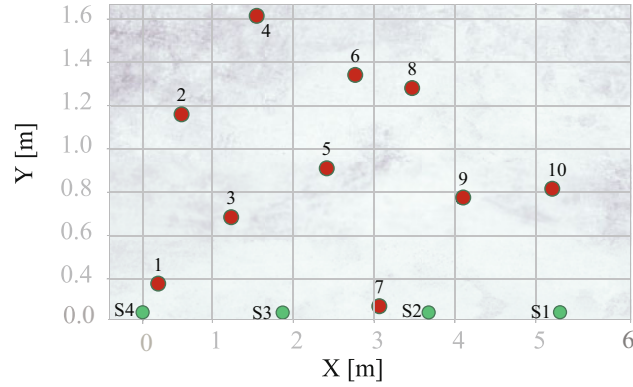
Fig. 4.1 Experimental setup, calibration test with ten impact locations and four sensors

The model presented in this chapter was formulated using a Bayesian framework to provide credible intervals for the model parameters to represent the lack of information on the structural system and the energy dissipation mechanisms. This proposed model aims to find the distribution of the unattenuated acceleration amplitude A_o, the energy dissipation constant α, and the localization of the event X and Y that best represents the wave propagation described by the maximum acceleration captured by each independent sensor. This model combines the information from multiple sensors to improve the estimations. We use a Bayesian framework because it allows modeling parameters to express knowledge, and prior or expert knowledge is vital in the inference process [8–12]. The location model was evaluated using floor vibration data collected at ten locations in a hallway. The first part of this chapter presents the probabilistic formulation of the model and its implementation using impact hammer testing. The second part presents the preliminary results of the localization and its limitations.

4.2 Methods

A calibration experiment was performed in a hallway at the structural laboratory of the University of South Carolina. Calibration consisted of recording acceleration data from four sensors linearly located beside the hallway. Figure 4.1 shows the ten locations impacted with a large-sledge PCB impulse hammer model 086D50 (numbered dots). Floor accelerations were collected using four PCB 393B31 seismic accelerometers positioned with a separation of 1.88 m, as presented in Fig. 4.1 by S1, S2, S3, and S4. Data were collected at a sampling frequency of 2048 Hz.

4.3 Probabilistic Localization Model

The exponential function in Eq. (4.1) describes the attenuation of the acceleration's amplitude, where $i = \{1, 2, 3, ...n\}$ for n sensors, R_i represents the distance between the event location and the sensors' locations, and it is described by Eq. (4.2), with $[x_i, y_i]$ representing the coordinates of the sensor i and $[X, Y]$ the coordinates of the unknown location of the event.

$$A_i = A_o e^{-\alpha R_i} \tag{4.1}$$

$$R_i = \sqrt{(X - x_i)^2 + (Y - y_i)^2} \tag{4.2}$$

Prior distributions of model parameters X and Y were chosen using the principle of maximum entropy [13, 14]. Considering that the information at hand is only the hallway space, these parameters' prior knowledge was defined using a uniform distribution. The prior distribution for the model parameters A_0 and α was defined using the information collected from an optimization procedure where data from impact repetitions at the different locations were used. Figure 4.2 presents the distribution of the model parameters according to the optimization. Thus, two exponential distributions best represent A_0 and α parameters.
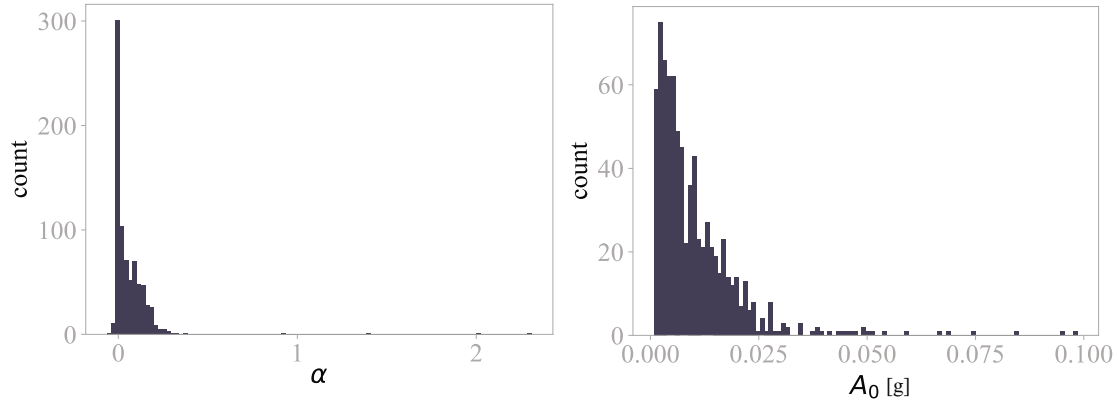
Fig. 4.2 Optimization for α and A_0 to define the prior distribution. Optimization using repetitions of impacts at all locations

$A_o \sim$	$f(x \mid \lambda = 1/0.01031)$
$\alpha \sim$	$f(x \mid \lambda = 1/0.0614)$
$X[m] \sim$	$f(x \mid lower = 0, upper = 6)$
$Y[m] \sim$	$f(x \mid lower = 0 = upper = 1.63)$
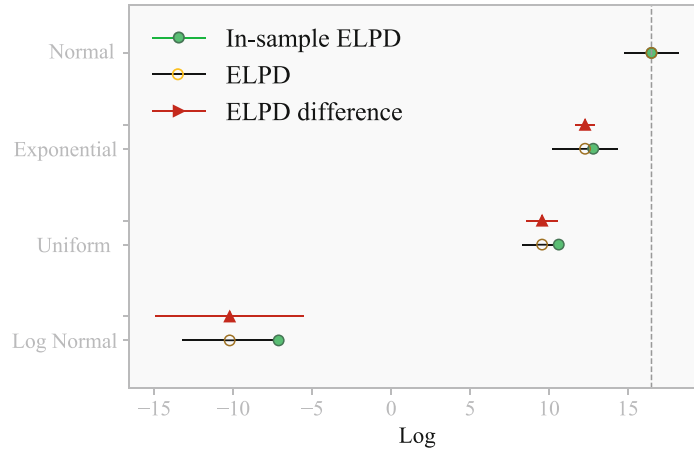
Fig. 4.3 Expected log point-wise predictive density (ELPD) for different likelihood functions. In-sample ELPD: no penalization for the number of parameters. ELPD difference: standard error between the model and the best model. Dashed line: the best assumption

The likelihood used for this analysis is the Gaussian likelihood function presented in Eq. (4.3), which describes the probability of the data D according to the model M and the parameters $\boldsymbol{\theta}$. D represents the known maximum amplitude of the acceleration captured by each sensor. Thus $D = \{A_1, A_2, A_3, A_4\}$. On the other hand, $\boldsymbol{\theta} = \{A_o, \alpha, X, Y\}$ and M is the model, whose mathematical formulation is described in Eq. (4.1). Given the unknown nature of the likelihood function, an evaluation of multiple distributions using leave-one-out cross-validation (LOO) and Pareto Smoothed importance sampling (PSIS) [15] was performed under the same prior distributions. The estimated out-of-sample predictive accuracy was determined in each case and is presented in Fig. 4.3 and Table 4.1. The expected log point-wise predictive density (ELPD) indicates that the Gaussian distribution can be selected as the best assumption for this inference process. Parameter σ_A from Eq. (4.3) was considered in the inference process as a random variable.

$$P(\boldsymbol{D}|\boldsymbol{\theta} M) = \prod_{i=1}^{n} \frac{1}{\sigma_A \sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(A_i - \bar{A}_i(\boldsymbol{\theta} M))^2}{\sigma_A{}^2} \right] \qquad (4.3)$$

Table 4.1 Log point-wise predictive density for different likelihood functions

Model	$P(D|\theta M)$	ELPD
Normal	$f(x \mid \mu, \sigma)$	16.49
Exponential	$f(x \mid \lambda)$	11.58
Uniform	$f(x \mid a, b)$	10.16
Log normal	$f(x \mid \mu, \tau)$	-10.93

4.4 Results and Discussion

The joint posterior distribution was approximated using 9000 samples. A convergence analysis was performed using five parallel chains evaluated using the potential scale reduction factor (PSRF) [16]. The criteria evaluate variances from multiple sequences as $PSRF = \hat{V}/W$. $\hat{V}$ represents the posterior marginal variance, and W represents the within-chain variance. Markov Chain Monter Carlo simulations were obtained using [17] and [18] libraries in Python 3.8 [19].

Results of this analysis are presented in Fig. 4.4 for each impact location presented in Fig. 4.1. The actual localization of the event seems to be within the neighborhood of higher density for parameters X and Y. However, the results further indicate that the 95HDP is too broad to make a final decision on the predicted localization of the event. From the results presented, it is also clear that in all cases the posterior distribution for the model parameter X changed considerably from the prior distribution, according to the data used to inform the model, in comparison with the parameter Y. The last results can be associated with the fact that the information provided to the model comes from the sensors aligned in the -x- direction, which indicates that there is more information for X compared to Y. The authors hypothesize that the results' undecided nature can be linked to the attenuation parameter α, which was selected as a single parameter for all sensors and the fact that the sensors were placed in a line along the x direction. The authors believe this can misrepresent the dissipation from the event location to each sensor location associated with the structural configuration and the sensors' placement. A future approach to improve the results can be modeling the dissipation parameter as a function of X and Y.

4.5 Conclusions

A probabilistic approach to establish the localization of an event using floor vibrations was presented. The model was formulated considering the unattenuated acceleration amplitude, the dissipation rate, and the localization of the event as random variables. The model's formulation is associated with the dissipation of the wave from the event location to the sensors' location. The model results showed that although the actual localization of the event is within the area of higher density in the joint posterior distribution for the localization parameters X and Y, the decision must be withheld considering the significantly higher uncertainty on these estimations. Additional model parameters could help represent the multipath-fading nature of the wave across the structural system in a future implementation.

4.6 Disclaimers

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
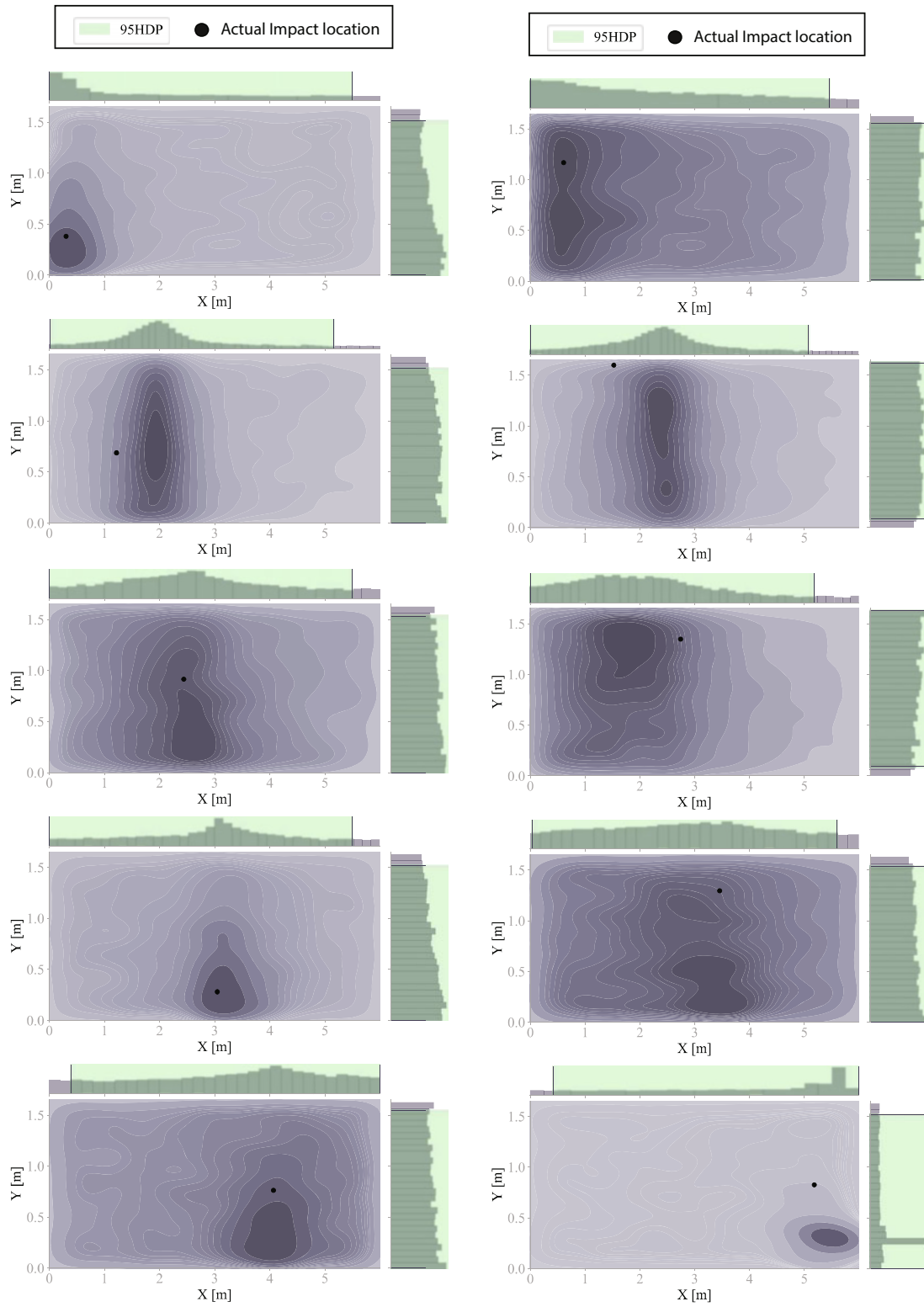
Fig. 4.4 Joint posterior distribution for parameters *X* and *Y* representing the localization of the event for Impacts at locations as referenced in Fig. 4.1

References

1. Kessler, E., Sriram Malladi, V.V.N., Tarazaga, P.A.: Vibration-based gait analysis via instrumented buildings. Int. J. Distrib. Sens. Netw. **15**(10), 1550147719881608 (2019)
2. MejiaCruz, Y., Franco, J., Hainline, G., Fritz, S., Jiang, Z., Caicedo, J.M., Davis, B., Hirth, V.: Walking speed measurement technology: a review. Current Geriatrics Reports **10**, 1–10 (2021)
3. King, J., Perry, C.: Smart buildings: using smart technology to save energy in existing buildings. In: Amercian Council for an Energy-Efficient Economy (2017)
4. Chen, W., Guan, M., Wang, L., Ruby, R., Wu, K.: FLOC: Device-free passive indoor localization in complex environments. In: 2017 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE, New York (2017)
5. Tang, X., Huang, M.-C., Mandal, S.: An "internet of ears" for crowd-aware smart buildings based on sparse sensor networks. In: 2017 IEEE SENSORS, pp. 1–3. IEEE, New York (2017)
6. MejiaCruz, Y., Jiang, Z., Caicedo, J.M., Franco, J.M.: Probabilistic force estimation and event localization (PFEEL) algorithm. Eng. Struct. **252**, 113535 (2022)
7. MejiaCruz, Y., Davis, B.T.: Event reconstructing adaptive spectral evaluation (ERASE) approach to removing noise in structural acceleration signals. Exp. Tech. **48**, 1–11 (2022)
8. Bayarri, M.J., Berger, J.O.: The interplay of Bayesian and frequentist analysis. Stat. Sci. **19**, 58–80 (2004)
9. Efron, B.: Bayesians, frequentists, and scientists. J. Am. Stat. Assoc. **100**(469), 1–5 (2005)
10. González-Torre, F.J.G., Bas, E.M.: On the frequentist and Bayesian approaches to hypothesis testing. SORT: Stat. Oper. Res. Trans. **30**(1), 3–54 (2006)
11. Moreno, E., Girón, F.J.: On the frequentist and Bayesian approaches to hypothesis testing (invited article with discussion: George casella, daniel peña and christian p. robert). SORT-Stat. Oper. Res. Trans. **30**(1), 3–54 (2006)
12. Vallverdú, J.: The false dilemma: Bayesian vs. frequentist. arXiv preprint arXiv:0804.0486 (2008)
13. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
14. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review **5**(1), 3–55 (2001)
15. Vehtari, A., Gelman, A., Gabry, J.: Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. **27**(5), 1413–1432 (2017)
16. Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**(4), 457–472 (1992)
17. Salvatier, J., Wiecki, T.V., Fonnesbeck, C.: Probabilistic programming in python using PYMC3. PeerJ Computer Science **2**, e55 (2016)
18. Kumar, R., Carroll, C., Hartikainen, A., Martin, O.A.: ArviZ a Unified Library for Exploratory Analysis of Bayesian Models in Python (2019)
19. Van Rossum, G., Drake F.L., Jr.: Python Reference Manual. In: Centrum voor Wiskunde en Informatica Amsterdam (1995)

Chapter 5
Advancing Model Credibility for Linked Multi-physics Surrogate Models Within a Coupled Digital Engineering Workflow of Nuclear Deterrence Systems

Sofie W. Schunk, Shane McMurray, and Jake A. Gonzales

Abstract Sandia National Laboratories' (SNL) digital engineering transformation initiative to accelerate product realization of nuclear deterrence (ND) systems has institutionalized quick turn modeling and simulation solutions. Surrogate modeling, coupled with model-based systems engineering (MBSE) using commercial-off-the-shelf (COTS) tools, moves the start line forward to inform design and requirements. Yet, this paradigm shift poses a large challenge in a high-security environment: quick-turn credibility solutions and verification, validation, and uncertainty quantification (VVUQ) to match the rate at which models are developed.

This project demonstrates a model credibility process generating evidence to obtain buy-off from key stakeholders for rapidly developed (<2–3 hours) surrogate models within MATLAB/Simulink that interface with SNL-developed codes and MBSE in an extended integrated digital engineering workflow. The pilot project under the test of this process utilizes legacy higher fidelity and computationally expensive codes to inform mass/stiffness matrices for a structural and aero-dynamics trade study problem that verifies requirements—all on a standard desktop used by the customer vs. need for high-computing power and/or subject matter experts (SMEs).

Our technical approach is directed at risk-informed decision-making for design engineers waiting on requirements, up to program leadership making key decisions. The steps include: (1) benchmarking against current VVUQ processes guided by SMEs; (2) uncertainty inventory including source definition, quantification, and mapping (model form, parametric, numerical, and environmental boundary conditions); (3) mapping of uncertainties to modeling activities; and (4) aggregation of evidence to fill gaps identified (e.g., peer review of methodology) or identify risks where additional testing or data may be required. This approach is underpinned by data engineering and configuration management that face need-to-know security challenges creating innovative capability adaptation for national security defense applications.

In summary, digital engineering workflows utilizing multi-physics surrogate models integrated with MBSE and data management are the way of the future for SNL—assuming associated credibility evidence, accessibility, and usability advances in parallel. The techniques discussed are an integral step in this process and how these types of models can help inform higher fidelity models, qualification, and beyond.

Keywords Surrogate model · Digital engineering · Model credibility · Multi-physics modeling

5.1 Introduction

Moving toward a digital engineering product lifecycle (Fig. 5.1) in order to accelerate product realization of nuclear deterrence systems requires integrated and innovative, yet *credible*, capabilities within a model-based ecosystem that are usable and accessible. A generalized credibility process was created in order to address apprehension and challenges of using surrogate commercial-off-the-shelf (COTS)-based models early in the design process (Phases 1 and 2) to support risk-

S. W. Schunk (✉) · S. McMurray · J. A. Gonzales
Virtual Technologies and Engineering Department, Sandia National Laboratories, Albuquerque, NM, USA
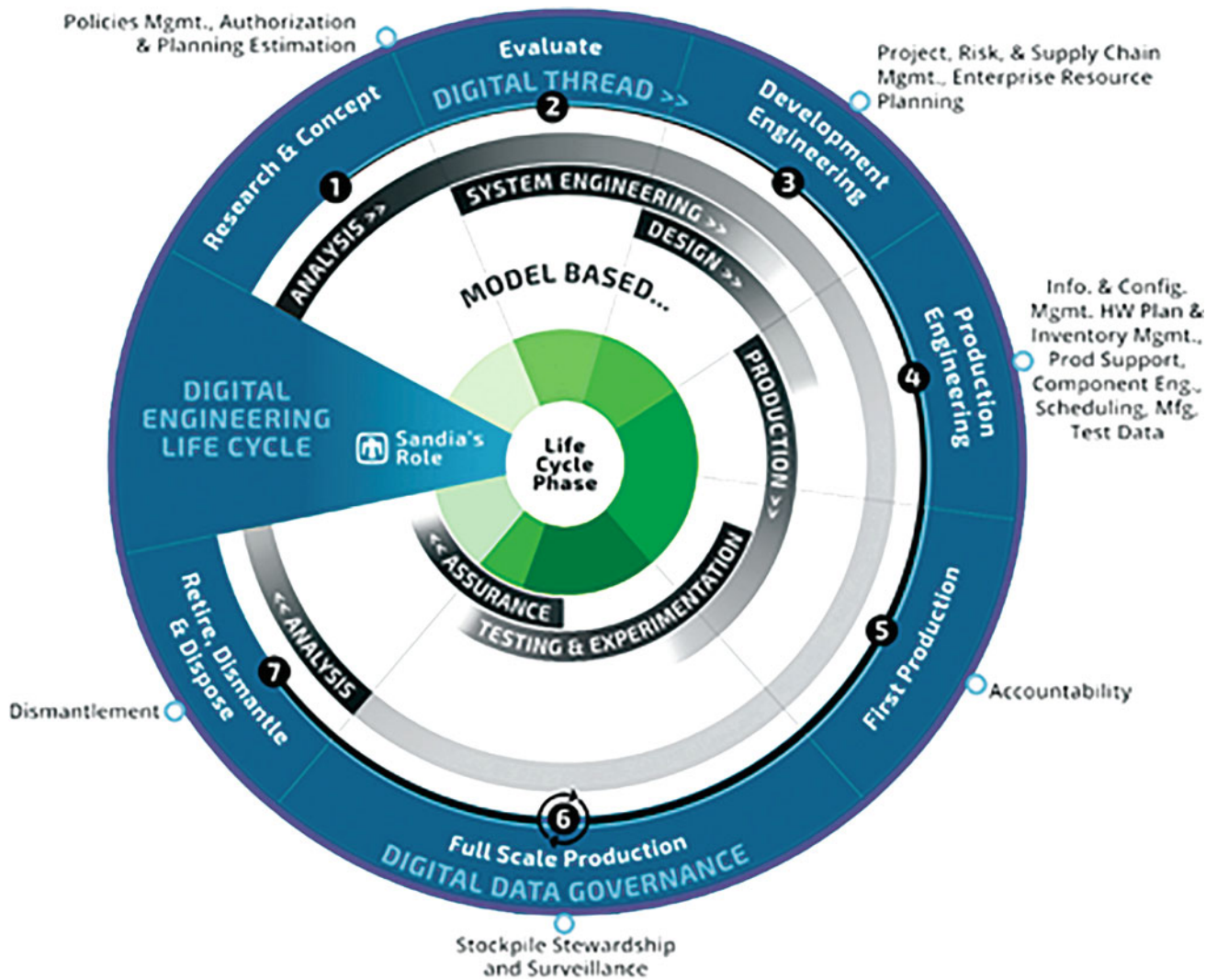
Fig. 5.1 The digital engineering product lifecycle at Sandia National Laboratories

informed decision-making, quicker. Further, this process breaks down silos by integrating into the digital thread, including traceability to requirements via model-based systems engineering (MBSE), creating an iterative environment that can be updated and matured as a product develops, ultimately informing higher fidelity model development and qualification. Challenges surrounding need-to-know and security remain but building an adaptable credibility evidence package for models *early* moves the goal of accelerating product realization at Sandia via digital engineering transformation one step closer.

5.2   Background

Historically, most of the Modeling and Simulation (M&S) at Sandia has focused on high-fidelity computational codes developed by SMEs geared toward areas where physical test and qualification is not possible. Hence, the Predictive Capability Maturity Model (PCMM) [1] displayed in Fig. 5.2 along with other extensive verification, validation, and uncertainty quantification (VVUQ) and credibility processes were established for evaluating model credibility. As surrogate and other reduced-order modeling efforts increased, including the increased use of COTS codes, the need for adapting current processes

MATURITY / ELEMENT	Maturity Level 0 — Low Consequence, Minimal M&S Impact, e.g. Scoping Studies	Maturity Level 1 — Moderate Consequence, Some M&S Impact, e.g. Design Support	Maturity Level 2 — High-Consequence, High M&S Impact, e.g. Qualification Support	Maturity Level 3 — High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification
Representation and Geometric Fidelity — What features are neglected because of simplifications or stylizations?	• Judgment only • Little or no representational or geometric fidelity for the system and BCs	• Significant simplification or stylization of the system and BCs • Geometry or representation of major components is defined	• Limited simplification or stylization of major components and BCs • Geometry or representation is well defined for major components and some minor components • Some peer review conducted	• Essentially no simplification or stylization of components in the system and BCs • Geometry or representation of all components is at the detail of "as built", e.g., gaps, material interfaces, fasteners • Independent peer review conducted
Physics and Material Model Fidelity — How fundamental are the physics and material models and what is the level of model calibration?	• Judgment only • Model forms are either unknown or fully empirical • Few, if any, physics-informed models • No coupling of models	• Some models are physics based and are calibrated using data from related systems • Minimal or ad hoc coupling of models	• Physics-based models for all important processes • Significant calibration needed using separate effects tests (SETs) and integral effects tests (IETs) • One-way coupling of models • Some peer review conducted	• All models are physics based • Minimal need for calibration using SETs and IETs • Sound physical basis for extrapolation and coupling of models • Full, two-way coupling of models • Independent peer review conducted
Code Verification — Are algorithm deficiencies, software errors, and poor SQE practices corrupting the simulation results?	• Judgment only • Minimal testing of any software elements • Little or no SQE procedures specified or followed	• Code is managed by SQE procedures • Unit and regression testing conducted • Some comparisons made with benchmarks	• Some algorithms are tested to determine the observed order of numerical convergence • Some features & capabilities (F&C) are tested with benchmark solutions • Some peer review conducted	• All important algorithms are tested to determine the observed order of numerical convergence • All important F&Cs are tested with rigorous benchmark solutions • Independent peer review conducted
Solution Verification — Are numerical solution errors and human procedural errors corrupting the simulation results?	• Judgment only • Numerical errors have an unknown or large effect on simulation results	• Numerical effects on relevant SRQs are qualitatively estimated • Input/output (I/O) verified only by the analysts	• Numerical effects are quantitatively estimated to be small on some SRQs • I/O independently verified • Some peer review conducted	• Numerical effects are determined to be small on all important SRQs • Important simulations are independently reproduced • Independent peer review conducted
Model Validation — How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy?	• Judgment only • Few, if any, comparisons with measurements from similar systems or applications	• Quantitative assessment of accuracy of SRQs not directly relevant to the application of interest • Large or unknown experimental uncertainties	• Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs • Experimental uncertainties are well characterized for most SETs, but poorly known for IETs • Some peer review conducted	• Quantitative assessment of predictive accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application • Experimental uncertainties are well characterized for all IETs and SETs • Independent peer review conducted
Uncertainty Quantification and Sensitivity Analysis — How thoroughly are uncertainties and sensitivities characterized and propagated?	• Judgment only • Only deterministic analyses are conducted • Uncertainties and sensitivities are not addressed	• Aleatory and epistemic (A&E) uncertainties propagated, but without distinction • Informal sensitivity studies conducted • Many strong UQ/SA assumptions made	• A&E uncertainties segregated, propagated and identified in SRQs • Quantitative sensitivity analyses conducted for most parameters • Numerical propagation errors are estimated and their effect known • Some strong assumptions made • Some peer review conducted	• A&E uncertainties comprehensively treated and properly interpreted • Comprehensive sensitivity analyses conducted for parameters and models • Numerical propagation errors are demonstrated to be small • No significant UQ/SA assumptions made • Independent peer review conducted

Fig. 5.2 General descriptions for PCMM table entries which act as part of Sandia's current process

was apparent. An initial exercise of applying historic processes to a SOLSTICE[1] surrogate aerodynamics and structural dynamics highlighted the following:

• The PCMM can be used as a technical basis for elements to consider for surrogate multi-physics credibility (e.g., model form error when test data is available, which oftentimes is not the case).
• Uncertainty Quantification and Sensitivity Analysis techniques can be applied, but they are often performed by built-in toolboxes within MathWorks products (for the SOLSTICE example) which may require additional tool verification.
• Code commenting and documentation, peer review (expert judgment and customer feedback), model boundary conditions, and interfaces with other codes are gaps that must be considered.
• Credibility evidence evolves over time along with the model and must be a dynamic and flexible process.
• A VVUQ expert is often required, slowing down the process due to funding and resource constraints.

Based on this benchmarking exercise, a hybrid approach was developed for such models, enabling modelers to establish credibility quickly (<2–3 hours) and without the need for an expert (other than consultation), that can be passed along to design engineers. The intended use is for quick-turn surrogate COTS-based models used early in the design phase of the product lifecycle and not to replace current processes, but provide alternative solutions based on the application and use of the model.

---

[1] SOLSTICE is an acronym for Simulation Of Linked multi-physics Surrogate Time-domain models In Combined Environments, a modeling methodology used at Sandia National Laboratories.

5.3 Process

A generalized credibility process was created for use when creating and implementing surrogate models within a digital engineering workflow, as in Fig. 5.3. While not limited to this framework, a SOLSTICE model is used to demonstrate the necessary elements of this process, and the upstream/downstream effects of credibility based on the many integrations and interfaces. Model-based systems engineering, described as a "Descriptive Model" below, provides a means of tracing and providing requirements for verification activities, whether it be through M&S (e.g., SOLSTICE) or test data.

The process as applied to a SOLSTICE aerodynamics/structural dynamics model is as follows:

Part 1: Create Model Credibility Document
- If within MATLAB/Simulink, model preparation includes adding all variables and parameters to the workspace and ensuring no errors are present. The model credibility script can then be run (in less than 2 minutes!), which generates a document printout. The following sections are then completed:
- *Introduction*: Includes customer needs, scope of model, and optimal use case for decision-making with the model, important to defining boundary conditions around model usage.
- *Model Overview*: Describes basic functions, including inputs, outputs, toolboxes used (e.g., Simulink blocks, MATLAB commands, etc.), and other code interfaces or origins of the surrogate model (CAD, for instance, or other higher-fidelity models).
- *Model Equations and Assumptions*: All equations and behaviors modeled to generate output are listed along with all variables and units defined. Assumptions are documented for each equation, including why it was made, and how the model might improve if such assumptions are addressed. Additionally, this is where simplifications to improve run time are documented and uncertainties surrounding the simplification are quantified—for instance, if the model is a surrogate of a higher fidelity model, associated uncertainties inherited from that model are provided here.
- *Model Parameters*: A table is created by the model credibility script, listing all parameters (inputs) with names, units, and values used to create outputs.
- *Sensitivity Analysis*: Evaluates how input parameters of a model influence the model output or specific design requirements. For this use case (or any MATLAB/Simulink model), MATLAB's Sensitivity Analyzer[2] can be used to
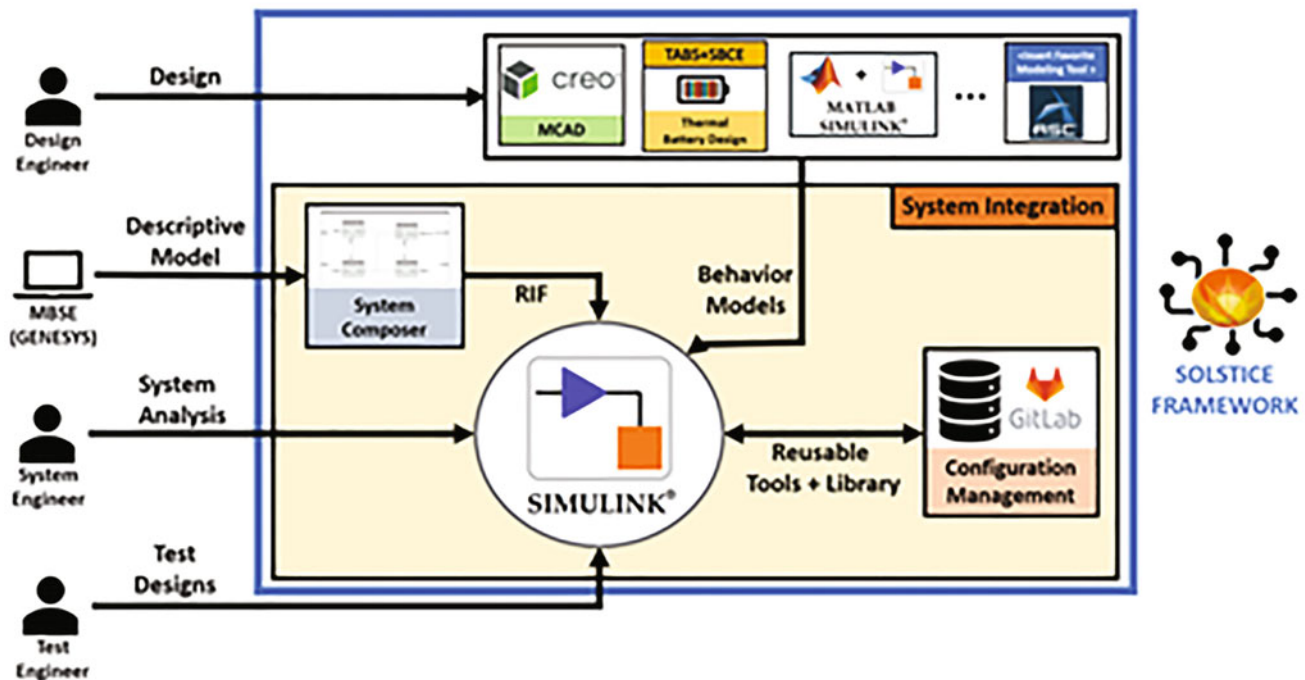
Fig. 5.3 SOLSTICE framework within the digital engineering workflow

---

[2] Reference MathWorks documentation on sensitivity analysis: https://www.mathworks.com/help/sldo/sensitivity-analysis.html
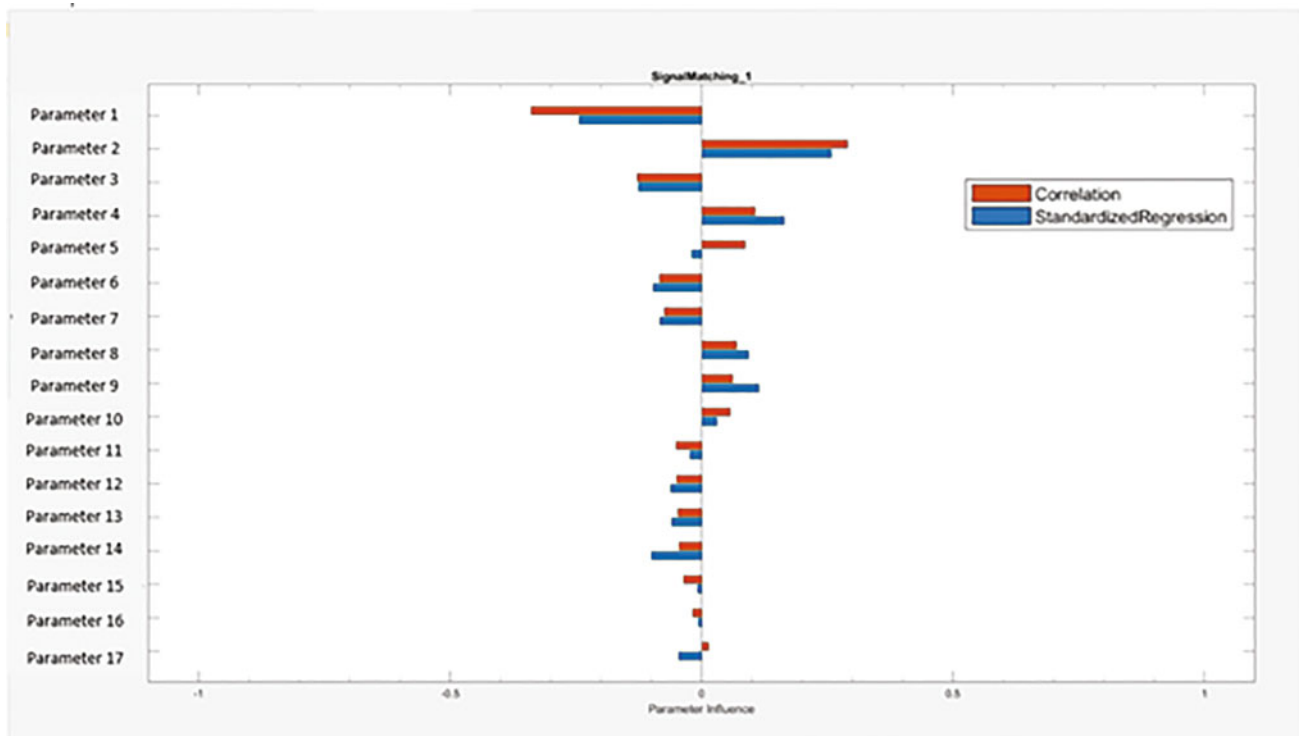
Fig. 5.4 Example tornado plot with statistical analysis of the sensitivity analysis showing correlation coefficients between each parameter and the requirement (*red*) and the normalized regression slope of the parameters vs. requirement (*blue*)

perform this analysis. Steps include (1) selecting the input parameters that will be used in the analysis and generating N number of samples for each parameter depending on a selected mean and standard deviation, (2) performing Monte Carlo simulations to evaluate design requirements at parameter values, and (3) using visual analysis to plot cost function evaluations against parameter samples in order to identify trends. Statistical analysis (correlation, partial correlation, and standardized regression) is then used to compute correlation coefficients quantifying the relations with either linear (Pearson) analysis or ranked (Spearman) analysis, depending on whether the relationship between the cost function and parameter values is a linear or nonlinear monotonic relationship. A tornado plot visualizes the results (see Fig. 5.4 as an example), showing parameters with the most influence (and magnitude). Results are used to make decisions about the model and each parameter, including the use of the parameter estimation tool to optimize the model.

- *MATLAB/Simulink Simulation Form Error*: Runs the model with various time-steps, solvers, and relative tolerances to get simulation run time, providing insight into optimal model use based on customer needs. If test data is available to compare simulation results, the root-mean square error (RMSE) is also included. Table 5.1 is outputted from the script to discover large discrepancies in run time or RMSE based on the choice of solver, time-step, or relative tolerance (Note: "Time-step" is replaced with solver, or relative tolerance).
- *Optimal Selections for Simulation Solver Options*: Evaluate the recommended options for time-step, solver, and relative tolerance when running the model. Inform these decisions from the information obtained. Document any SME judgment or peer review required in making these decisions.

Part 2: Simulink Report Generator
A Graphical User Interface (GUI) was created and is used to auto-generate a customizable system design description and report on Simulink models within a preferred format (PDF, Word, HTML) documenting the model in detail from all blocks used, system, sub-system, components, properties, formulas, code comments, inputs, outputs, etc., to supplement the model overview and assumption sections with the credibility document. Impact includes near-automatic model documentation to provide to the model customer, stakeholder, or any individual that may not understand the model details without the need to install Simulink.

Table 5.1 Sample table output from executing model credibility script within a SOLSTICE model

Time-step	Run time (s)	RMS Error (power)
Baseline (Fixed step 10^{-6} s)	57.23	–
Fixed step (1 s)	0.48	3.68e44
Fixed step (0.1 s)	0.59	2.24e36
Fixed step (0.01 s)	0.63	1.37e-2
Fixed step (0.001 s)	0.71	4.32e-4
Variable step (max 1 s)	0.60	8.98e-3
Variable step (max 0.1 s)	0.72	3.92e-3
Variable step (max 0.01 s)	0.64	1.36e-3
Variable step (max 0.001 s)	1.06	4.32e-4

Part 3: Model User Guide

Generated based on outputs of the model credibility process in order to provide model usage details and limitations that were discovered (for instance, most sensitive parameters, optimal model solvers, and analysis tolerances or time-steps) to the user or customer so that they can make the most optimal risk-informed decisions based on the model.

Additional credibility elements include model peer review of methodology, traceability to all data sources and/or other codes or interfaces, and configuration management of the model and simulation results as visualized in Fig. 5.3.

5.4 Conclusion

The creation of a quick-turn hybrid generalized credibility process (derived from historic VVUQ techniques at Sandia) enables a model credibility document to be generated for any SOLSTICE model (or, more generally, a surrogate model) within minutes. This greatly impacts the ability to accelerate product realization by providing design engineers, key stakeholders, and program leadership a means to make risk-informed decisions early in design without the need for higher fidelity models and/or physical tests. Additionally, this process can be iterated as often as needed based on changes upstream/downstream within the digital thread and/or requirements. The process is currently being piloted and in use for aerodynamic/structural dynamics and thermal battery applications at Sandia, with the goal to increase the user base exponentially as programs work through the digital engineering product lifecycle and digital engineering becomes the way of the future at Sandia.

Reference

1. Oberkampf, W.L., Pilch, M., Tucano, T.G.: Predictive Capability Maturity Model for Computational Modeling and Simulation prepared for *Sandia National Laboratories*, SAND2007-5948, October 2007

Chapter 6
Estimating the Effect of Noise on Various ARMA-Based Damage-Sensitive Features

Emmett Lepp and Thomas Matarazzo

Abstract Vibration-based damage-sensitive features (DSFs) are a useful diagnostic tool in condition assessment of structural systems. While many DSFs are designed to be insensitive to random noise, in practice the presence of measurement noise can impede the identification of structural damage. In this study, four DSFs for use with autoregressive moving average (ARMA) models were examined with signal-to-noise ratios (SNR) ranging from 40 to 0.5 dB to evaluate the accuracy of damage estimation for increasing levels of noise. All structural responses were generated from a simulated cantilever beam model. Student t-tests were utilized to compare DSFs between the healthy state and an unknown state. The results indicated that the Cosh spectral distance DSF outperforms the other DSFs for SNRs less than 20 (i.e., greater than 1% of noise added). These results are consistent with prior work that demonstrated the robustness of the Cosh spectral distance as a DSF.

Keywords Structural health monitoring · Damage detection · Sensors · Noise · Uncertainty quantification

6.1 Introduction

Sensors emplaced within a variety of different structures can help make decisions on the necessity of structure maintenance, overall health, as well as possibly prevent catastrophic failure of these structures. The field of structural health monitoring attempts to model and predict the states of various structures to prepare for and prevent these failures from occurring. By improving the understanding of how accurate different detection methods are with various levels of noise, a true level of damage detection can be determined. These findings have a broad applicability to predicting damage, and possibly preventing catastrophic failure, given that the level of noise is known at the source.

This research attempts to determine the damage-sensitive feature (DSF) that is the least sensitive to added signal noise while still being able to accurately determine damage. Through the addition of random white noise to the signal through simulations, the noise sensitivity of different damage-sensitive features can be calculated. By repeating this process for a wide range of random noise seeds and averaging the results, a Monte Carlo simulation is performed, determining the long-run average of noise allowance given the manipulation of noise seeds.

This approach can be used to better understand and apply sensor data from structures such as bridges, automobiles, or aircraft to detect and determine damage based on the levels of noise. By understanding how accurate these calculations are at estimating damage over various levels of noise, this research aids in their application within physical structures. Given the level of sensor noise is known, this research attempts to provide a general accuracy within the damage decision for various damage-sensitive features.

This research adds to the field by applying concepts that are used within civil structural health monitoring, such as those seen in Nair et al. [1]. Nair investigates the use of these time-based models to determine the damage to the ASCE structure with deliberate damage introduced in the form of bolt loosening and removal. This research, while applying the same fundamental concept of comparing an unknown state to a healthy state to make a decision as to whether or not damage is present, adds the addition of random white noise to determine how sensitive these calculations are to added noise.

Previous research in this field has validated the approach of structural health monitoring on a variety of civil structures such as highway bridge spans and generic supported beams [2, 3]. This research applies the concept of damage decision-making using acceleration data in combination with statistical testing to determine the likelihood of induced damage. Autoregressive

E. Lepp (✉) · T. Matarazzo

Department of Civil and Mechanical Engineering, United States Military Academy, West Point, NY, USA
e-mail: emmett.lepp@tufts.edu

modeling with exogenous input (ARX) has also been used as a method of damage detection utilizing an unknown input to determine both damage presence and location [4]. This past research has utilized ARX-based coefficients to determine the change in the dynamic response of civil structures, which is indicative of induced damage. Research done by [5] has analyzed some applications of estimating effects of signal noise, to create a confidence interval of damage decision, yet lacks direct comparison for different forms of damage detection under high levels of noise.

Further improvements in the field of structural health monitoring include the increased understanding of sensor likelihood to detect damage. Research conducted by [6] introduces a process of determining how to best detect damage based on the probability of detection (POD) curves, promoting further implementation of sensor networks and damage detection metrics targeted to detect probable forms of damage. These advancements in the field can be used in coordination with advanced sensing techniques, such as those seen in [7] to implement more cost-effective methods of detecting damage.

6.2 Background

Sensors attached to physical structures, such as accelerometers or inertial measurement units (IMU), are capable of reading and recording acceleration to ascertain the structure's response over a given period of time. This data can then be used to make a damage decision based on the time series–based model of a fit autoregressive moving average (ARMA) equation, utilizing damage-sensitive feature calculations. This time-based model, as seen in Eq. 6.1, fits an equation to dictate the system response as gathered by sensors [1].

$$Y(t) + \sum_{i=1}^{n} \alpha_i Y(t-i) = \sum_{i=1}^{n} \beta_i (t-i) + \sum_{i=1}^{q} c_i e(t-i) \tag{6.1}$$

This ARMA equation involves using α, β, and c coefficients of various model orders to equate the system response. By selecting a sufficiently high model order, an accurate representation of the data can be determined with high fit percentage and a relatively low Akaike Information Criterion (AIC) indicating sufficient fit. The ARMA equation α and β coefficients are then used in DSF calculations [1]. The equation for DSF1 was gained from Nair et al. [1] and the equations for DSFs 2 and 3 were inspired by this equation, utilizing similar α and β coefficients to attempt to determine the presence of damage.

$$\text{DSF1} = \frac{\alpha_2}{\sqrt{\alpha_2^2 + \alpha_3^2 + \alpha_4^2}} \tag{6.2}$$

$$\text{DSF2} = \frac{\alpha_2 + \beta_1}{\sqrt{\alpha_2^2 + \alpha_3^2 + \alpha_4^2 + \beta_1^2 + \beta_2^2 + \beta_3^2}} \tag{6.3}$$

$$\text{DSF3} = \frac{\alpha_2 + \beta_1}{\sqrt{(\sum \alpha)^2 + (\sum \beta)^2}} \tag{6.4}$$

These equations represent an arbitrary value that can be compared to other known states to determine whether or not damage has been introduced. Additionally, a Cosh spectral distancing method compares the power spectral density (PSD) of an unknown state to a known healthy state to make a damage decision similar to the above DSF. In the DSF equation, which was developed in [8], $S(\omega_j)$ represents the power spectral density of the unknown state and $\overline{S}(\omega_j)$ represents the power spectral density of the known healthy state.

$$C\left(S, \overline{S}\right) \approx \frac{1}{2N} \sum_{j=1}^{N} \left[\frac{S(\omega_j)}{\overline{S}(\omega_j)} - \log \frac{S(\omega_j)}{\overline{S}(\omega_j)} + \frac{\overline{S}(\omega_j)}{S(\omega_j)} - \log \frac{\overline{S}(\omega_j)}{S(\omega_j)} - 2 \right] \tag{6.5}$$

After determining the DSF for each state, a comparison of the DSF means for the respective states can be examined to determine whether or not there is a statistically significant difference in the means corresponding to an addition of damage to the structure. The null hypothesis is that the mean of the DSFs for the healthy state, μ_{Healthy}, is equal to the mean of DSFs

for the unknown state, μ_{Unkown}, as seen in Eq. 6.6. Rejection of this null hypothesis indicates that damage has been added to the structure. A student t-test was carried out at the 90% significance level to either support or reject the null hypothesis.

$$H_0 : \mu_{\text{Healthy}} = \mu_{\text{Unknown}} \qquad H_a : \mu_{\text{Healthy}} \neq \mu_{\text{Unknown}} \tag{6.6}$$

When noise is added to the signal, the distribution of the DSFs becomes random; however, the means of the DSFs can still be calculated and compared to make a damage decision using the student t-test. In real-world scenarios, noise can be added from a variety of sources such as electric or magnetic interference within the sensor itself. The level of noise present within the signal can be defined by the signal to noise ratio (SNR), as seen in Eq. 6.7, which compares the power of the signal to the power of the noise in decibels.

$$\text{SNR} = 10 \log \frac{P_{\text{signal}}}{P_{\text{noise}}} \tag{6.7}$$

6.3 Methods

This research involves the process of simulating data as generated by accelerometers placed about a cantilever beam, adding white noise to the data, and fitting a time series ARMA equation to the data. These ARMA equations are then used to determine the DSFs based on the α and β coefficients of the ARMA equation. Then a student t-test is utilized to compare the sample mean of the healthy state to the sample mean of the damaged state and determine whether there is a statistically significant difference indicating damage within the structure.

The data generated for this analysis simulates the accelerations at the masses of a four degree-of-freedom cantilever beam model, pictured in Fig. 6.1. This beam simulated was 10 m in length, with a lumped mass spaced evenly at 2.5 m intervals along the span of the beam. The beam was given material properties of Titanium alloy, with a modulus of Elasticity of 114 GPa, a density of 77,000 $\frac{\text{kg}}{\text{m}^3}$. The cross section of the beam is 0.05 m by 0.50 m depth, to represent a generic Euler-Bernoulli beam element. This beam's motion was simulated using a single applied force at the end of the beam, which allowed for acceleration data to be generated as a response of free vibrations of the beam.

The data simulates 30 days of recordings, where damage is deliberately added on Day 11 and an increased amount of damage is added on Day 21. By simulating the cantilever beam with known introduction of damage, it provides the ability to determine how accurate the damage-sensitive feature calculations are, given that they should be detecting damage. Since the four sensors are spread out on the cantilever beam, an accuracy of damage estimation was calculated by finding the number of sensors that were predicting damage out of all sensors present. While this process is rather straightforward with no noise addition, by adding noise the accuracy estimation will change.

The data was collected at a rate of 5 kHz and downsampled to 500 Hz based on knowledge of the band of the modal frequencies. After the data was simulated for each individual sensor, random white noise was added at various SNR levels. The process of adding random noise seeds was then repeated 30 times for each SNR and the results averaged to remove an error which may be a result of the random noise seed. This process of adding noise to the data can be seen in the difference between Figs. 6.2 and 6.3, which represent the average DSFs for the various days without and with added noise, respectively.

From these figures, it can be seen that as noise is added to the sensor, the DSFs display less of a linear pattern and are more spread out. The means of the distributions of DSFs can then be calculated to allow for comparison between damage states. The same application of a student t-test of the means can be carried out to compare the different damage states and determine if the damage is present within the beam. This process was executed for all four sensors in the beam at SNRs ranging from 50 dB to 0.5 dB to represent the range of noise which may be possible within the sensors.
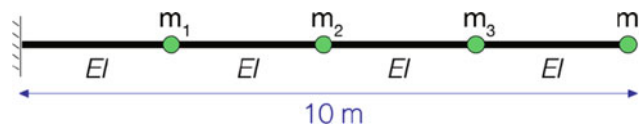
Fig. 6.1 Simulated cantilever beams diagram
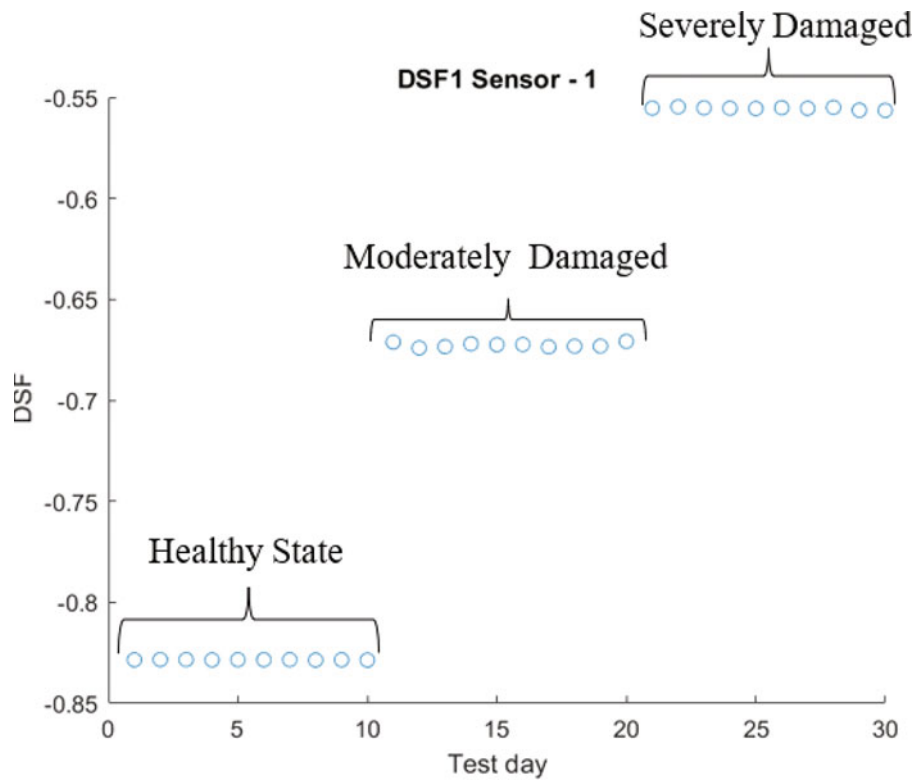
Fig. 6.2 Comparison of damage-sensitive feature groupings with no signal noise for one sensor
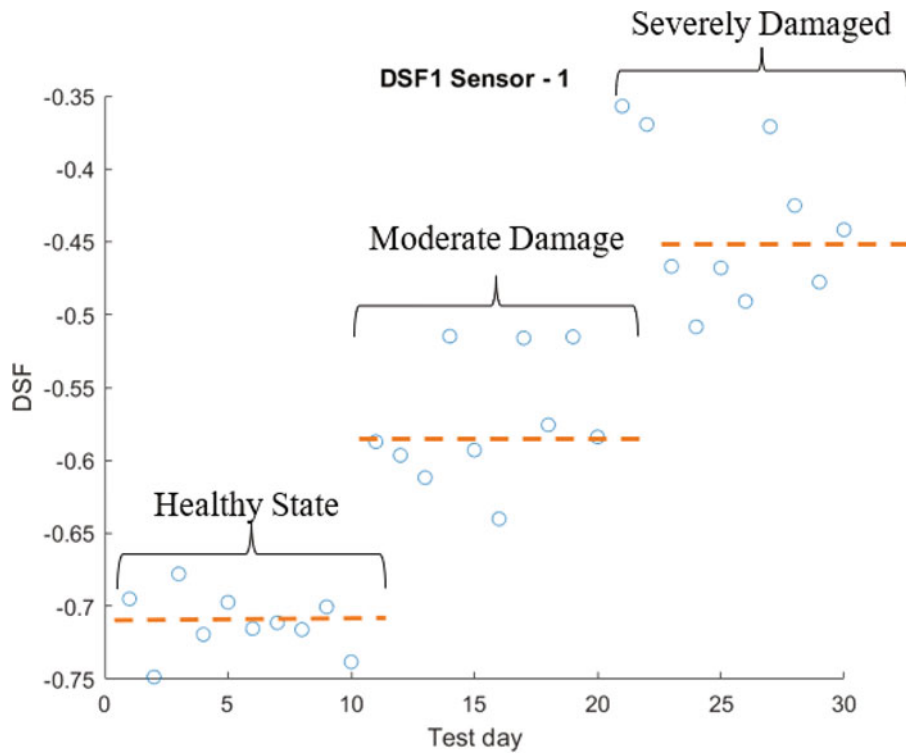
Fig. 6.3 Comparison of damage-sensitive feature groupings with signal noise at an SNR of 20 dB for one sensor

6.4   Results

As the amount of noise is increased within the sensor, it is expected that the damage detection methods become less accurate and, as a result, are able to predict damage less correctly within the beam. Additionally, damage detection for a severely damaged beam is expected to be easier to predict compared to the moderately damaged beam. Figures 6.4 and 6.5 display the results from this research estimating the accuracy of various damage calculations with an increased amount of noise.

When examining the trends from Figs. 6.4 and 6.5, the expected results hold true with the DSFs becoming less accurate in damage estimation for increased levels of noise. Both figures demonstrate that the Cosh spectral distancing method far out preforms the DSF calculations based on the ARMA coefficients alone, as the Cosh spectral distancing method has higher accuracy of damage estimation with increased noise. Additionally, when comparing the results from Fig. 6.4 to the results of Fig. 6.5, it can be seen that the severely damaged case has higher levels of accuracy compared to the moderately damaged case, reflecting the expected results. When comparing the performance of the ARMA-based DSFs, it can be seen that DSF1 and DSF2 preform very closely, regardless of the noise level, with DSF3 trailing behind slightly. Lastly, it can be seen for relatively low amount of noise (SNR of 30), all damage detection methods are capable of detecting damage to the same level.

These results illustrate the robustness of the Cosh spectral distance DSF. Although it is more computationally demanding, this method is capable of estimating damage more accurately at higher levels of noise, correctly predicting damage in above 80% of simulations at a SNR of 5 dB, which is representative of 30% added noise. When compared to damage metrics based solely on the coefficients of the ARMA model, the Cosh spectral distancing method accurately identifies damage approximately in 50% more simulations than its alternative at an SNR of 5 dB. In application, if a sensor concluded there was damage within a physical structure, it should prompt an investigation into the structure to confirm the data-based damage decision. If more than one imbedded sensor was alerting that damage was introduced to a structure, it could prevent the need for manual inspections on a time-basis and move toward using these sensor networks to prompt the need for structure maintenance.
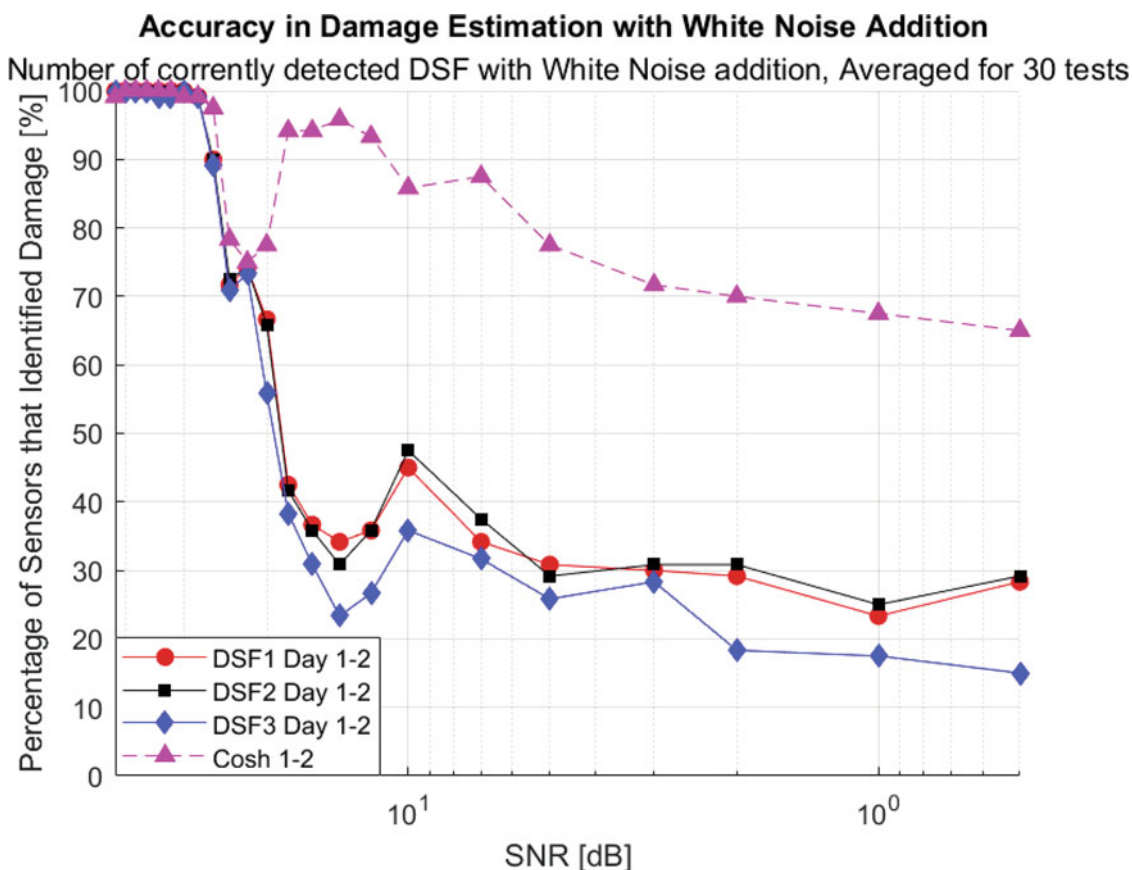
Fig. 6.4 Accuracy of damage-sensitive feature calculations comparing the healthy state to a moderately damaged state with increased amount of signal noise
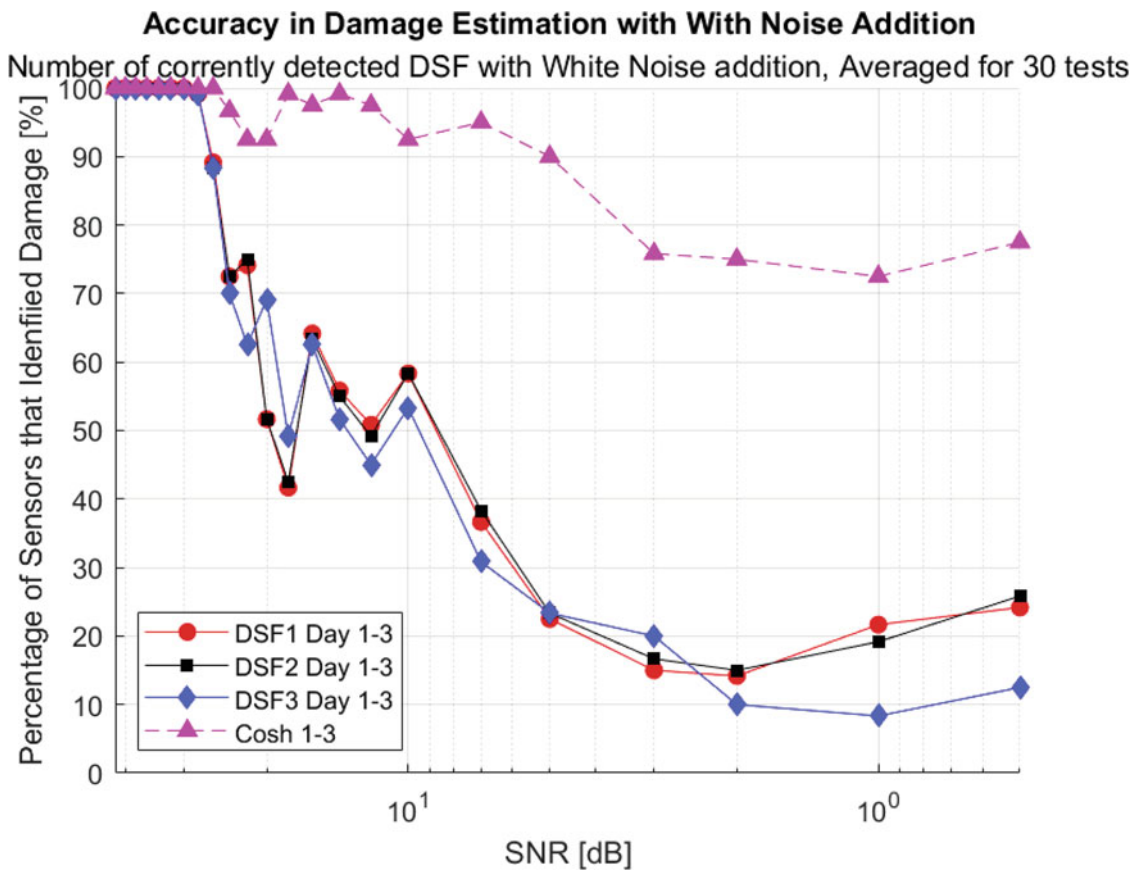
Fig. 6.5 Accuracy of damage-sensitive feature calculations comparing healthy state to a severely damaged state with increased amount of signal noise

6.5 Conclusion

In a case study involving a four degree-of-freedom cantilever beam, four ARMA-based damage-sensitive features were evaluated for their sensitivity to noise. The results of this research demonstrate the strength of the Cosh spectral distancing method as a damage-sensitive feature. For all levels of added noise, the Cosh spectral distancing method outperformed other ARMA-based DSFs. The results from Figs. 6.4 and 6.5 highlight this conclusion in which the Cosh spectral distancing method detects damage in roughly 80–90% of simulations compared to the 20–30% by the ARMA coefficient-based model, when 5 dB of noise has been added to the signal. To confirm the validity of this model, a Monte Carlo simulation was carried out varying the noise seed to remove the random aspect of noise and determine the long-run damage accuracy at each SNR.

This research serves to guide the implementation of damage detection methods within the field of structural health monitoring. By knowing the amount of noise that is added to a signal, a confidence interval in the damage decisions can arise, alerting the true state of structural health within a system. Additionally, the performance of the Cosh spectral distancing method highlights that this method should be used to make damage estimations, due to the robustness of its accuracy at high levels of noise. This system is hopefully implementable in physical structures such as bridges or buildings to monitor and predict when damage has been introduced, prompting maintenance or in severe cases, shut down of these structures. To further improve on this model, the addition of other types of noise such as gray or brown noise or a combination of these sources of noise would aid the realistic simulation of real-world sources of noise. Modifying the model order of the ARMA equation additionally provides a difference in results of the damage accuracy estimation. If the model order were increased, more accurate results would be expected as the data is in theory better fit with a higher model order equation. Further research will include a comparison of SNR as well as the ARMA model in order to determine which combination of the variables produces the most accurate results.

References

1. Nair, K.K., Kiremidjian, A.S., Law, K.H.: Time series-based damage detection and localization algorithm with application to the ASCE benchmark structure. J. Sound Vib. **291**(1–2), 349–368 (2006)
2. Catbas, F.N., Gul, M., Burkett, J.L.: Conceptual damage-sensitive features for structural health monitoring: laboratory and field demonstrations. Mech. Syst. Signal Process. **22**(7), 1650–1669 (2008). https://doi.org/10.1016/J.YMSSP.2008.03.005
3. dos Santos, A.D.F.: Output-Only Methods for Damage Identification in Structural Health Monitoring, Universidade Federal do Pará p. 164 (2017). https://repositorio.ufpa.br/jspui/handle/2011/9076
4. Roy, K., Bhattacharya, B., Ray-Chaudhuri, S.: ARX model-based damage sensitive features for structural damage localization using output-only measurements. J. Sound Vib. **349**, 99–122 (2015). https://doi.org/10.1016/j.jsv.2015.03.038
5. Yao, R., Pakzad, S.N.: Damage and noise sensitivity evaluation of autoregressive features extracted from structure vibration. Smart Mater. Struct. **23** (2014). https://doi.org/10.1088/0964-1726/23/2/025007
6. Mendler, A., Döhler, M., Grosse, C.U.: Selection of damage-sensitive features based on probability of detection curves. In: IOMAC 2022-9th International Operational Modal Analysis Conference, Vancouver, Canada (2022)
7. Lynch, J.P., Loh, K.J.: A summary review of wireless sensors and sensor networks for structural health monitoring. Shock Vibr. Dig. **38**(2), 91+ (2006). https://link.gale.com/apps/doc/A142047382/AONE?u=west10360&sid=bookmark-AONE&xid=803d5474
8. Yao, R., Pakzad, S.N.: Autoregressive statistical pattern recognition algorithms for damage detection in civil structures. Mech. Syst. Signal Process. **31**, 355–368 (2012). https://doi.org/10.1016/j.ymssp.2012.02.014

Chapter 7
Bayesian Model Updating for System and Damage Identification of Bridges Using Synthetic and Field Test Data

Niloofar Malekghaini, Farid Ghahari, Hamed Ebrahimian, Vinayak Sachidanandam, Eric Ahlberg, Matthew Bowers, and Ertugrul Taciroglu

Abstract The finite element (FE) models of bridges are vastly used for structural analysis. However, these initial models – developed from as-built drawings – cannot adequately present the real-world bridges due to the inherent modeling uncertainties, irregularities during construction, or aging. The uncertain model parameters can be estimated using Bayesian model updating techniques wherein the initial model is updated using measured responses. The updated finite element model can be used for structural health monitoring and damage diagnosis. This study presents a new framework for operational monitoring and damage diagnosis of bridges through the integration of finite element models with bridge vibration responses and vehicle tracking data using a Bayesian FE model updating method. First, the framework is verified in a simulation environment via synthetic data obtained from a finite element model of the San Roque Canyon Bridge located in Santa Barbara, California. Then, the proposed method is employed using real-world data collected from a pair of full-scale girders at the Turner-Fairbank Highway Research Center in McLean, Virginia. The performance of the employed approach is evaluated through a comparison of the estimated model and previously observed damage in the structure. Taken together, the results demonstrate the capabilities of the proposed framework to diagnose potential damages in real-world bridge structures.

Keywords Bridge · Structural health monitoring · Operational monitoring · Bayesian inference · Finite element model updating

7.1 Introduction

The growing number of aged in-service bridges necessitates the development of new technologies for operational health monitoring of these structures [1]. Non-destructive testing methods are widely used for structural damage identification, but there are limitations in their applications [2–4]. To overcome these limitations, the Bayesian finite element (FE) model updating in time domain is introduced in literature and is mainly used for damage identification of building structures subjected to earthquake to jointly estimate the model parameters and input acceleration time histories [5–7]. This study extends the application of the available Bayesian FE model updating methods to develop a new damage identification framework for bridges under operational conditions. In this framework, first, the known location of vehicles (obtained using computer vision techniques) and bridge measured responses are synchronized and integrated. Then, this information is used to jointly estimate the vehicular loads and the FE model parameters. The updated FE model is used to identify damage in the bridge. Development of formulations, verification, and validation of the proposed framework are the novelties of this work.

N. Malekghaini (✉) · H. Ebrahimian
Department of Civil and Environmental Engineering, University of Nevada, Reno, NV, USA
e-mail: nmalekghaini@nevada.unr.edu

F. Ghahari · E. Ahlberg · E. Taciroglu
Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA

V. Sachidanandam · M. Bowers
SC Solutions Inc., Sunnyvale, CA, USA

7.2   Formulation

The proposed framework and Bayesian inference are schematically shown in Fig. 7.1a and b, respectively. As can be seen in Fig. 7.1a, the measured responses of the bridge and the tracked location of vehicles are the input to the Bayesian inference to update the mechanics-based model of the bridge. In this process, the unknown model parameters as well as the load of vehicles, presented as θ in Fig. 7.1b, are treated as random variables. The uncertainties of these parameters are characterized by a joint Probability Density Function (PDF) – shown as $\mathbf{P_\theta}$. The prior uncertainties in unknown parameters, including the mean and vector $\hat{\boldsymbol\theta}^-$ and covariance matrix $\hat{\mathbf{P}}_\theta^-$, are propagated through the FE model of the bridge and stochastic FE-predicted responses, $\hat{\mathbf{Y}}(\theta)$, are estimated. Minimizing the discrepancies between $\mathbf{Y}$ (measured responses) and $\hat{\mathbf{Y}}(\theta)$ using Bayes' theorem results in the posterior estimates of mean vector and covariance matrix, known as $\hat{\boldsymbol\theta}^+$ and $\hat{\mathbf{P}}_\theta^+$. In this study, the rolling estimation method is employed in which the time domain is divided into n_s windows while the sth window is defined between time step $t_0 = 1$ and t_s. The above process is iterated at each estimation window until convergence of unknown parameters. Then, the estimation window rolls out in time and the process is repeated using the new batch of data. It is noteworthy that the number of unknown parameters varies between different estimation windows as the number of vehicles traversing on the bridge varies among the estimation windows. The updated model can be used for damage identification and decision-making related to asset management.

7.3   Verification

The verification study is performed through damage identification of a numerically simulated damage scenario. For this purpose, a mechanics-based nonlinear FE model of a prestressed box girder bridge is developed in OpenSees [8]. The identifiable model parameters and location for measurement channels are determined using an information-theoretic approach for identifiability analysis [9]. A damage scenario – including concrete delamination, reinforcement corrosion, concrete degradation, and tendon corrosion in five regions along the length of the bridge – is introduced to the bridge. The introduced damage state is shown in Fig. 7.2a and the corresponding color code is stated in Fig. 7.2b. A random traffic is generated on the bridge and the bridge acceleration responses are collected. Then, the framework of Fig. 7.1b is implemented to estimate the unknown model parameters – including concrete compressive strength and prestressing force – and load of vehicles. The final estimates of unknown model parameters are used to infer damage state of the bridge. The identified damage state of the bridge is shown in Fig. 7.2c and is similar to the true damage state. In addition to the correct estimation of location and severity of damage, the vehicular loads are estimated with a maximum of 11% error. For the sake of presentation, the updating process of the unknown parameters is not shown here.
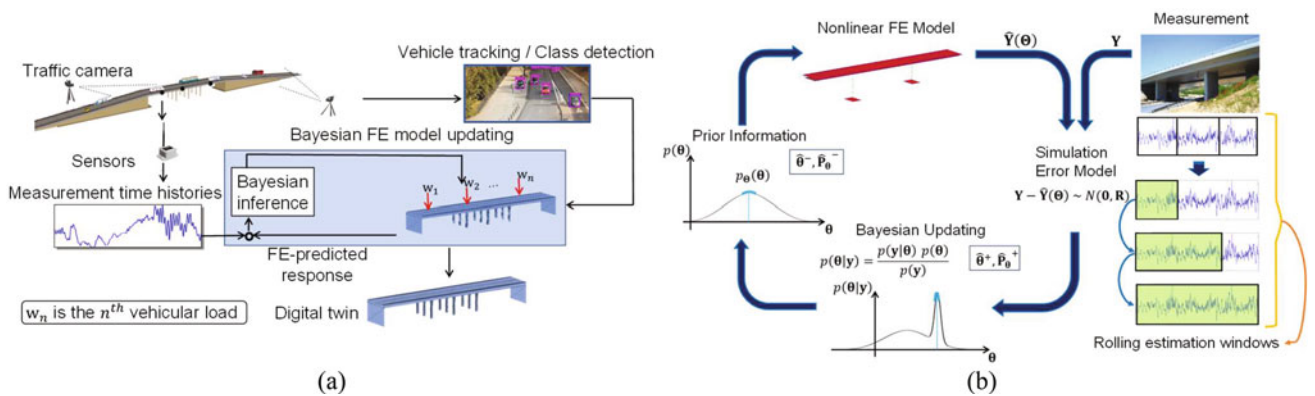
Fig. 7.1 Proposed approach: (**a**) The overall view of the framework, and (**b**) Bayesian inference
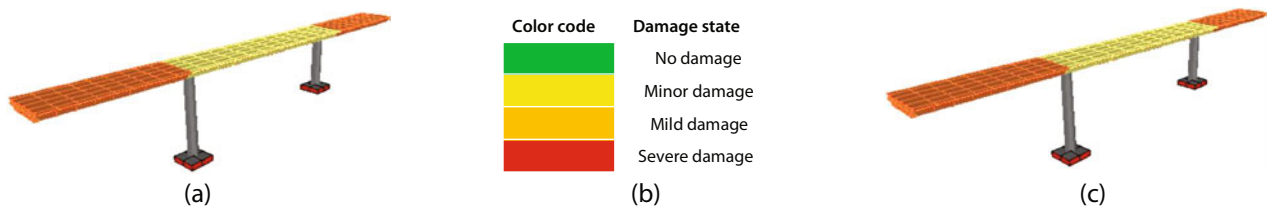
Fig. 7.2 Verification study: (**a**) True damage state, (**b**) Color code for damage states, and (**c**) Estimated damage state
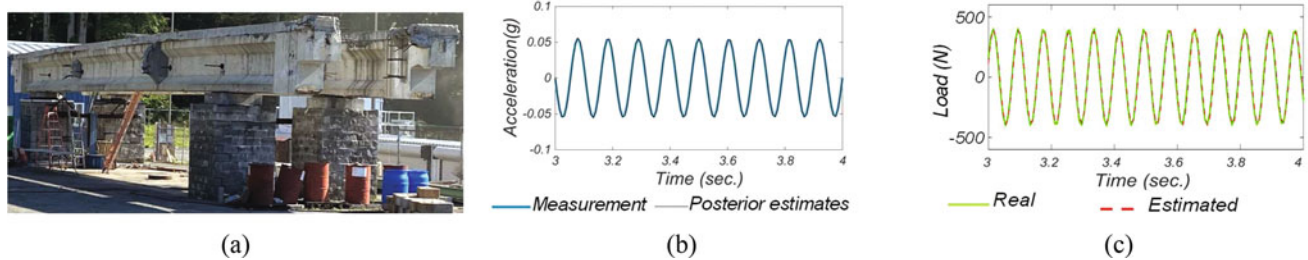
Fig. 7.3 Verification of output-only Bayesian inference: (**a**) Testbed structure, (**b**) Estimation of measurement, and (**c**) Estimation of input load

7.4 Validation

Validation is carried out in two steps. First, the applicability of the output-only Bayesian model updating technique in a real-world setting is examined. For this purpose, a pair of decommissioned prestressed I-girders excited by forced vibrations was used as a testbed. The measured acceleration responses of the girders were used for joint estimation of unknown model parameters – including concrete compressive strength and damping parameters – and input load. The testbed structure is presented in Fig. 7.3a. A comparison between updated FE-predicted responses and measurements, and a comparison between estimates of input load and the measured input load are shown in Fig. 7.3b, c. The final estimates of the concrete compressive strength were used to infer damage in the studied girders. The results confirmed the applicability of output-only Bayesian model updating technique in a real-world setting.

The next step to validate the proposed framework includes application to real-world bridges, wherein tri-axle dump trucks traveling at various speeds and traffic configurations are utilized to induce bridge acceleration responses. The results are not presented here for brevity and are the subject of future publications.

7.5 Conclusions

In this study, a new Bayesian model updating framework for the identification of damage in bridges under operational conditions was proposed. The proposed framework was first verified in a numerically simulated environment. A nonlinear FE model of a prestressed box girder bridge was developed in OpenSees. A damage scenario was introduced to the bridge model. A random traffic scenario was run on the bridge and the bridge's acceleration response was simulated. Using the developed Bayesian model updating framework, the damage state of the bridge was estimated. The strong correlation between the true and estimated damage state verified the feasibility of the framework. Then, the output-only Bayesian inference was validated in a real-world setting for damage identification of a pair of decommissioned prestressed I-girders. The results showed the applicability of the technique in a real-world setting.

References

1. ASCE: ASCE's 2021 Infrastructure Report Card. https://infrastructurereportcard.org/cat-item/bridges/ (2021)
2. Nair, A., Cai, C.S.: Acoustic emission monitoring of bridges: review and case studies. Eng. Struct. **32**, 1704–1714 (2010). https://doi.org/10.1016/j.engstruct.2010.02.020
3. McCann, D.M., Forde, M.C.: Review of NDT methods in the assessment of concrete and masonry structures. NDT E Int. **34**, 71–84 (2001). https://doi.org/10.1016/S0963-8695(00)00032-3
4. Scott, M., Rezaizadeh, A., Delahaza, A., Santos, C.G., Moore, M., Graybeal, B., Washer, G.: A comparison of nondestructive evaluation methods for bridge deck assessment. NDT E Int. **36**, 245–255 (2003). https://doi.org/10.1016/S0963-8695(02)00061-0
5. Ebrahimian, H., Astroza, R., Conte, J.P., Papadimitriou, C.: Bayesian optimal estimation for output-only nonlinear system and damage identification of civil structures. Struct. Control Health Monit. **25**, 1–32 (2018). https://doi.org/10.1002/stc.2128
6. Ebrahimian, H., Kohler, M., Massari, A., Asimaki, D.: Parametric estimation of dispersive viscoelastic layered media with application to structural health monitoring. Soil Dyn. Earthq. Eng. **105**, 204–223 (2018). https://doi.org/10.1016/j.soildyn.2017.10.017
7. Ghahari, F., Malekghaini, N., Ebrahimian, H., Taciroglu, E.: Bridge digital twinning using an output-only Bayesian model updating method and recorded seismic measurements. Sensors. **22** (2022). https://doi.org/10.3390/s22031278
8. McKenna, F.: OpenSees: a framework for earthquake engineering simulation. Comput. Sci. Eng. **13**, 58–66 (2011). https://doi.org/10.1109/MCSE.2011.66
9. Ebrahimian, H., Astroza, R., Conte, J.P., Bitmead, R.R.: Information-theoretic approach for identifiability assessment of nonlinear structural finite-element models. J. Eng. Mech. **145**, 1–14 (2019). https://doi.org/10.1061/(ASCE)EM.1943-7889.0001590

Chapter 8
Static and Dynamic Characterization of a Vibration Decoupling Element Based on a Metamaterial Structure

Alessandro Annessi, Valentina Zega, Paolo Chiariotti, Milena Martarelli, and Paolo Castellini

Abstract When targeting vibration isolation, metamaterials do represent very powerful solutions, given the extreme design flexibility that can be introduced. However, dealing with metamaterials ensuring wide bandgaps in vibration stop-band filter configuration is highly challenging. This chapter discusses the static and dynamic characterization of a decoupling mechanical system designed to act as a metamaterial stop-band filter in the 1500–15000 Hz frequency range. The system is produced in PA2200 polyamide powder by exploiting Selecting Laser Sintering technology. The external envelope of the system is a parallelepiped of $4 \times 4 \times 12$ cm. Both contact and noncontact vibration approaches have been investigated to properly identify the effective bandwidth of the filter and the eventual structural resonances that might spoil its filtering capabilities in the target frequency range. The experimental data is used to update the linear numerical model exploited for the structure design.

Keywords Metamaterial · Mode separation · Laser Doppler vibrometry · Modal analysis · Vibration suppression

8.1 Introduction

A metamaterial is a conventional material (e.g., nylon or steel) characterized by specific properties concerning wave propagation brought on by its peculiar geometry, which is mostly but not necessarily periodic. Due to this geometrical complexity, metamaterials are usually manufactured by additive manufacturing technology, such as stereolithography. Vibrational metamaterials are emerging as a valid and feasible alternative to traditional materials (e.g., foams, elastomers, rubbers) for vibration isolation and absorption due to their extreme versatility and high attenuation capability. Nowadays, new metamaterial designs are emerging constantly in the field of passive vibration suppression and isolation, aiming at better performances [1–6]. Lu et al. used a nonlinear resonator consisting of a rolling-ball sliding in a spherical cavity together with a magnetic energy harvester for simultaneous vibration isolation and energy production [7]. Essink et al. designed a three-dimensional metamaterial capable of passive vibration suppression in three-dimensional excitation (longitudinal, transverse, and torsional), verifying this behavior experimentally [3]. Srivatsa et al. studied the correlation between the natural frequency characterizing tuned mass-dampers of a metamaterial and any observed local stress concentrations, which could reveal fatigue issues decreasing the metamaterial lifespan [8]. Liu et al. developed a zero Poisson's ratio metamaterial for shock absorption using shape memory polymer [9].

In a recent work [10], we have proposed a compact metamaterial design for vibration isolation, bearing in mind applications in vibration testing. Initially, the unit cell design is presented. The unit cell (UC) is planned to be as compact as possible by filling a cubic unit cell with 4 cm side. Half cube corners work as a mass connected to the center of the UC by cylinders. The remaining ones are linked to the UC center by cylindric beams embedding spring-like behavior.

A. Annessi (✉) · M. Martarelli · P. Castellini
Department of Industrial Engineering and Mathematical Sciences, Università Politecnica delle Marche, Ancona, Italy
e-mail: a.annessi@pm.univpm.it; m.martarelli@univpm.it; p.castellini@univpm.it

V. Zega
Department of Civil and Environmental Engineering, Politecnico di Milano, Milano, Italy
e-mail: valentina.zega@polimi.it

P. Chiariotti
Department of Mechanical Engineering, Politecnico di Milano, Milano, Italy
e-mail: paolo.chiariotti@polimi.it
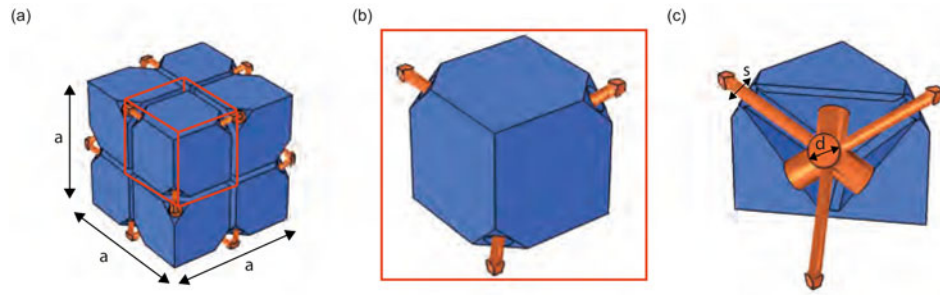
Fig. 8.1 (**a**) Schematic view of the unit cell. (**b**) One-eighth of the unit cell. (**c**) Sectional view of one-eighth of the unit cell. Rigid masses are colored in blue while deformable connections are in orange [10]

Table 8.1 Material properties for fine polammide PA 2200 for EOSINT P SLS printers

PA 2200			
Particle dimension	Laser diffraction	60	μm
Density	Laser sinterized	0.9-0.95	g/cm^3
Tensile modulus	DIN EN ISO 527	1700 ± 150	N/mm^2
Shear modulus	DIN EN ISO 527	1240 ± 130	N/mm^2
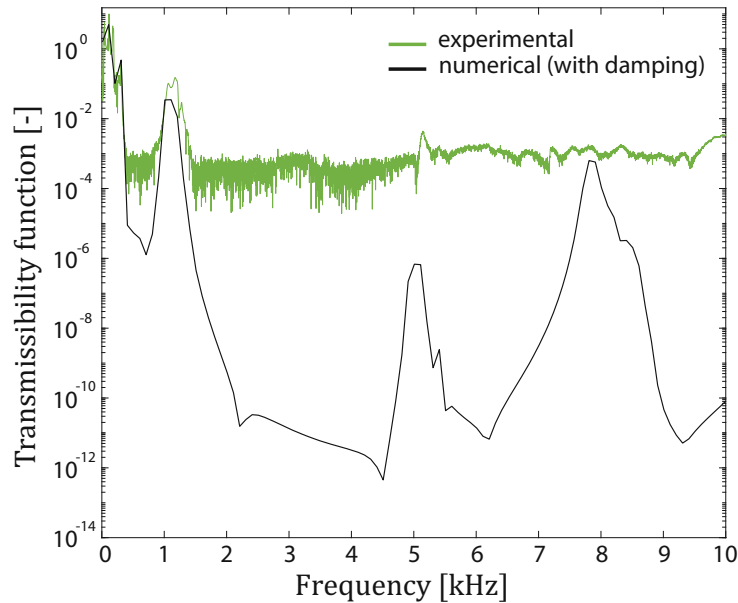
Fig. 8.2 Transmission diagram: comparison between numerical predictions (black line) and experiments (blue and green lines) [10]

The UC geometry is depicted in Fig. 8.1a whilst its internal complex structure is shown in Fig. 8.1b. The material chosen for prototyping is the PA 2200, a typical polymeric material employed in additive manufacturing processes. In Table 8.1, material properties available from the manufacturer are summarized for the sake of clarity. The Poisson ratio is evaluated as $v = E/2G - 1 = 0.31452$. Thus, longitudinal wave speed can be computed as $v = \sqrt{\frac{E}{\rho}} = 1374$ m/s.

By assuming infinite periodicity and choosing a unit cell size of 4 cm, the metamaterial behaves as a mechanical stop-band filter in the frequency range between 1478.1 Hz and 15239 Hz (these are the frequencies related to the opening and closing modes of the bandgap).

Hereafter, a metastructure built as a $1 \times 1 \times 3$ repetition of the unit cell is numerically studied, though with a simplified model, and experimentally validated in terms of its attenuation capability by means of transmissibility function estimation, as shown in Fig. 8.2.

8.2   Simplified Preliminary Numerical Analysis

The finite metastructure is composed of the repetition of three unit cells. The external dimension of the system is $4 \times 4 \times 12$ cm. Since the prototype sample has to be made through additive manufacturing, it is important to check that there are no problems at static level due to self-weight. A simplified numerical model has been developed for the purpose. The model exploits linear elastic formulation, even though this might not be the condition best matching with the real-world scenario. Authors are aware of this extreme simplification of the numerical problem, and they are working on a more complex model that could be used for predictive analysis of the metamaterial static and dynamic behavior. Nevertheless, the approach presented in this chapter holds for the purpose of easing the preparation of the model and speeding up the calculation process. A linear static analysis is then performed in COMSOL Multiphysics© to simulate self-weight-induced deformations of the metastructure. The base surface is clamped to reproduce operating conditions, while self-weight is simulated through a volume load of amplitude 1g applied on the full geometry. A compression of 0.13 mm is computed when the axial self-weight is applied (Fig. 8.3) while a bending deflection of 3.36 mm is obtained for the transversal gravity load (Fig. 8.4).

A finite element-based modal analysis has been performed by means of COMSOL Multiphysics©, trying to reproduce, even though in a simplified and ideal manner, the constraints imposed on the metastructure for the experimental evaluations. The base of the component is fixed, that is, the metastructure is clamped on the bottom surfaces of the first layer of masses (mimicking double adhesive tape) whilst the rest of the structure is free to vibrate. Computed mode shapes are always in pairs due to the symmetry of the metastructure. In Fig. 8.5, the modal shape functions of the first 12 modes of the metastructure under study are reported. The contour of the normalized displacement field is shown in color for the sake of clarity.

Fig. 8.3 Numerical displacement due to the gravity along the axial direction
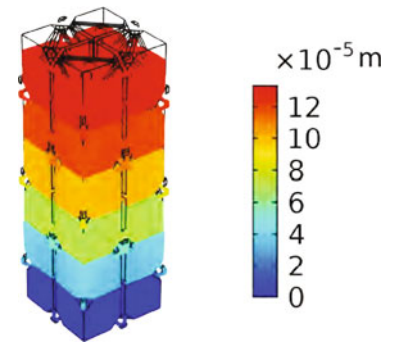
Fig. 8.4 Numerical displacement due to the gravity along the load direction

Fig. 8.5 Examples of mode shapes retrieved from finite element modal analysis .

10.411 Hz	53.595 Hz	118.75 Hz	134.26 Hz
174.75 Hz	198.22 Hz	209.24 Hz	230.69 Hz
281.04 Hz	869.04 Hz	1116.2 Hz	1467.6 Hz
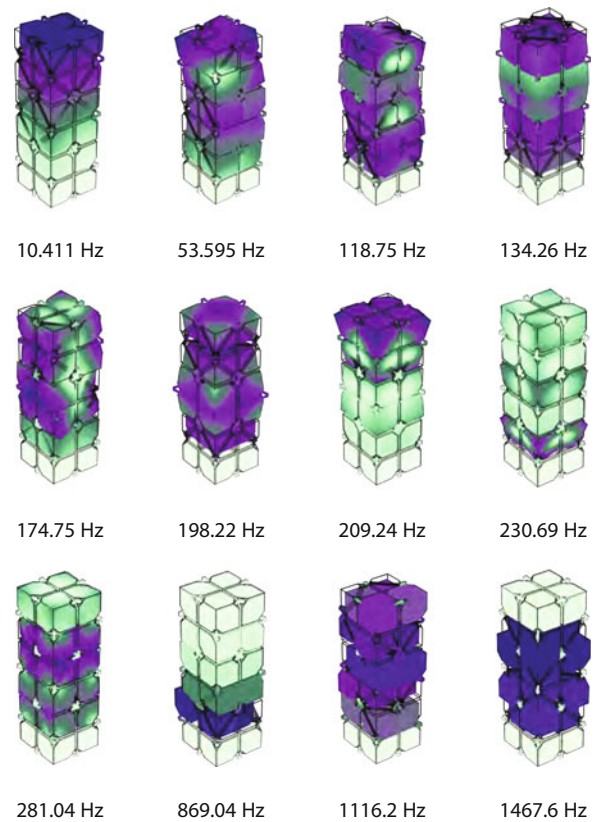
8.3 Experimental Analysis

The metastructure static behavior is characterized by compression testing. Reaction forces at different imposed displacement values are measured during the test. Experiments have been carried out using an MTS810 material testing system, equipped with an MTS force transducer model $661.22D - 01$ with a load capacity of 250 kN. During the test, the metastructure has undergone one load/unload cycle, the displacement speed was equal to 0.01 mm/s. Test results are shown in Fig. 8.6. The evident slope changing around 8 mm of displacement is due to the bundling of the masses composing the metastructure. In fact, in the interval with lower slope, only the slanting beams take charge of the deformation, while the masses behave as rigid elements. Increasing the displacement, the masses come into contact and thus the prototype starts to work as a bulk structure. Moreover, hysteresis is visible in the force–displacement curves. This is caused by the intrinsic viscoelastic behavior of the polymeric constitutive material.

The dynamic behavior of the metastructure is then experimentally assessed using scanning laser Doppler vibrometry (SLDV), which allows to perform contact-less experimental evaluation of the operational deflection shapes (ODS) of a target structure. ODS have been measured by performing a scanning on the side and on the top of the metastructure. The test setup is presented in Fig. 8.7. The metastructure is mounted with double adhesive tape on the plate of the electrodynamic shaker (Data Physics V20), to experimentally recreate the same constraint used for numeric simulations and to allow proper excitation. A miniaturized accelerometer (PCB model 352C23 with a sensitivity equal to 5.39 mV/g) is then mounted on the electrodynamic shaker plate (bee-wax connection, to capture the shaker table acceleration) as reference signal to evaluate a velocity/acceleration transmission function, which would be namely m/s (velocity measured by the SLDV) over m/s^2 (acceleration measured by the reference accelerometer). To scan the top and the side of the structure, a Polytec OFV 055 scanning head with the relative OFV 3001 S controller is used; acquisition parameters are listed in Table 8.2.

A custom grid is drawn for both top and side faces of the metastructure. In the former case, a grid of 248 points, shown in Fig. 8.8a, is used. In the latter case, 405 points are used in the grid shown in Fig. 8.8b (spatial grid resolution is adequate to solve mode shapes of interest). Lastly, the electrodynamic shaker excites the structure in the band 2 Hz–20 kHz. Transmissibility functions between the vibrometer response (velocity) and the reference accelerometer (acceleration) are saved as well as spectra. Even if the transmissibility function is not a frequency response function (we do not measure the input force, but only the acceleration at the shaker plate), modal analysis can still be performed. The aim of this experimental
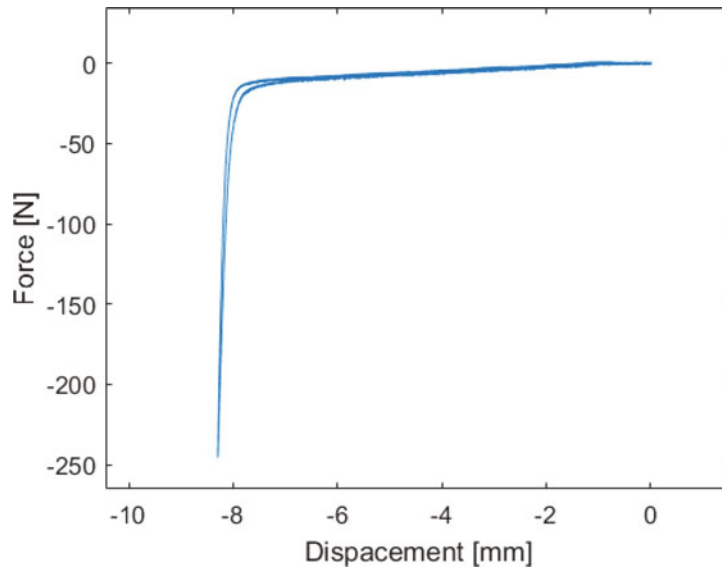
Fig. 8.6 Compression test results using MTS810 testing machine
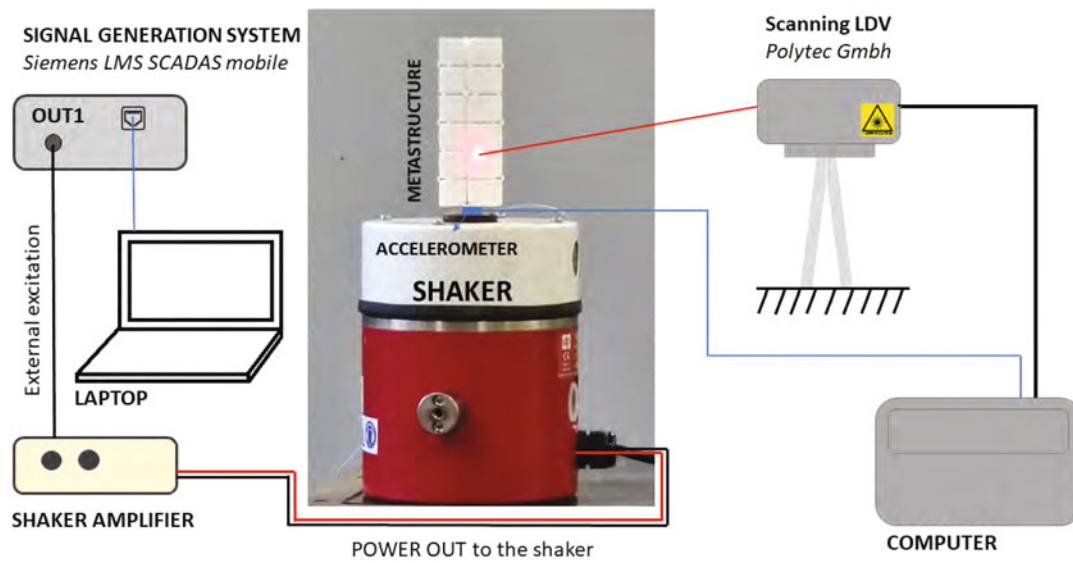
Fig. 8.7 Experimental setup for experimental modal analysis with scanning laser Doppler vibrometer and reference accelerometer

Table 8.2 Scanning laser Doppler vibrometer equipment used in the measurement campaign

Vibrometer	Polytech OFV 055
Controller	OFV 3001
Sensivity	$5 \frac{mm/s}{V}$
Tracking filter	Fast
Averages	16
Bandwidth	5000 Hz
Freq. res.	0.78125 Hz
Spectral lines	6400
Windowing	Hanning

(a)   (b)
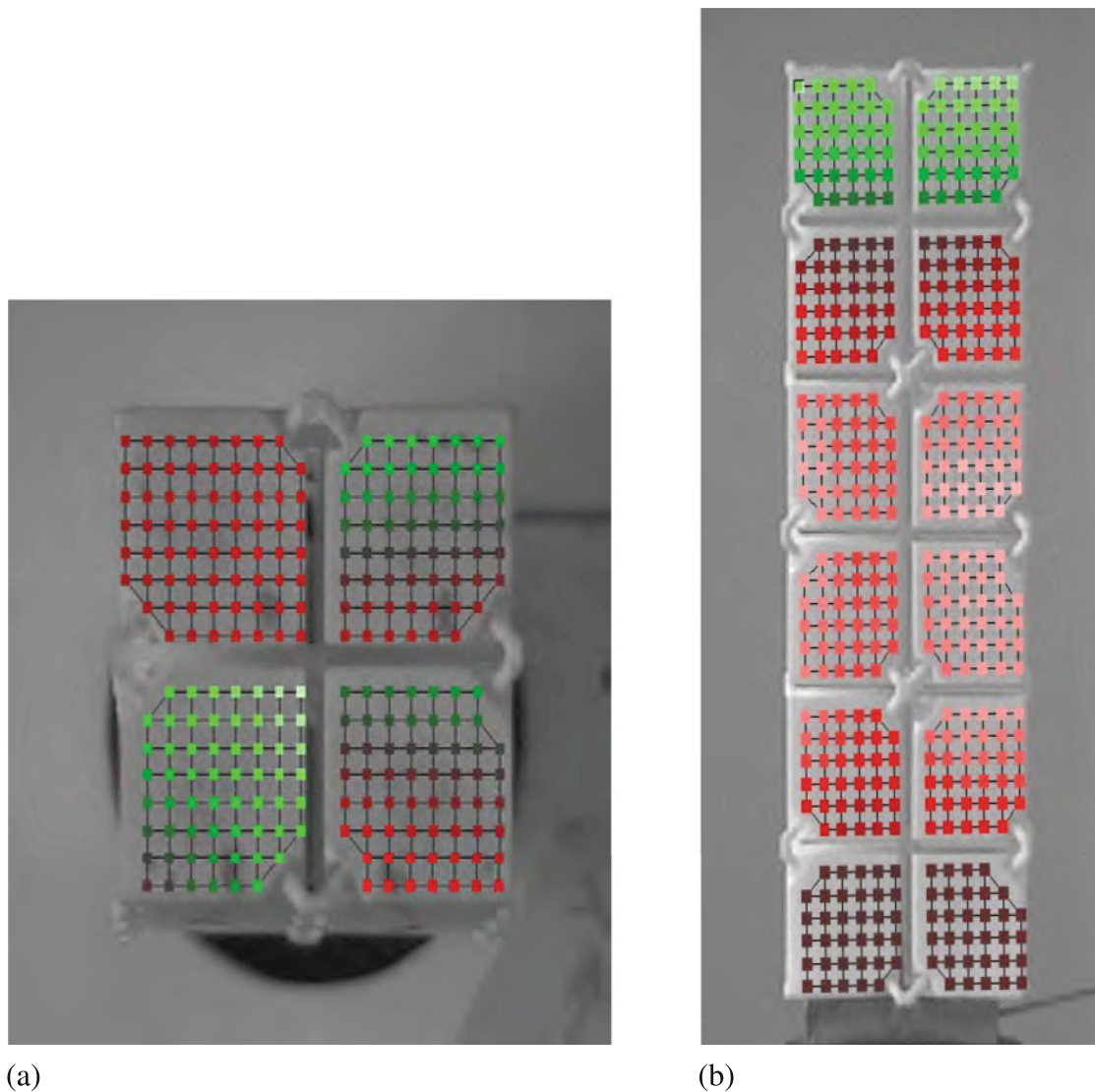
Fig. 8.8 Scanning grid for laser Doppler vibrometer measurement campaign. (**a**) Scanning grid of the metastructure top face composed of 248 points. Color scale referred to preliminary peak picking analysis, showing one of the computed operational deflection shapes. (**b**) Scanning grid of the metastructure side face composed of 405 points. Color scale referred to preliminary peak picking analysis, showing one of the computed operational deflection shapes

campaign is to assess the dynamic behavior of the metastructure under known conditions, without focusing on magnitudes but on the extraction of modal parameters. The PolyMAX algorithm embedded in SIEMENS Simcenter TestLab was used for the purpose. The main issue related to the use of a laser Doppler vibrometry is the amount of light scattered back that highly affects the velocity signal quality and hence its signal-to-noise ratio (SNR). In this specific case, the PA 2200 material, from which the metastructure is fabricated, is opaque. The laser beam impacting the metastructure surface is transmitted in the sample, leading to a low signal return, not sufficient for measuring. Therefore, a glass powder is glued on the target surfaces to enhance light backscatter and then velocity signal SNR. The added powder has a negligible mass compared to the metastructure itself.

Looking at the experimental transmissibility function of Fig. 8.2, the designed metastructure has a relevant amplitude only in the first (from 0 to approximately 310 Hz) and second (from 900 to approximately 1420 Hz) transmission bands due to its engineered design. The second transmission band has a lower magnitude with respect to the first. The high amplitude peak related to the first transmission band causes the saturation of the vibrometer demodulator, resulting in a strong noise base and hiding global modes of the second transmission band. Therefore, an additional analysis is carried out shrinking the excitation
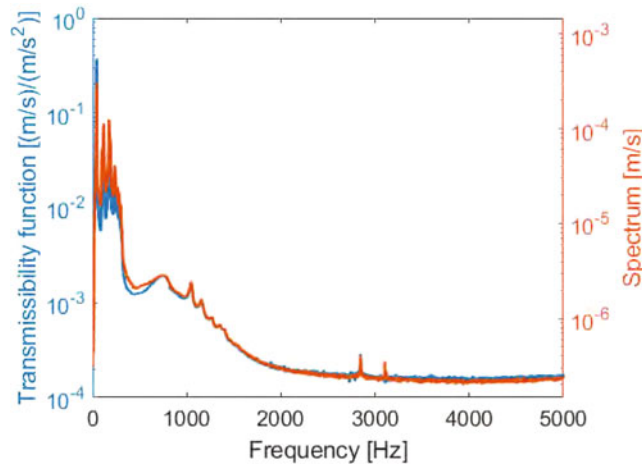
Fig. 8.9 Sum over whole transmissibility functions over average spectrum from vibrometry data for full bandwidth excitation

band to 614–10,240 Hz. This expedient allows us to excite properly and retrieve modes on the second transmission band in a better way.

It can be pointed out that we are performing modal analysis on a function that is a ratio between two response signals from the vibrometer and the accelerometer. A comparison between the amplitude of the sum of the transmissibility functions and the average velocity spectrum is shown in Fig. 8.9. There is no relevant frequency shift between the curves, suggesting a proper estimation of the structure's natural frequencies. It should be highlighted that the data reported in Fig. 8.9 have different measurement units and the scales in the figure have been properly chosen to make the two curves easily comparable. Modal analysis is then performed on transmissibility functions using the PolyMAX algorithm. Experimental mode shapes thus computed are reported in Fig. 8.10.

8.4 Young's Modulus Estimation

The viscoelastic behavior of the polymeric constitutive material is taken into account through a complex Young's modulus ($E = E_0(1 + i\tau\omega)$), where E_0 is the linear Young's modulus and τ is identified through a sensitivity analyses as 3 μs. The PA220 material characteristics specified by the manufacturer (Table 8.1) are used as a preliminary assumption on the linear Young's modulus E_0. However, since a static characterization of the metastructure was performed, the elastic modulus can be also retrieved from the force–displacement curve reported in Fig. 8.6. Fixing the Poisson ratio value, Young's modulus can be estimated from the linear fit of the compression data shown in Fig. 8.11 in the range between about −4.8 mm and −7.5 mm. This interval is chosen considering the data that is less noise-affected, before the change in slope due to the bundling of the metastructure.

The slope, that is, the metastructure stiffness, found out from the linear fit is 2.44 N/mm. To retrieve the Young's modulus, as a first approximation (i.e., not considering more complicated structural analytical models), a numerical compression test of the metastructure has been carried out with an E equal 1 MPa and a preload Δz of 1 mm. A coefficient relating the metastructure stiffness and the Young's modulus of the material, defined as k_E, can be computed from the surface reaction force $F_z = -0.0018$ N where displacement has been applied. Overall stiffness can be computed as $K = F_z/\Delta z = k_E E$, thus $E = 1352 \pm 32$ MPa (with 95% confidence interval on linear fit). To confirm the value retrieved from the former approximation, a numerical analysis of the displacement-controlled compression test is carried out, exploiting the same displacement range used for the linear fit of Fig. 8.11, virtually measuring the force as the reaction force on the upper boundary. Force–displacement curves for a spectrum of Young's modulus values are obtained with their relative slopes. For a stiffness value of 2.44 N/mm, retrieved experimentally, the associated numerical Young's modulus is included between 1350 and 1360 MPa, upholding the initial assumption. Therefore, the experimentally retrieved Young's modulus is approximately 20% smaller compared to the initial value from the datasheet.

Both numerical transmissibility functions, computed with Young's modulus from datasheet and the updated one, and experimental data are shown in Fig. 8.12. A much better match with experimental data is clear by observing the red curve plotted in Fig. 8.12b, which represents the transmission function obtained exploiting the Young's modulus from the
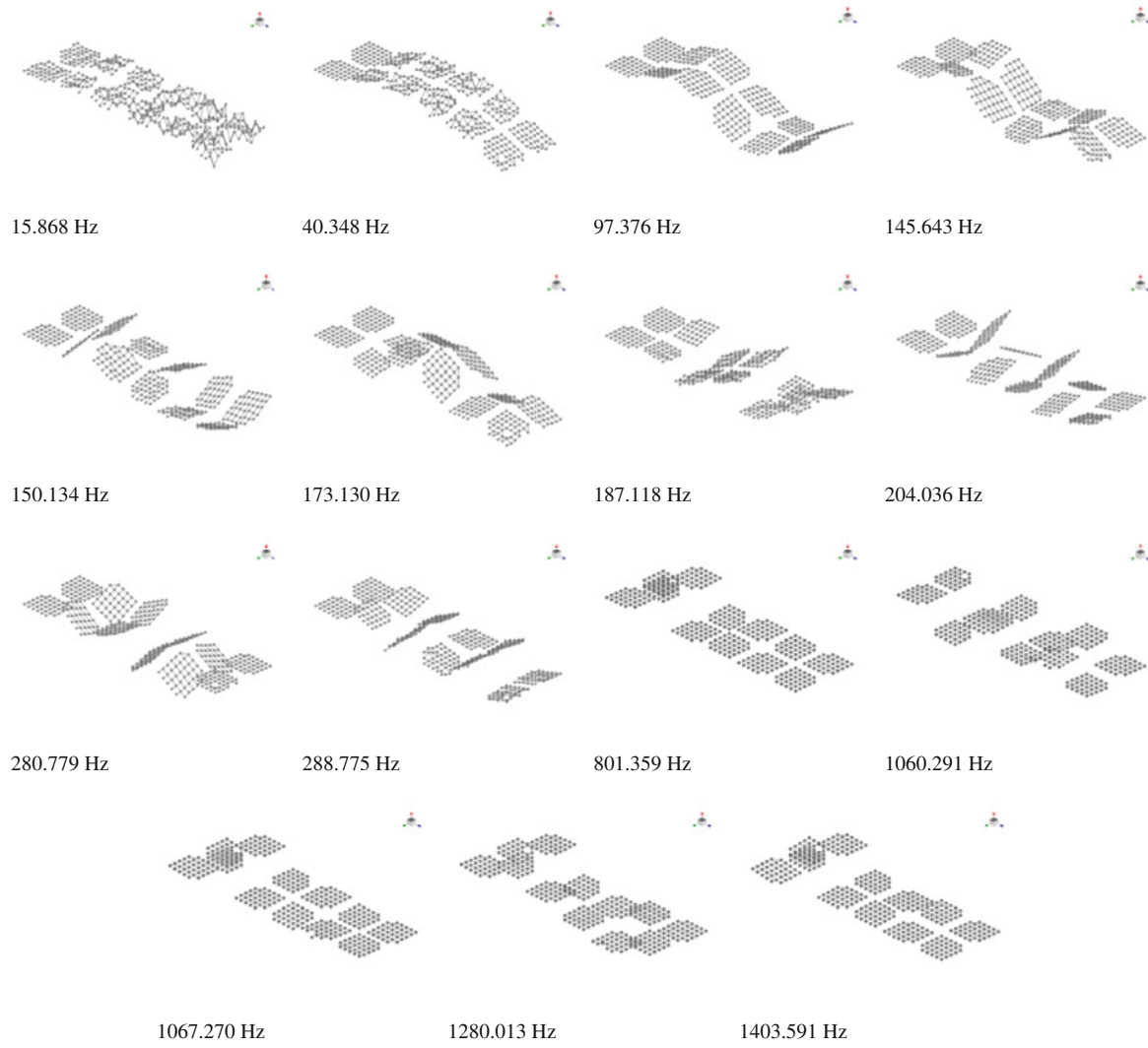
Fig. 8.10 Examples of mode shapes retrieved from experimental modal analysis
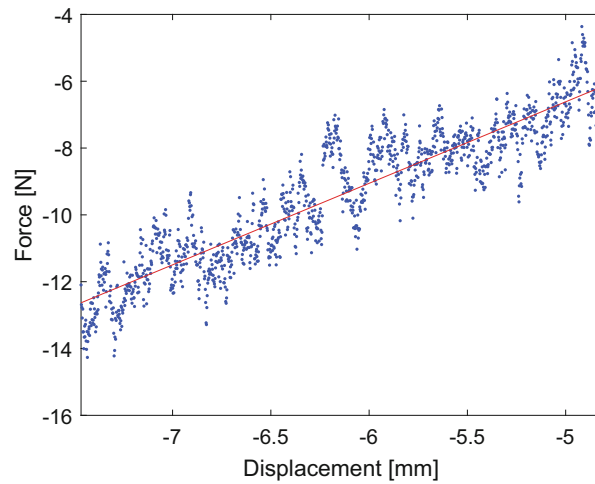
Fig. 8.11 Linear fit for the compression data before the metastructure gets full packed
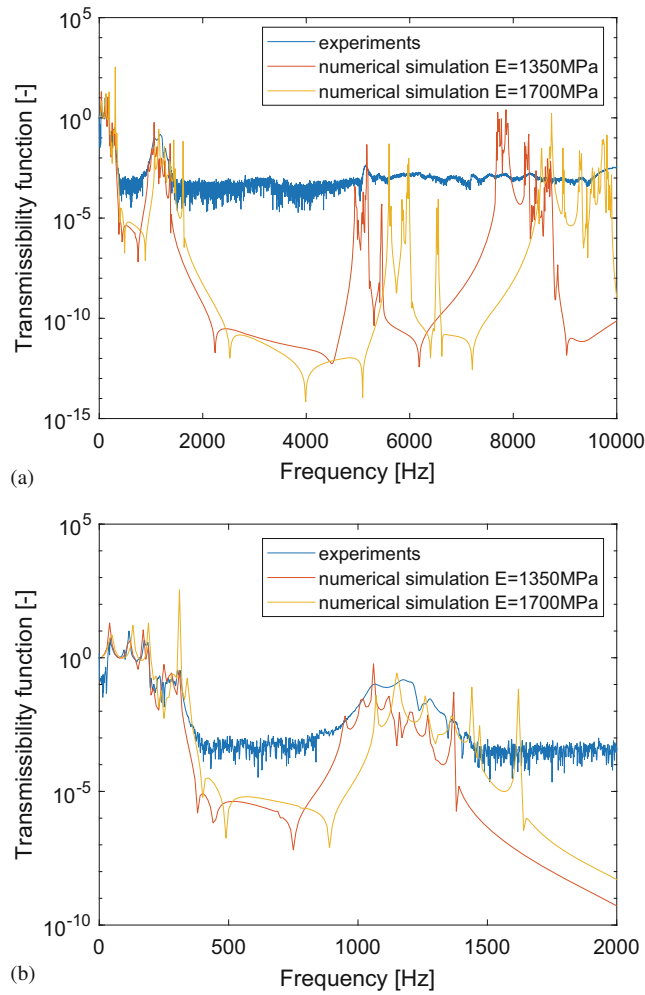
Fig. 8.12 Comparison between numerical and experimental transmissibility function after Young's modulus parameter updating. (**a**) Comparison between numerical and experimental transmissibility function in the whole frequency range. (**b**) Close up on the low frequency range

experimental load test by drawing attention to the metastructure transmission bands. Also, the small peak present at 5145 Hz in the experimental data of Fig. 8.12a, due to local modes of the metastructure represented in Fig. 8.13, is picked properly after updating the material properties.

Moreover, density can be retrieved from mass and volume of the first metastructure specimen. The mass value is 92.3 g with 0.01 g accuracy. To get the volume of the metastructure, the specimen tested was scanned via computed tomography, thus obtaining a dense point cloud representation of the structure through the MeshLab software. The volume estimated value is 90.73 cm^3. Therefore, the mass density for the selected sample is around 1 g/cm^3, which is in good agreement with the datasheet provided by the manufacturer. Performing the same procedure over a set of 10 specimens and averaging the resulting values, we obtain an average density of 0.97 g/cm^3 with a standard deviation of 0.02 g/cm^3. Uncertainties in density estimation are related to the presence of dust remaining inside the samples. This causes primarily the presence of an additional weight to be considered. Secondly, it creates issues when the point cloud is imported into MeshLab open-source software. In fact, surfaces defining the edges of the metamaterial are not sharp, resulting in a less-than-optimal volume reconstruction and estimation. Numerical modal analysis is then computed considering the updated material properties (E = 1350 MPa and G = 984.71 MPa, retrieved from ν previously estimated). Numerical and experimental mode shapes are compared visually, and natural frequency difference (NFD) is carried out. In Fig. 8.14, a linear fit with intercept passing through the origin is shown facing numerical and experimental resonance frequencies for the manufacturer and the updated Young's modulus. Slopes resulting from both cases are respectively equal to 1.0531 and 0.9406, which are 5.31% and 5.94% off the diagonal. The linear fit for both cases oscillates near the diagonal and is similar in terms of NFD. The corrected Young's modulus is too small, resulting in a trend line with a slope less than the diagonal, obtaining a less stiff material than it should be. The reason behind this behavior is related to the linearity assumption on the material behavior and the Young's modulus retrieving
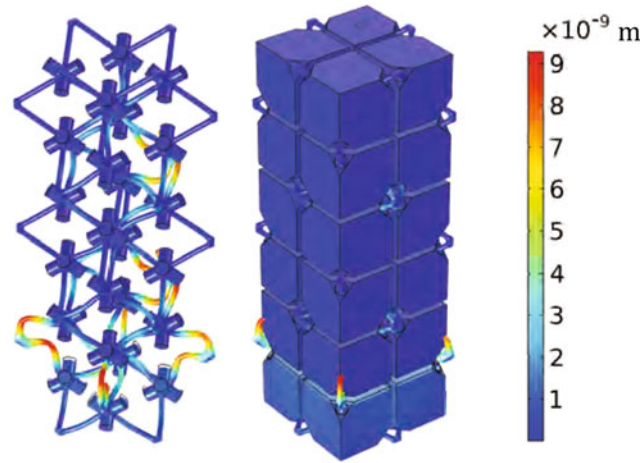
Fig. 8.13 Numeric displacement field for the resonance peak at 5145 Hz after Young's modulus parameter updating
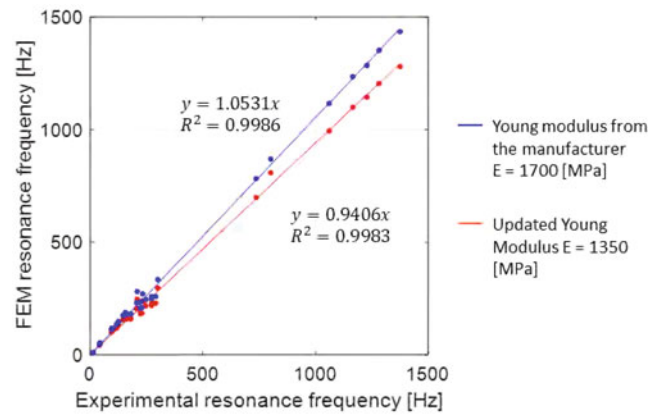
Fig. 8.14 Linear fit between numerical and experimental resonance frequencies for the updated Young's modulus and the one from manufacturer material datasheet

approach, which only considers static information. Moreover, static and dynamic data should be used to tune the nonlinear model. The main issue related to this type of analysis is the difficulty in the visual comparison between experimental and numerical mode shapes due to the complexity of the geometry that gives a high modal density (and therefore the presence of local and global modes), the material nonlinearity and non-ideal constraint conditions in the experiments, resulting in uncertainty on mode matching and data dispersion around the linear fit. The low-frequency vibration modes are dependent on the constraint used in the experiment and its efficiency in relation to the numerical case. Discrepancies between the hypothesis of clamped bottom surfaces of the metastructure used in the FEM model and its realization in the experiment lead to variations of resonance frequencies in the low range. In addition, we have scanned the top and side faces of the metastructure, which is not enough for a comprehensive classification of mode shapes. Having in mind these assumptions and limitations, a good correspondence between experimental and numerical data is present for our own purposes.

8.5   Conclusion

Vibrational metamaterials are emerging as a valid alternative to traditional materials for vibration isolation and absorption. In this chapter, we have focused on updating the simple linear numerical model of the metamaterial design recently proposed, which could then be used for further static and dynamic performance optimization. Material and mechanical properties have been estimated according to experimental data retrieved from compression tests and computed tomography. Numerical transmissibility function is compared with experimental data, considering both manufacturer and updated parameters, highlighting a better result in the latter case. A measurement campaign by scanning laser Doppler vibrometry

is done to retrieve the metastructure modal parameters. Furthermore, a comparison between numerical and experimental eigenfrequencies is carried out before and after parameter updating, obtaining a good agreement of the linear fit with the ideal diagonal case. Authors are currently working on a more rigorous numerical model with the aim of getting a better match with the experimental data and enabling the exploitation of the model for predicting the behavior of the metamaterial in different working conditions.

References

1. Wagner, P.-R., Dertimanis, V.K., Chatzi, E.N., Antoniadis, I.A.: Design of metamaterials for seismic isolation. In: Conference Proceedings of the Society for Experimental Mechanics Series, pp. 275–287. Springer International Publishing, Berlin (2016)
2. Reichl, K.K., Inman, D.J.: Constant mass metastructure with vibration absorbers of linearly varying natural frequencies. In: Mains, M., Blough, J.R. (eds.) Topics in Modal Analysis & Testing, vol. 10, pp. 153–158. Springer International Publishing, Cham (2017)
3. Essink, B.C., Inman, D.J.: Three-dimensional mechanical metamaterial for vibration suppression. In: Dervilis, N. (ed.). Special Topics in Structural Dynamics & Experimental Techniques, vol. 5, pp. 43–48. Springer International Publishing, Cham (2020)
4. Chondrogiannis, K.A., Dertimanis, V., Chatzi, E.: Application of geometrically nonlinear metamaterial device for structural vibration mitigation. In: Nonlinear Structures & Systems, vol. 1, pp. 109–118. Springer International Publishing, Berlin (2022)
5. Wang, K., Zhou, J., Tan, D., Li, Z., Lin, Q., Xu, D.: A brief review of metamaterials for opening low-frequency band gaps. Appl. Math. Mech. **43**(7), 1125–1144 (2022)
6. Zolfagharian, A., Bodaghi, M., Hamzehei, R., Parr, L., Fard, M., Rolfe, B.F.: 3d-printed programmable mechanical metamaterials for vibration isolation and buckling control. Sustainability **14**(11), 6831 (2022)
7. Lu, Z.-Q., Zhao, L., Ding, H., Chen, L.-Q.: A dual-functional metamaterial for integrated vibration isolation and energy harvesting. J. Sound Vib. **509**, 116251 (2021)
8. Srivatsa, S.M., Inman, D.J.: Characterization of fatigue in integrated tuned mass-dampers. In: Epp, D.S. (ed.). Special Topics in Structural Dynamics & Experimental Techniques, vol. 5, pp. 25–35. Springer International Publishing, Cham (2021)
9. Liu, K., Han, L., Hu, W., Ji, L., Zhu, S., Wan, Z., Yang, X., Wei, Y., Dai, Z., Zhao, Z., Li, Z., Wang, P., Tao, R.: 4d printed zero poisson's ratio metamaterial with switching function of mechanical and vibration isolation performance. Mater. Des. **196**, 109153 (2020)
10. Annessi, A., Zega, V., Chiariotti., P., Martarelli, M., Castellini, P.: An innovative wide and low-frequency bandgap metastructure for vibration isolation. J. Appl. Phys. **132**(8), 084903 (2022)

Chapter 9
Incorporating Uncertainty in Mechanics-Based Synthetic Data Generation for Deep Learning–Based Structural Monitoring

M. Cheraghzade and M. Roohi

Abstract This chapter presents a hybrid deep-learning methodology for seismic structural monitoring, damage detection, and localization of instrumented buildings. The proposed methodology develops mechanics-based structural models to generate sample response datasets by accounting for the uncertainty of model parameters that can highly affect the estimation of baseline model nonlinear responses. The mechanics-based models are developed considering uncertainties in the stiffness, strength, and geometry of the baseline numerical model's characteristics and elements. The baseline model is run multiple times with defined assumptions and variations in the selected parameters of the model. The uncertainty of model parameters is evaluated through the design of experiments methodology by employing the central composite design for sampling. A parameter effect analysis is used to assess the significance of the modeling input parameters on the selected structural output response, such as inter-story drifts. The generated sample response dataset is utilized for training a hybrid data-driven model for feature extraction. To select the damage-sensitive features, a convolutional neural network as the main feature extraction body of the network is used. In addition, wavelet packet–based nodal first temporal moments (energies) are also employed to boost the feature extraction power of the network as a complementary body. This data-driven model is designed to use global story-level noise-contaminated response measurements are employed as input for the data-driven model to perform damage detection and localization in a manner consistent with performance-based design criteria. The performance of the proposed methodology is studied in the context of numerical and experimental case studies developed based on the shake table testing of a concentrically braced frame subject to various input ground motion intensities at the E-Defense facility in Miki, Japan. The results show that the proposed methodology provides high accuracy in classifying and localizing various damage patterns.

Keywords Seismic monitoring · Deep learning · Model-based uncertainty · Wavelet packet transform · Central composite design

9.1 Introduction

Structural integrity and risk assessment of civil infrastructure assist decision-makers in the prior, during, and following extreme damaging events with maintenance, resource allocation, and planning. Risk and resilience assessment methodologies (such as FEMA P-58) and platforms [such as Interdependent Networked Community Resilience Modeling Environment (IN-CORE)] quantify physical damage, functionality, and consequences using probabilistic component and system-level fragility and vulnerability functions [1, 2]. On the other hand, structural monitoring systems provide near-real-time data and measurements (such as conventional vibration data and visual inspection data), which can be incorporated into structural integrity assessment procedures to improve the accuracy of physical damage assessment decision-making. Therefore, the accurate estimation of damage states as a critical element in loss estimation can be reliably achieved by adopting monitoring systems and methodologies.

The development of more feasible and practical SHM approaches will be enabled by recent advances in computational and big-data analytics methods, as well as low-cost sensors for data acquisition. Various vibrational methods for civil

M. Cheraghzade
International Institution of Earthquake Engineering and Seismology, Structural Research Center, Tehran, Iran

M. Roohi (✉)
Durham School of Architectural Engineering and Construction, University of Nebraska–Lincoln, Omaha, NE, USA
e-mail: milad.roohi@unl.edu

infrastructure have been proposed in the past two decades, including methods based on models, methods based on model-data fusion, and methods based on data [3–5]. Recent SHM initiatives have concentrated on creating data-driven (model-free) techniques that use machine learning algorithms such as deep learning (DL) to extract damage-sensitive characteristics exclusively from measured data that are hand-crafted or automated. Notably, practical uses of the DL approaches are seen in a variety of seismic monitoring projects. An image classifier model is used to quickly assess post-earthquake damage to structures by using convolutional neural network (CNN) deep models with transformed wavelet data as inputs [6]. Based on various evaluation matrices, the DL methods are used to predict structural and earthquake features used in performance-based and seismic design [7]. The estimate of responses and the computation of various structures have been the subject of substantial numerical and empirical models in previous years. With the aid of deep models for seismic monitoring, a time-series predictor is suggested to reconstruct the responses of nonlinear structural models [8]. Researchers have developed deep models to identify unknown damage patterns that can detect damage and stiffness loss with excellent generalization ability. For example, an application of CNN for detecting damage caused by unknown seismic excitations is conducted using wavelet-based transmissibility [9].

In the context of earthquake engineering and structural damage assessment, numerous studies have focused on the uncertainties of finite element (FE) models and damage localization models. Using a response surface, Monte Carlo sampling has been used to assess probabilistic collapse risk assessments and predict structural responses [10]. There is, however, a greater need to pay attention to variation and uncertainty in FE models since model-based uncertainty affects the evaluation of engineering demand parameters (i.e., inter-story drift) needed to perform collapse risk assessments more than other uncertainty sources. A comparison of the distribution functions of intact and damaged structures was used to evaluate the uncertainty of FE models for damage assessment using modal data such as frequencies and shapes [11]. FE model-based uncertainty has also been used for damage assessment using a comprehensive set of non-probabilistic methods, such as response surface models [12]. Existing literature that explores the significant influence of model-based uncertainty assessment does not sufficiently address the notable impact of model-based uncertainty assessment integrated with data-driven methods. As discussed previously, it is imperative to develop DL algorithms that can capture the FE modeling error and provide an accurate estimate of the structural damage.

This chapter presents a DL-based structural monitoring methodology that incorporates model-based sample generation with consideration of uncertainties in the process of data-driven monitoring. The performance of data-driven algorithms for structural monitoring is highly dependent on the availability and accuracy of training data, which is difficult to obtain from real-world structures due to the extensive instrumentation of various structures being impractical (primarily due to budget and maintenance constraints) and the low probability of high-consequence events like earthquakes. The performance of data-driven methods can be considerably enhanced by using robust numerical techniques that consider model-based uncertainty. As a result, the main objectives of this research are summarized as follows:

- Accounting for the uncertainty of model parameters that can highly affect the accuracy of dynamic response estimation and complimentary sample generation estimation using the baseline nonlinear structural model, which are developed based on highly idealized engineering assumptions which might not accurately represent the physical aspects of the actual structures and lead to various levels of modeling error
- Augmenting real-world measurements with simulated physics-based measurements to improve the robustness and accuracy of data-driven seismic monitoring

The experimental validation of the proposed approach demonstrates its capability and effectiveness in helping structural engineers make informed and swift decisions regarding post-earthquake assessment of critical instrumented building structures and improving earthquake resiliency of communities.

9.2   Methodology

In accordance with performance-based design objectives, this section proposes a hybrid DL-based approach for post-earthquake damage identification and localization of instrumented structures. The approach divides damage quantification into five steps, as shown in Fig. 9.1. By taking parameter uncertainty into account, mechanics-based structural models are first created in the first two steps. Each model is then examined while being subjected to a variety of ground movements, producing and augmenting a sample response dataset in the third step. The sample response datasets will then be used in the fourth and fifth steps to train a hybrid data-driven model for damage identification and localization utilizing global story-level
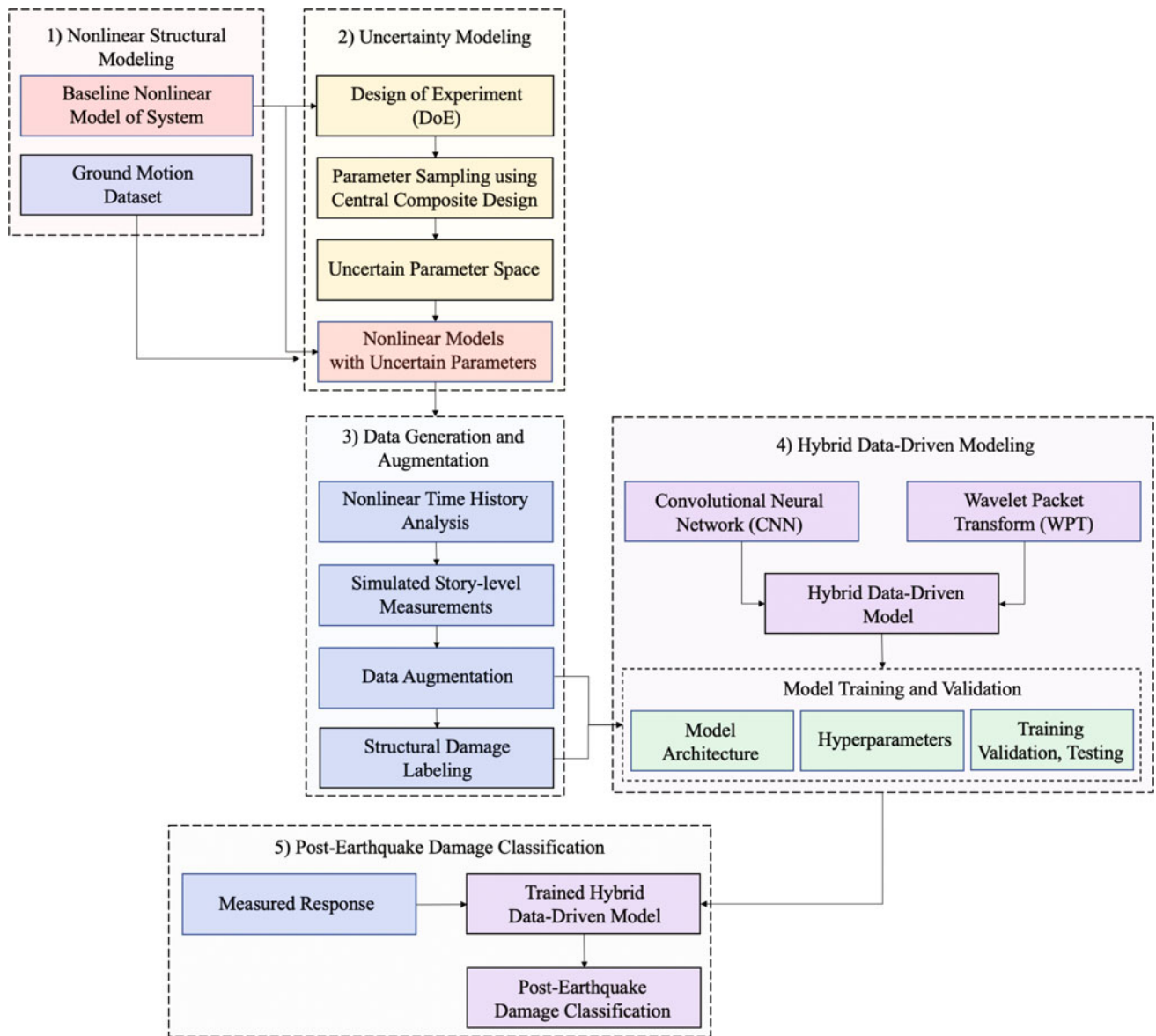
Fig. 9.1 Proposed methodology for deep learning–based structural monitoring

noise-contaminated response measurements. This research employs accelerometers as the preferred sensor because of their widespread usage, long lifespan, and dependability. It then applies a hybrid DL model for damage estimation based on code-based seismic performance levels to the recorded story-level acceleration responses. Interested readers are directed to [17] regarding further details about the proposed methodology and its application validated through numerical and experimental case studies.

9.2.1 Structural Modeling

The simulation of the nonlinear structural models and sample generation is the first step in putting the proposed methodology into practice. From the initial stages of damage to total collapse, the nonlinear structural models can effectively simulate all key modes of deformation and degradation in the structure. It is, therefore, important that these models encompass local and global inelastic behavior as well as the structure's actual behavior.
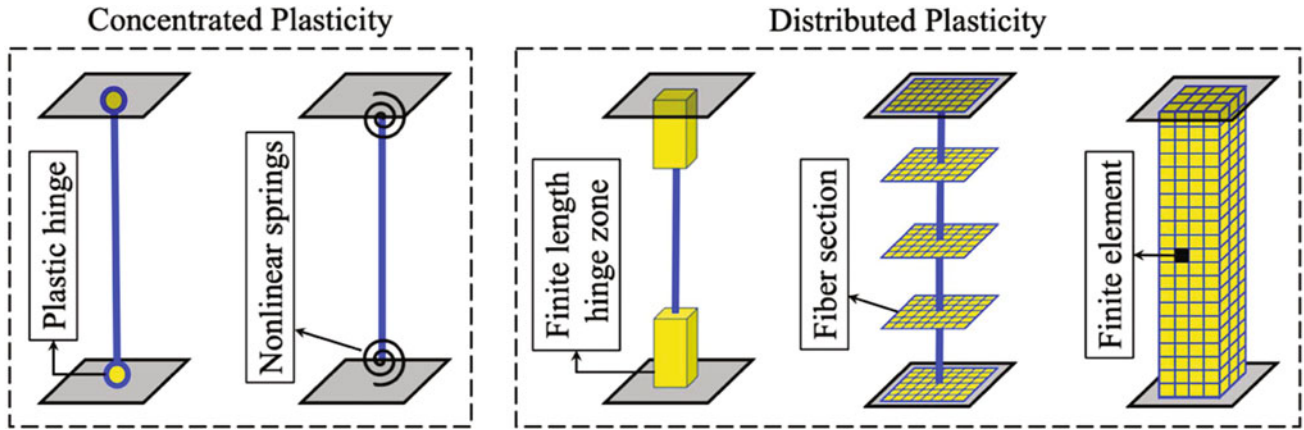
Fig. 9.2 Idealized nonlinear models for simulation of various structural elements. (Adopted from [13])

The modeling of non-linearity can generally be classified into concentrated and distributed plasticity. The concentrated plasticity models involve plastic hinge and nonlinear spring hinge models. On the other hand, the distributed plasticity models are also categorized as finite length hinges, fiber section, and finite element continuum models [13]. Figure 9.2 shows a comparison of five idealized model types for simulating the inelastic response of the structural frame. It is recommended to simulate nonlinear structural behavior using advanced structural modeling software (such as OpenSees [16]) that provides a wide range of nonlinear simulation material and element models, solution algorithms, data processing procedures, and distributed computing approaches.

9.2.2 Uncertainty Assessment and Data Augmentation

Several uncertainties can be considered, including those in the simulation models, measurements, or damage detection models. Specifically, this study focuses on assessing the uncertainty of baseline numerical model parameters, such as strength, stiffness, and geometry, which can strongly influence the simulated responses that are used for training data-driven models.

It is normal for modeling parameter values to vary from their intended values. It is important for the classification of post-earthquake damage to examine the randomness brought by these variations. The central composite design (CCD) sampling technique is used in this work to assess the consequences of modeling uncertainty using the criteria of design of experiments (DOEs) techniques. CCD contains a fractional factorial design $2k$ with ±1 levels; $2k$ axial points with $\pm\alpha$ star points and n_0 center points. Different types of CCDs are proposed based on the value of α, namely, rotatable, spherical, face-centered, and practical. As a result of selecting α and n_0, the CCD is characterized by certain desirable properties. For rotatable designs, the α factor is calculated in Eq. (9.1).

$$\alpha = \frac{2^{k-f}\left(n_f\right)_{\frac{1}{4}}}{n_s} \tag{9.1}$$

where n_f is the number of replicates of rows in the original factorial design, n_s is the number of replicates at the axial points, f is the fractional factorial, and k is the number of parameters. Mechanic-based models are developed based on the variation of variables (i.e., modeling parameters such as f_y) within the DOEs criteria. Experimental and statistical studies can determine the amount of variation in baseline values. In other words, various samples for the baseline values should be statistically evaluated to determine the amount of mean, variance, and statistical distribution of the variables. Moreover, the CCD method determines the variation of parameters in each test which allows for the augmentation of data for various structural responses conducted in each sampling with different sets of modeling parameters. Additionally, the generated data are split into same-length sub-fragments of the raw samples, as shown in Fig. 9.3, as another approach for data augmentation.
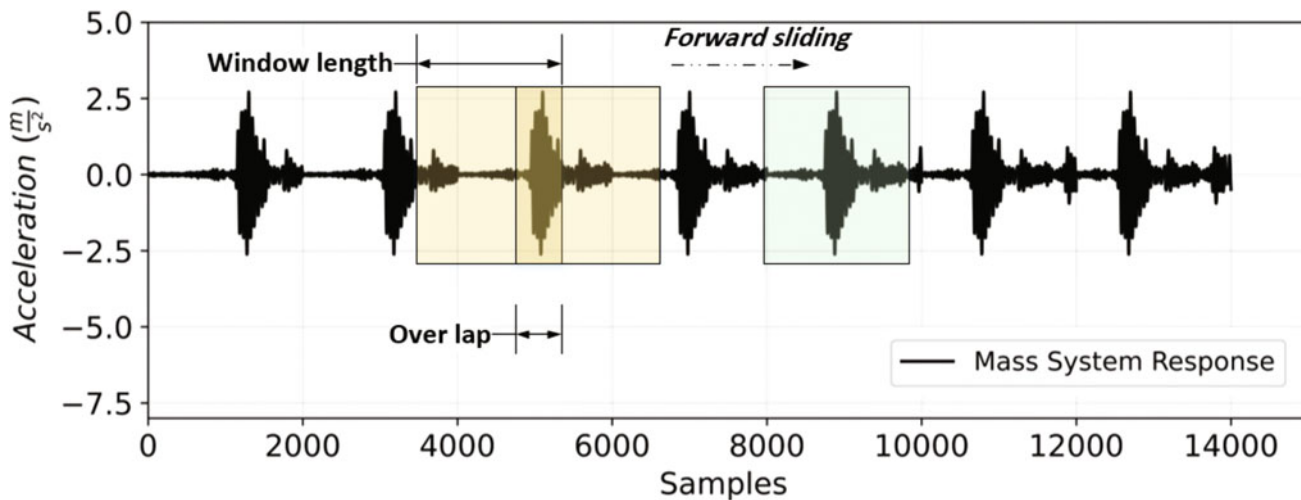
Fig. 9.3 Sliding window strategy for data augmentation on generating samples
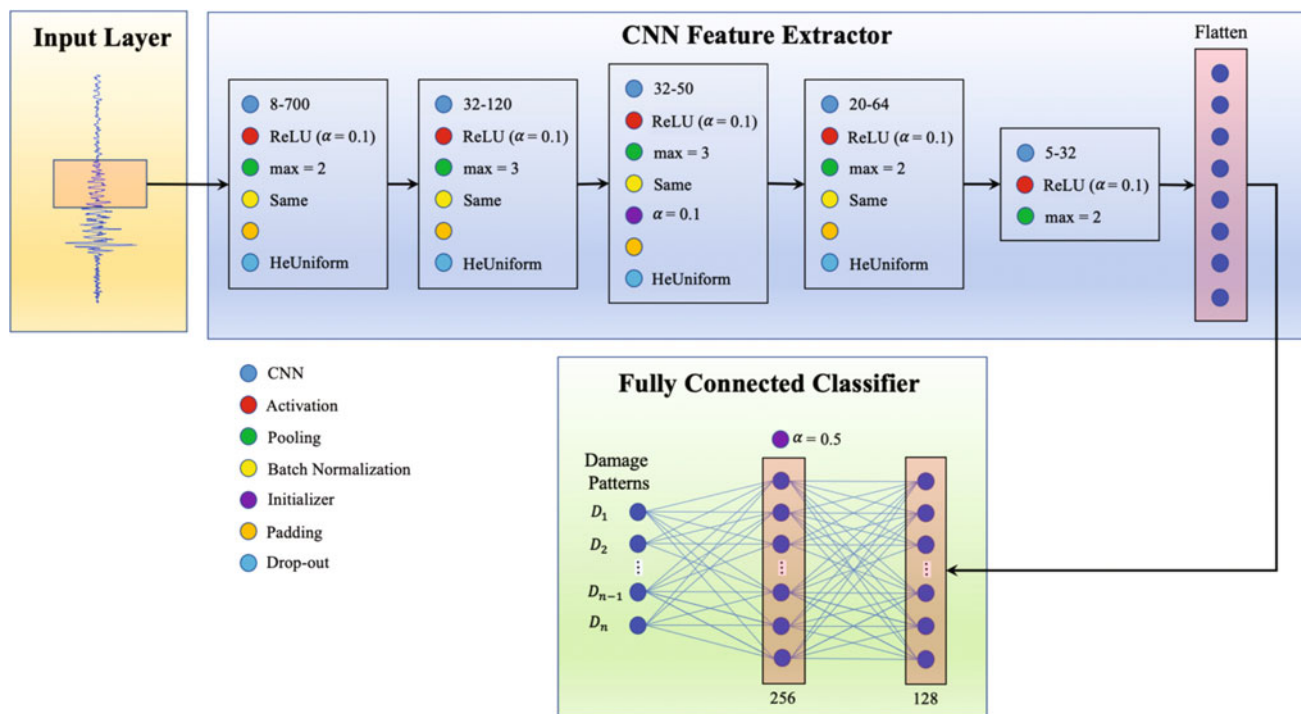
Fig. 9.4 Proposed CNN architecture for damage state classification

9.2.3 Damage State Classification

For reliable feature extraction, a hybrid classifier composed of a CNN and a wavelet packet transform (WPT) is suggested in this study. The last flattened layer of the CNN model and the WPT nodal energies are combined in the hybrid model to feed the fully connected classifier, as shown in Fig. 9.4. In order to improve classification outcomes, retrieved features from the CNN and WPT modules are employed as numerous inputs for the fully connected layers. In comparison to a CNN model, it is anticipated that the hybrid model will require fewer training iterations, making network convergence simpler.

CNNs are the commonly used type of deep neural network for automatic pattern recognition (i.e., mapped feature extraction). These networks use learnable units called kernels or filters for abstracting detail extraction in vector or grid

form through the convolution procedure as element-wise multiplications. The mathematical operation of convolution is a well-known and regularly used function defined in Eq. (9.2).

$$f(i) = \int_{-\infty}^{+\infty} x(n)k(i-n)\, d_n \tag{9.2}$$

The filter function k slides over the input data x element-by-element through different convolutional layers, and the results are added up. Interested readers are referred to [17] for further detail regarding the background theories of CNN and other associated algorithms.

In order to assess the architecture of the suggested model, many hyperparameters need to be tested. To do this, many factors, including the CNN network's learning rate and the number of layers, are assessed via sensitivity try-and-error analysis. With hyperparameters chosen at random, nearly 40 networks are built. An Intel i7-10750H CPU, 16GB of RAM, and Nvidia GeForce RTX 2060 graphics cards are used for the numerical calculations, which are performed simultaneously.

9.3 Case Study of Experimental Concentrically Braced Frame

9.3.1 Experimental and Numerical Modeling

A full-scale single-story chevron concentrically braced frame (CBF) structure, as shown in Fig. 9.5, was used as a test case in this section to demonstrate the effectiveness of the proposed damage assessment methodology. This study utilizes seismic test data, such as acceleration responses, despite the fact that the structure was also exposed to other monotonic loads. For mass system and shake table acceleration response measurements, eight accelerometers with sample rates of 1000 samples per second (i.e., 1000 Hz) are used with four sensors on the shake table and four sensors on the mass system. For the 12%, 14%, 2 × 28%, 42%, and 70% amplification levels, respectively, the acceleration response data of the East-West component of the JR Takatori seismic motion are measured and accessible. For further information on the shaking table testing of the case study framework, the readers are directed to [14].

The experimental responses from the shaking table test are used to validate the numerical modeling for further data generation for DL model training and validation. The shaking table model's numerical CBF specimen with a single-story, single-span configuration is created first, as shown in Fig. 9.6. The model is developed using the OpenSeesPy python library of the open system for earthquake engineering simulation (OpenSees [16]). The validation results for various structural
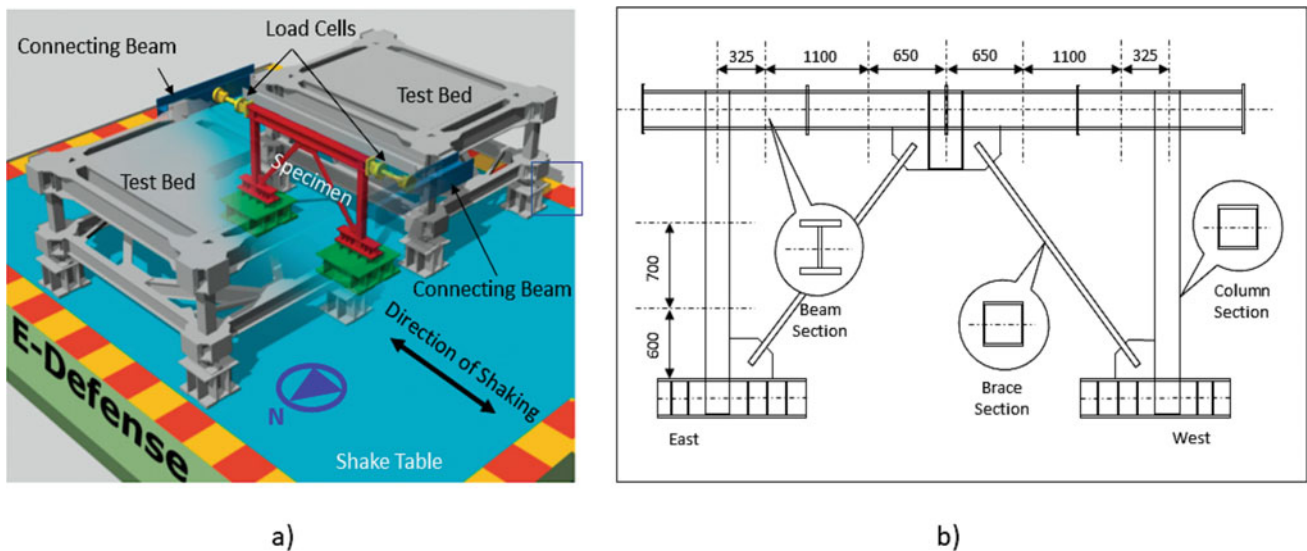
Fig. 9.5 (**a**) Schematic figure of the shake table test bed; (**b**) Schematic figure of the frame geometry (Adapted from [14])
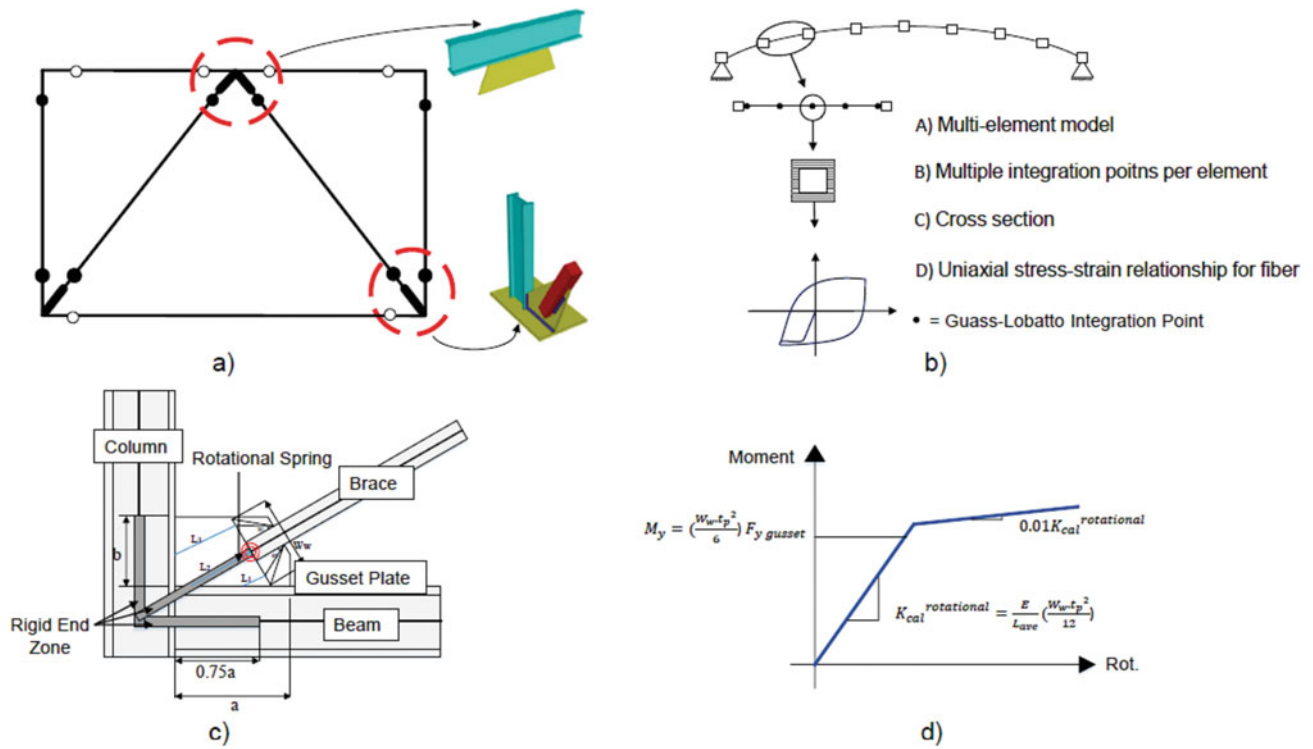
Fig. 9.6 (**a**) Schematic figure of the simulated model validated with experimental data; (**b**) Methodology toward modeling of the braces; (**c**) Modeling of gusset plates with concentrated hinges; and (**d**) Nonlinear modeling of the hinges
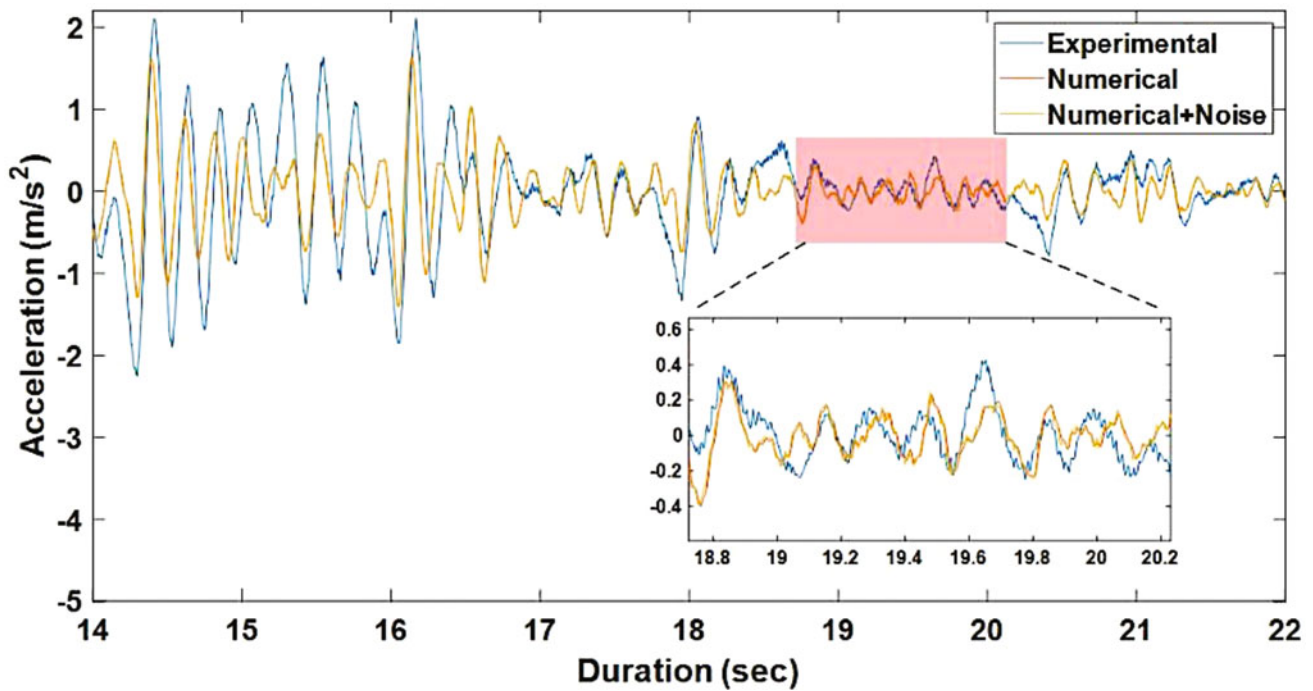
Fig. 9.7 Validation result for acceleration response simulation of 42% amplification level of the Takatori earthquake for numerical modeling

responses were captured. As shown in Fig. 9.7, the acceleration as a high-frequency response is simulated and is compared to experimental data for the 42% amplification level of the Takatori earthquake.

Table 9.1 Damage state labeling for classification

Damage state	Amplification level (%)	Performance label	Damage description
D_1	14	No damage	No damage and nonlinear behavior have been observed during the test
D_2	28	Slight damage	A small out-of-plane displacement occurred
D_3	42	Moderate damage	A residual out-of-plane displacement occurred near the mid-length of the left brace
D_4	70	Extensive damage	Both steel frames fractured near their mid-length due to low-cycle fatigue

Fig. 9.8 The confusion matrix for the incorporated experimental and numerical trained hybrid model tested with the experimental datasets

9.3.2 Damage Classification

Incorporating numerical and experimental data into the proposed methodology, the performance of the data-driven hybrid model is assessed in this section. The proposed method is examined for damages associated with various Takatori earthquake amplification levels, as shown in Table 9.1. The numerical model is applied to four distinct Takatori earthquake amplification levels, with 14%, 28%, 42%, and 70% with consideration of noise effects, respectively. The recorded simulated responses for various modeling parameters [i.e., the module of elasticity (E), yield stress (f_y), and geometry of brace (thickness (t))] in different test sets in accordance with CCD sampling are generated and augmented for training and validation. Additionally, a significant parameter analysis is also conducted to evaluate the significance of every parameter in structural response, such as story drift ratios. This study utilizes the ground truth damages associated with the experimental response as a test dataset in the hybrid model.

On experimental test data, the hybrid model achieves an overall accuracy of 82%, as shown in Fig. 9.8. The hybrid model can successfully classify damage patterns corresponding to various amplification levels.

9.4   Conclusions

This chapter presents a deep learning–based structural monitoring methodology that incorporates uncertainty in mechanics-based models to generate synthetic data and train hybrid data-driven models. The implementation of the proposed methodology starts by developing nonlinear simulated models that account for model-based parameter uncertainty. The developed models with determined sets of parameters analyzed through parameter significant analysis produce and augment sample response datasets using central composite design sampling by subjecting them to various sets of ground motions within the criteria of design of experiments. Hybrid DL-based models are then trained using noise-contaminated global story-level response measurements as generated sample response datasets. This hybrid model includes convolutional neural networks and wavelet packet transforms for damage-sensitive feature extraction. Damage feature extraction and classification are achieved in this methodology by utilizing and incorporating robust numerical models for training and validation with experimental data as test datasets. The numerical model used in this study for sample generation is validated with experimental responses recorded from shaking table concentrically braced frame test data conducted at E-Defense, Miki, Japan, analyzed for various ground motions. Afterward, the proposed hybrid data-driven model is trained based on the sample dataset for assessing post-earthquake damages. The proposed methodology is verified and validated using global dynamic response measurements, demonstrating that the proposed approach can detect and localize structural post-earthquake damage patterns.

References

1. FEMA: FEMA P-58-1: Seismic Performance Assessment of Buildings. Volume 1—Methodology. Federal Emergency Management Agency, Washington, DC (2012)
2. Roohi, M., van de Lindt, J.W., Rosenheim, N., Hu, Y., Cutler, H.: Implication of building inventory accuracy on physical and socio-economic resilience metrics for informed decision-making in natural hazards. Struct. Infrastruct. Eng. **17**(4), 534–554 (2020)
3. Roohi, M., Hernandez, E.M.: Performance-based post-earthquake decision making for instrumented buildings. J. Civ. Struct. Heal. Monit. **10**(5), 775–792 (2020)
4. Yu, Y., Wang, C., Gu, X., Li, J.: A novel deep learning-based method for damage identification of smart building structures. Struct. Health Monit. **18**(1), 143–163 (2019)
5. Zhao, X., Lang, Z.: Baseline model based structural health monitoring method under varying environment. Renew. Energy. **138**, 1166–1175 (2019)
6. Mangalathu, S., Jeon, J.S.: Ground motion-dependent rapid damage assessment of structures based on wavelet transform and image analysis techniques. J. Struct. Eng. **146**(11), 04020230 (2020)
7. Gharehbaghi, S., Gandomi, M., Plevris, V., Gandomi, A.H.: Prediction of seismic damage spectra using computational intelligence methods. Comput. Struct. **253**, 106584 (2021)
8. Xu, Z., Chen, J., Shen, J., Xiang, M.: Recursive long short-term memory network for predicting nonlinear structural seismic response. Eng. Struct. **250**, 113406 (2022)
9. Lei, Y., Zhang, Y., Mi, J., Liu, W., Liu, L.: Detecting structural damage under unknown seismic excitation by deep convolutional neural network with wavelet-based transmissibility data. Struct. Health Monit. **20**(4), 1583–1596 (2021)
10. Liel, A.B., Haselton, C.B., Deierlein, G.G., Baker, J.W.: Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. Struct. Saf. **31**(2), 197–211 (2009)
11. Bakhary, N., Hao, H., Deeks, A.J.: Damage detection using artificial neural network with consideration of uncertainties. Eng. Struct. **29**(11), 2806–2815 (2007)
12. Fathnejat, H., Ahmadi-Nedushan, B.: An efficient two-stage approach for structural damage detection using meta-heuristic algorithms and group method of data handling surrogate model. Front. Struct. Civ. Eng. **14**(4), 907–929 (2020)
13. Roohi, M., Hernandez, E.M., Rosowsky, D.: Reconstructing element-by-element dissipated hysteretic energy in instrumented buildings: application to the Van Nuys Hotel testbed. J. Eng. Mech. **147**(1), 04020141 (2021)
14. Okazaki, T., Lignos, D.G., Hikino, T., Kajiwara, K.: Dynamic response of a chevron concentrically braced frame. J. Struct. Eng.-ASCE. **139**(4), 515–525 (2013)
15. Hsiao, P.C., Lehman, D.E., Roeder, C.W.: Improved analytical model for special concentrically braced frames. J. Constr. Steel Res. **73**, 80–94 (2012)
16. McKenna, F.T.: Object-Oriented Finite Element Programming: Frameworks for Analysis, Algorithms and Parallel Computing. University of California, Berkeley (1997)
17. Cheraghzade, M., Roohi, M.: Deep learning for seismic structural monitoring by accounting for mechanics-based model uncertainty. J. Build. Eng. **57**, 104837 (2022)

Chapter 10
Aerodynamic Load Estimation in Wind Turbine Drivetrains Using a Bayesian Data Assimilation Approach

Mohammad Valikhani, Vahid Jahangiri, Hamed Ebrahimian, Sauro Liberatore, Babak Moaveni, and Eric Hines

Abstract This work explores a Bayesian data assimilation approach to estimate the aerodynamic input force on a wind turbine drivetrain. To this end, a high-fidelity wind turbine drivetrain model is created in SIMPACK to generate synthetic data. The NREL 5 MW wind turbine is considered as a case study in this work. The synthetic data include generator rotational speed, generator torque, rotor rotational speed, which are simulated and used as measurement data. A low-fidelity model of the drivetrain is developed for load estimation to accelerate the estimation process. The Bayesian data assimilation approach is employed to integrate synthetic data with the low-fidelity model to estimate aerodynamic input load.

Keywords Input load estimation · Wind turbine drivetrain · Bayesian data assimilation

10.1 Introduction

The wind energy industry has gained remarkable attention in the past decade due to climate change concerns [1]. Wind turbines are subjected to harsh environmental loadings with limited maintenance access resulting in increased levelized cost of energy. Among different wind turbine components, drivetrains are critical subsystems in terms of operation and maintenance [2]. Wind turbine drivetrains are subjected to uncertain aerodynamic loads resulting in fatigue damage in different drivetrain components. Aerodynamic load estimation is required to predict the remaining useful life of the components. Direct measurement of the load is costly if not impossible, so inverse modeling can be used to estimate the input loads.

Limited studies have focused on the aerodynamic load estimation of drivetrains. In a case study, Perisic et al. proposed an indirect method to estimate the main shaft torque from the simulated rotor speed and generator torque [3]. For a mechatronic powertrain, Forrier et al. developed a method to estimate the input torque of a driveline [4]. They used the Kalman-based method to fuse multi-sensor data with the powertrain model for torque estimation. In this study, the effect of sensor type on the estimated input torque was investigated, and the results indicated that the acceleration data can broaden the frequency band of the estimated load.

In the present work, a Bayesian inference method is employed to estimate the input aerodynamic torque of the wind turbine drivetrain using synthetic data [5, 6]. The synthetic data is generated from a high-fidelity model of the drivetrain developed in SIMPACK. The Bayesian inference method is used to infuse the data with a lumped-mass model of the drivetrain to infer unknown aerodynamic torque.

10.2 Bayesian Inference Method

Governing differential equation of a linear dynamic system at step *t* can be expressed as

M. Valikhani · V. Jahangiri (✉) · H. Ebrahimian
Department of Civil and Environmental Engineering, University of Nevada, Reno, NV, USA

S. Liberatore · B. Moaveni · E. Hines
Department of Civil and Environmental Engineering, Tufts University, Medford, MA, USA

Table 10.1 The Bayesian inference method for input estimation

1. Set the window parameters and initial values of the input vector and associated covariance matrix
1.1. Set the window counter $m = 1$.
1.2. Set the window size t_l and sliding size t_s.
1.3. Set the initial values of a prior mean estimate of the input vector and covariance matrix as $\hat{\boldsymbol{\varphi}}_{m,0}^+ = \left[\hat{\mathbf{u}}_{t_1^1:t_2^1,0}^T\right]^T$ and $\left(\hat{\mathbf{P}}_\varphi^+\right)_{m,0} = \left(\hat{\mathbf{P}}_\mathbf{u}^+\right)_{t_1^1:t_2^1,0}$
1.4. Set the perturbation covariance matrix $\mathbf{Q}$ and measurement error covariance matrix $\widetilde{\mathbf{R}}$; see Ref. [5] for more details.
2. Update the input vector and covariance matrix from the data in the mth **window**
2.1. Set the iteration counter $i = 1$
2.2. Set $\hat{\boldsymbol{\varphi}}_{m,i}^- = \hat{\boldsymbol{\varphi}}_{m,i-1}^+$ and $\left(\hat{\mathbf{P}}_\varphi^-\right)_{m,i} = \left(\hat{\mathbf{P}}_\varphi^+\right)_{m,i-1} + \mathbf{Q}$
2.3. Predict the response according to input vector $\hat{\boldsymbol{\varphi}}_{m,i}^-$ as: $\hat{\mathbf{y}}_{t_1^m:t_2^m} = \mathbf{h}_{t_1^m:t_2^m}\left(\hat{\boldsymbol{\varphi}}_{m,i}^-, \hat{\mathbf{u}}_{1:t_1^m-1}, , \dot{\mathbf{x}}_0, \mathbf{x}_0\right)$
2.4. Calculate the response sensitivity with respect to the input vector: $\mathbf{C} = \partial\mathbf{h}_{t_1^m:t_2^m}\left(\boldsymbol{\varphi}_{m,i}, \hat{\mathbf{u}}_{1:t_1^m-1}, , \dot{\mathbf{x}}_0, \mathbf{x}_0\right)/\partial\boldsymbol{\varphi}_{m,i}\Big|_{\boldsymbol{\varphi}_{m,i}=\hat{\boldsymbol{\varphi}}_{m,i}^-}$
2.5. Compute the Kalman gain matrix: $\mathbf{G} = \left(\mathbf{C}^T\widetilde{\mathbf{R}}^{-1}\mathbf{C} + \left(\hat{\mathbf{P}}_\varphi^-\right)_{m,i}^{-1}\right)^{-1}\mathbf{C}^T\widetilde{\mathbf{R}}^{-1}$
2.6. Update the input vector and associated covariance matrix: $\hat{\boldsymbol{\varphi}}_{m,i}^+ = \hat{\boldsymbol{\varphi}}_{m,i}^- + \mathbf{G}\left(\mathbf{y}_{t_1^m:t_2^m} - \hat{\mathbf{y}}_{t_1^m:t_2^m}\right)$ and $\left(\hat{\mathbf{P}}_\varphi^+\right)_{m,i} = (\mathbf{I} - \mathbf{GC})\left(\hat{\mathbf{P}}_\varphi^-\right)_{m,i}(\mathbf{I} - \mathbf{GC})^T + \mathbf{G}\widetilde{\mathbf{R}}\mathbf{G}^T$
3. Set the initial values of the input vector and covariance matrix for $(m+1)^{\text{th}}$ **window**
3.1. Set $m = m + 1$.
3.2. Compute the initial input vector: $\hat{\boldsymbol{\varphi}}_{m,0}^-$ (see Ref. [5] for more details).
3.3. Compute the covariance matrix: $\left(\hat{\mathbf{P}}_\varphi^-\right)_{m,0}$ (see Ref. [5] for more details).
3.4. Return to step 2.

$$\mathbf{M}\ddot{\mathbf{x}}_t + \mathbf{C}\dot{\mathbf{x}}_t + \mathbf{K}\mathbf{x}_t = \mathbf{u}_t \tag{10.1}$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are mass, damping, and stiffness matrices, respectively; $\ddot{\mathbf{x}}_t$, $\dot{\mathbf{x}}_t$, $\mathbf{x}_t$, and $\mathbf{u}_t$ denote acceleration, velocity, displacement, and input force vectors, respectively.

Generally, the measurement model can be represented by model prediction response with additive noise model as

$$\mathbf{y}_t = \mathbf{h}_t\left(\mathbf{u}_{1:t}, \dot{\mathbf{x}}_0, \mathbf{x}_0\right) + \mathbf{v}_t \quad \mathbf{v}_t \sim N\left(\mathbf{0}, \mathbf{R}_t\right) \tag{10.2}$$

where $\mathbf{y}_t$ is measured responses vector, $\mathbf{h}_t(.)$ Is predicted response based on Eq. (10.1), $\mathbf{u}_{1:t} = \left[\mathbf{u}_1^T, \mathbf{u}_2^T, \cdots, \mathbf{u}_t^T\right]^T$ is time history of input load, $\dot{\mathbf{x}}_0$, $\mathbf{x}_0$ are initial speed and displacement, and $\mathbf{v}_t$ denotes the measurement error vector assumed as a Gaussian white noise process with zero mean and covariance matrix $\mathbf{R}_t$.

The Bayesian inference method is employed to estimated unknown input force using measurement data [5]. In this method, data sets are partitioned into windows with starting time step t_1^m and ending time step t_2^m, and then the Bayesian inference method is used to integrate the model with the data across each window. Table 10.1 demonstrates the Bayesian inference method for input load estimation.

10.3 Wind Turbine Drivetrain Model

High fidelity model of the NREL 5 MW drivetrain [7] is developed in SIMAPCK software, which is a multibody system simulation software [8]. In this model, the drivetrain comprises of three gear sets, including two planetary gear sets and one parallel gear set. Figure 10.1 shows the model of the NREL 5 MW drivetrain. Details of the drivetrain components model are summarized in Table 10.2.

The simulated aerodynamic and generator torques and rotor and generator speeds are shown in Fig. 10.2. As can be seen, the generator speed, scaled with gear ratio $N = 97.8$, matches the rotor speed. This indicates a rigid body behavior in the drivetrain system within the excitation frequencies in this case study. This is because the external loading here has low-frequency content and cannot excite the higher-frequency modes of the system. On the other hand, internal loading such as mesh and impact loading (e.g., during the start and stop process) can excite higher frequency mode shapes; however, higher frequency response components are not usually observable in the speed response.
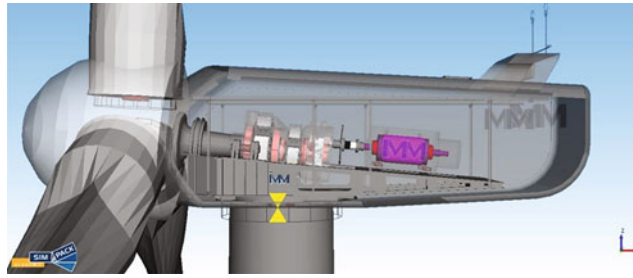
Fig. 10.1 Model of the NREL 5 MW wind turbine in SIMPACK software

Table 10.2 Details of the NREL 5 MW drivetrain model

Components	Model description
Main shaft	3D finite element model
Carrier in the first and second planetary sets	3D finite element model
Gears in the first and second planetary sets	Rigid gear bodies but flexible tooth contacts
Low-speed intermediate shaft	Cylindrical bushing (torsional spring)
High-speed intermediate shaft	Cylindrical bushing (torsional spring)
Gear in parallel set	Rigid gear bodies but flexible tooth contacts
High-speed shaft	Torsional spring
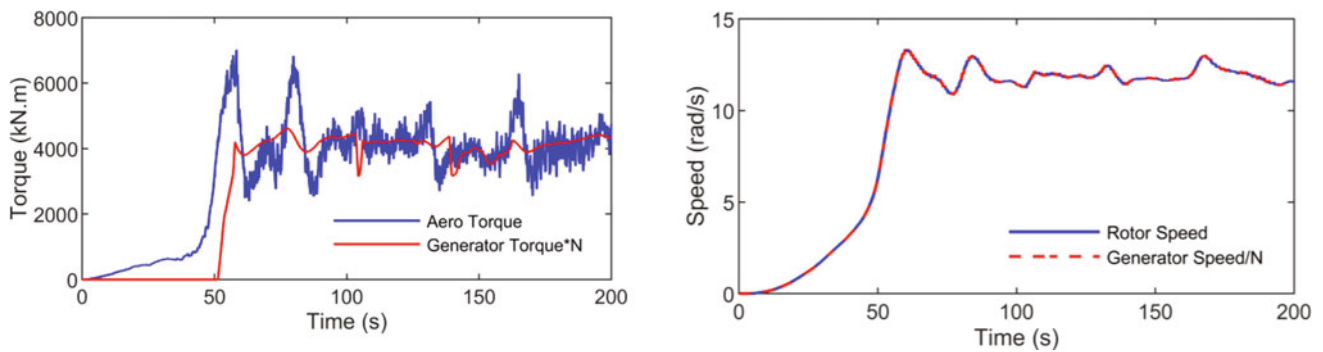
Fig. 10.2 Aerodynamic and generator torques (*left*), rotor and generator speeds (*right*)
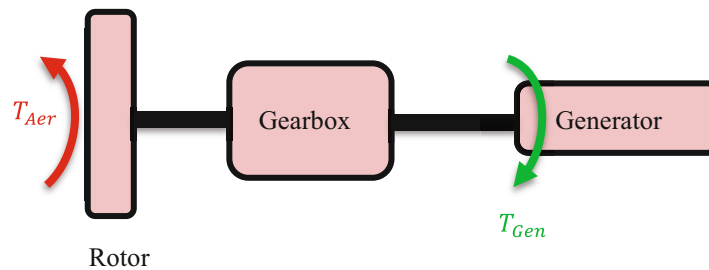
Fig. 10.3 Lumped mass model of the NREL 5 MW drivetrain

In addition to the SIMPACK model, a low-fidelity model of the drivetrain is developed using the mass method. The lumped mass model is computationally efficient and can accelerate the estimation process. Figure 10.3 shows the lumped mass model of the NREL 5 MW drivetrain. The model includes three masses connected by shafts. The drivetrain is subjected to aerodynamic and generator loads from the rotor and generator sides, respectively. The governing equation of motion can be derived through the Lagrange method [9].
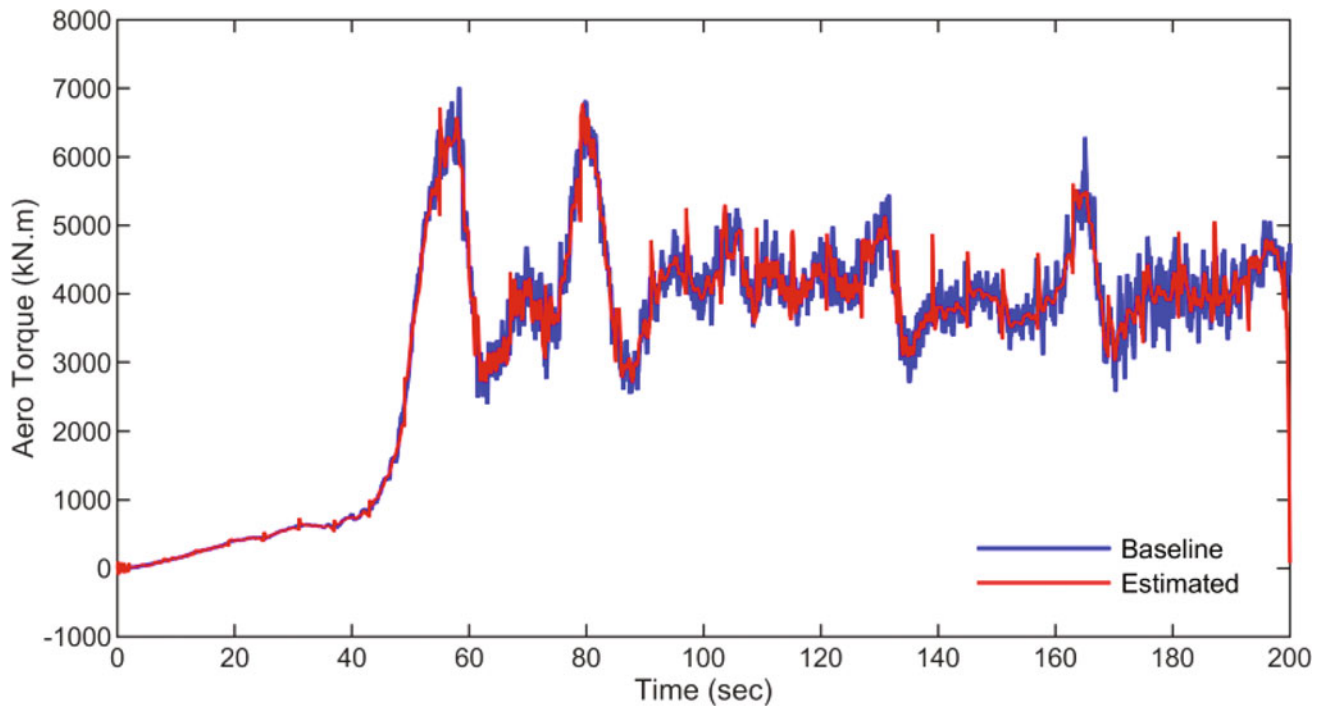
Fig. 10.4 The estimated aerodynamic torque versus the baseline

10.4   Results

The Bayesian inference method is employed to estimate aerodynamic load using lumped mass model and synthetic data. The data include rotor speed, generator speed, and generator torque as shown in Fig. 10.2. The Bayesian inference method is applied to integrate data with a low-fidelity model to estimate aerodynamic load. Figure 10.4 shows the time history of estimated aerodynamic loads, which is compared with the baseline. As can be observed, the estimated load has a reasonable agreement with the baseline load with Relative Root Mean Square Error [9] of 7.5%. This shows the potential applicability of the proposed approach for load estimation in drivetrain systems.

10.5   Conclusions

The load estimation of the wind turbine drivetrain is important for damage identification and remaining useful life prediction. Instrumentation to measure input load is costly and difficult. Alternatively, virtual sensing can be used to estimate aerodynamic load. In this work, a Bayesian inference approach was employed to estimate aerodynamic torque using synthetic data and a low-fidelity model. Synthetic data were generated using a high-fidelity model of a drivetrain system. A lumped mass model of the drivetrain was developed and used for load estimation. The results showed a good agreement between estimated and baseline input loads.

References

1. "Wind Vision: A New Era for Wind Power in the United States | Department of Energy." https://www.energy.gov/eere/wind/downloads/wind-vision-new-era-wind-power-united-states

2. Wiser, R., Bolinger, M., Lantz, E.: Assessing wind power operating costs in the United States: results from a survey of wind industry experts. Renew. Energy Focus. **30**, 46–57 (2019). https://doi.org/10.1016/J.REF.2019.05.003
3. Perišić, N., Kirkegaard, P.H., Pedersen, B.J.: Cost-effective shaft torque observer for condition monitoring of wind turbines. Wind Energy. **18**(1), 1–19 (2015). https://doi.org/10.1002/WE.1678
4. Forrier, B., Naets, F., Desmet, W.: Broadband load torque estimation in mechatronic powertrains using nonlinear Kalman filtering. IEEE Trans. Ind. Electron. **65**(3), 2378–2387 (2018). https://doi.org/10.1109/TIE.2017.2739709
5. Ebrahimian, H., Astroza, R., Conte, J.P., Papadimitriou, C.: Bayesian optimal estimation for output-only nonlinear system and damage identification of civil structures. Struct. Control. Health Monit. **25**(4), e2128 (2018). https://doi.org/10.1002/STC.2128
6. Valikhani, M., Younesian, D.: Bayesian framework for simultaneous input/state estimation in structural and mechanical systems. Struct. Control. Health Monit. **26**(9), e2379 (2019). https://doi.org/10.1002/STC.2379
7. Jonkman, J., Butterfield, S., Musial, W., Scott, G.: Definition of a 5-MW Reference Wind Turbine for Offshore System Development (2009). https://doi.org/10.2172/947422
8. "Simpack MBS Software | Dassault Systèmes." https://www.3ds.com/products-services/simulia/products/simpack/
9. Valikhani, M., Jahangiri, V., Ebrahimian, H., Moaveni, B., Liberatore, S., Hines, E.: Inverse modeling of wind turbine drivetrain from numerical data using Bayesian inference. Renew. Sustain. Energy Rev. **171**, 202301 (2023). [accepted]

Chapter 11
Rail Roughness Profile Identification from Vibration Data via Mixing of Reduced-Order Train Models and Bayesian Filtering

Charikleia D. Stoura, Konstantinos E. Tatsis, and Eleni N. Chatzi

Abstract The increasing demand for mobility worldwide has led to an ever-increasing expansion of railway networks. Such a rapid expansion poses a challenge to guaranteeing quality, reliability, efficiency, and, most importantly, safety of railway infrastructure. Under this perspective, continuous monitoring emerges as a promising alternative to the traditionally adopted visual inspections, or inspections via portable measuring devices, which aim at collecting geometric data for the diagnosis and prognosis of defects in tracks. Recently, railway operators worldwide have adopted the use of dedicated Track Recording Vehicles equipped with optical and inertial sensors to collect data from tracks and assess their condition. Such an approach revolutionized rail condition assessment, introducing a mobile data acquisition platform for track inspection. On the other hand, the deployment of such specialized vehicles requires disruption of regular rail service, which hinders their frequent operation and thus the continuous collection of rail data. This work aims to tackle this limitation by examining an onboard monitoring (OBM) method that hinges on collecting vibration data from in-service trains. The proposed methodology relies on the collection of acceleration data from axle boxes of trains running at normal speeds. Its novelty lies in the usage of realistic train models and the consideration of the dynamic interaction between rails and trains, which is usually simplistically ignored. The adopted train models are reduced so as to decrease the required computational effort. The identification task is based on sequential Bayesian inference for joint input and state estimation, thereby also accounting for uncertainties related to the train model. Estimating the input leads to the identification of the pertinent rail roughness profile, which can subsequently provide information on the existence of isolated defects, for example, welded joints and squats, along the track system. This study is limited to reliable prediction of the dynamics of the train–track system in the vertical direction, but proposes methods and tools of general value.

Keywords System identification · Onboard monitoring · Railway infrastructure · Bayesian inference · Hybrid modeling

11.1 Introduction

Efficient and effective monitoring of railway infrastructures is decisive for the regular maintenance of railways, which, accordingly, guarantees the quality and safety of rail transport. Traditional approaches based on visual inceptions and measurements on-site, although reliable, can no longer accommodate the increased need for monitoring along large portions of railways. Therefore, roving implementations have been widely explored during the last few years [1].

Such a roving method constitutes the use of Track Recording Vehicles. Those are vehicles equipped with optical (e.g., laser scanners, high-speed cameras) and inertial sensors (e.g., gyroscopes, inclinometers) that collect geometric data from the rails [2]. These geometric data can then deliver insight into specific irregularities of the rails. The usage of such vehicles can provide very accurate information regarding the characteristics and location of isolated defects on tracks. However, these are limited to operating while normal rail operation is suspended, thus cannot provide continuous information regarding the condition of the track [3].

A viable alternative comprises the usage of in-service trains equipped with simple monitoring systems, for example, accelerometers. Accelerometers can be mounted on different train vehicle components, such as axle boxes, bogies, and car bodies, allowing for a continuous supply of vibration data. On the other hand, railway operators increasingly improve the monitoring systems that are mounted on in-service trains, aiming primarily at vehicle condition monitoring, which can

C. D. Stoura (✉) · K. E. Tatsis · E. N. Chatzi
Institute of Structural Engineering, Department of Civil, Environmental and Geomatic Engineering, ETH Zürich, Zürich, Switzerland
e-mail: charikleia.stoura@ibk.baug.ethz.ch; tatsis@ibk.baug.ethz.ch; chatzi@ibk.baug.ethz.ch

though further facilitate the monitoring of the sustaining rail infrastructure (tracks, rail bridges). For instance, the ICN train of Swiss Federal Railways (SBB) is currently equipped with accelerometers in various locations (axle boxes, bogies, car body), while further including two pairs of tensiometric wheels, able to measure contact forces [4]. This offers an abundance of vibration-based monitoring data that need to be appropriately processed.

A first approach to handling such data pertains to the usage of signal decomposition techniques, including wavelet transform [5], or mixed filtering approaches, including Kalman, band-pass, and compensation filters [6]. These approaches rely on solely treating the obtained data without considering the dynamics of the underlying physical problem, that is, the dynamic interaction between trains and tracks. A recent study by Dertimanis et al. [3] incorporated the dynamic train–track interaction into the identification of defects on rails via acceleration data collected by the axle box of a simple train model. This study revealed promising results toward the identification of rail roughness profiles from in-service trains.

In this work, we propose an indirect approach for the identification of rail roughness profiles from data collected by traversing trains. The proposed scheme considers the physics behind the dynamic train–track interaction phenomenon [7] and couples the substructure-based dynamics with Bayesian inference methods [8] to perform identification of rail roughness [9]. The enabling tool is a realistic, three-dimensional (3D) train model running on tracks with rail roughness defined according to the German spectra for high-speed trains [10], respecting a realistic scenario setting.

11.2   Train–Track Interaction Model

The adopted 3D train vehicle is modeled with six rigid bodies corresponding to the wheelsets, bogies, and car body (Fig. 11.1). All bodies comprise five degrees of freedom (DOFs) each; two translations and three rotations. The longitudinal DOF is omitted assuming constant running speed. The equation of motion (EOM) of the train running on the track system is

$$\mathbf{m}^{\mathrm{v}}\ddot{\mathbf{u}}^{\mathrm{v}}(t) + \mathbf{c}^{\mathrm{v}}\dot{\mathbf{u}}^{\mathrm{v}}(t) + \mathbf{k}^{\mathrm{v}}\mathbf{u}^{\mathrm{v}}(t) = \mathbf{W}^{\mathrm{v}}\mathbf{p}(t) \tag{11.1}$$

where $\mathbf{u}^{\mathrm{v}}(t)$ corresponds to the response vector of the train and $\dot{\square}$ indicates differentiation with respect to time t. The mass matrix of the train is denoted by $\mathbf{m}^{\mathrm{v}}$, while $\mathbf{c}^{\mathrm{v}}$ and $\mathbf{k}^{\mathrm{v}}$ represent the damping and stiffness matrices, respectively, that correspond to the train's suspension system. The right-hand side is factorized with respect to $\mathbf{p}(t)$, the contact force vector between the train and the underlying track system, and $\mathbf{W}^{\mathrm{v}}$, which is the contact direction matrix connecting the DOFs of the train to the contact force elements. The contact force $\mathbf{p}(t)$ follows a Hertzian contact model and is expressed as

$$\mathbf{p}(t) = k^{\mathrm{H}}\mathbf{r}_{\mathrm{c}}(x) \tag{11.2}$$

where k^{H} is the contact stiffness between the train wheels and the rail, and $\mathbf{r}_{\mathrm{c}}(x)$ is the vector containing the roughness of the rails. The EOM of the train can be written in state-space form by initially transforming Eq. (11.1) to the state equation as follows:

$$\dot{\mathbf{x}}(t) = \mathbf{A}_{\mathrm{c}}\mathbf{x}(t) + \mathbf{B}_{\mathrm{c}}\mathbf{p}(t) \tag{11.3}$$

where $\mathbf{x}(t)$ is the state vector:

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{u}^{\mathrm{v}}(t) \\ \dot{\mathbf{u}}^{\mathrm{v}}(t) \end{bmatrix} \tag{11.4}$$

and $\mathbf{p}(t)$ is the input vector. Lastly, $\mathbf{A}_{\mathrm{c}}$ and $\mathbf{B}_{\mathrm{c}}$ are, respectively, the system and input matrices of the system, defined as

$$\mathbf{A}_{\mathrm{c}} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -(\mathbf{m}^{\mathrm{v}})^{-1}\mathbf{k}^{\mathrm{v}} & -(\mathbf{m}^{\mathrm{v}})^{-1}\mathbf{c}^{\mathrm{v}} \end{bmatrix}, \quad \mathbf{B}_{\mathrm{c}} = \begin{bmatrix} \mathbf{0} \\ -(\mathbf{m}^{\mathrm{v}})^{-1}\mathbf{W}^{\mathrm{v}} \end{bmatrix} \tag{11.5}$$

Assuming that acceleration measurement data are available, the output vector can be written as

$$\mathbf{y}(t) = \mathbf{C}_{\mathrm{c}}\mathbf{x}(t) + \mathbf{D}_{\mathrm{c}}\mathbf{p}(t) \tag{11.6}$$

where the output matrix $\mathbf{C}_{\mathrm{c}}$ and the feedforward matrix $\mathbf{D}_{\mathrm{c}}$ are, respectively:
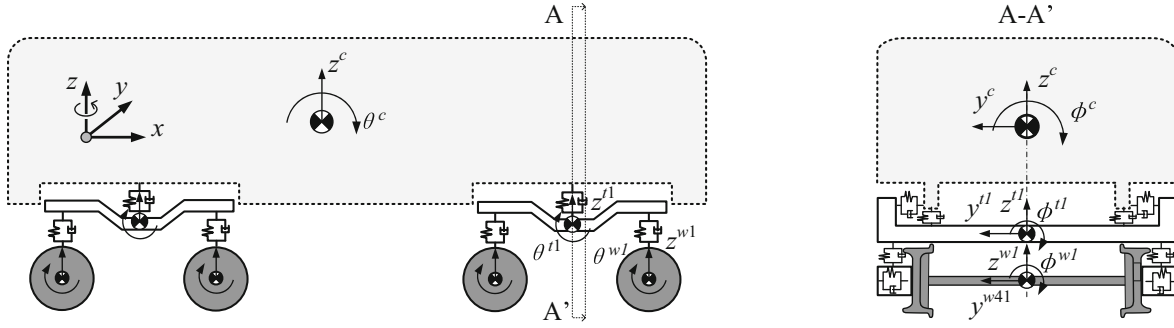
Fig. 11.1 A three-dimensional train vehicle model

$$\mathbf{C}_c = \left[-\mathbf{W}^a(\mathbf{m}^v)^{-1}\mathbf{k}^v \; -\mathbf{W}^a(\mathbf{m}^v)^{-1}\mathbf{c}^v\right], \quad \mathbf{D}_c = \mathbf{W}^a(\mathbf{m}^v)^{-1}\mathbf{W}^v \tag{11.7}$$

with $\mathbf{W}^a$ being the selection matrix of accelerations connecting the output data with the DOFs of the system.

11.3 Train Model Reduction

Train models can be rather complicated, consisting of a large number of DOFs. The number of DOFs can significantly increase the computational cost of the analysis without adding value to the identification task. To this end, a model reduction can be performed, without affecting the quality of the identified input. This study performs model order reduction based on the modal (eigenvalue) analysis of the train vehicle. Such an analysis is performed by the following factorization of the system matrix:

$$\mathbf{A}_c = \mathbf{V}\mathbf{L}\mathbf{V}^{-1} \tag{11.8}$$

where $\mathbf{L} = \mathrm{diag}\left(\mathbf{\Lambda}, \mathbf{\Lambda}^*\right)$ is a diagonal matrix consisting of the complex conjugate eigenvalues of the system and $\mathbf{V}$ is the eigenvector matrix written as

$$\mathbf{V} = \begin{bmatrix} \mathbf{\Psi} & \mathbf{\Psi}^* \\ \mathbf{\Psi}\mathbf{\Lambda} & \mathbf{\Psi}^*\mathbf{\Lambda}^* \end{bmatrix} \tag{11.9}$$

where $\mathbf{\Psi}$ is a matrix containing the modal vectors for the physical space. Note that the modal vectors of $\mathbf{\Psi}$ are complex as the system is not proportionally damped. The state-space system assumes the form

$$\dot{\boldsymbol{\zeta}}(t) = \mathbf{A}_m\boldsymbol{\zeta}(t) + \mathbf{B}_m\mathbf{p}(t) \tag{11.10}$$

$$\mathbf{y}(t) = \mathbf{C}_m\boldsymbol{\zeta}(t) + \mathbf{D}_m\mathbf{p}(t) \tag{11.11}$$

where $\boldsymbol{\zeta}(t)$ is the modal state vector. The state and measurement matrices, initially defined by Eqs. (11.5) and (11.7), become

$$\mathbf{A}_m = \mathbf{V}^{-1}\mathbf{A}_c\mathbf{V}, \quad \mathbf{B}_m = \mathbf{V}^{-1}\mathbf{B}_c, \quad \mathbf{C}_m = \mathbf{C}_c\mathbf{V}, \quad \mathbf{D}_m = \mathbf{D}_c \tag{11.12}$$

The entries of the state vector in modal basis $\boldsymbol{\zeta}(t)$ of Eq. (11.10) are decoupled. To reduce the size of the model, only a subset $\tilde{\boldsymbol{\zeta}}(t)$ of $\boldsymbol{\zeta}(t)$ is retained, while the rest of the states are truncated. Specifically, only the first few vertical and rolling, in case of different roughness in the left and right rails, modes of the train are considered as the goal is to identify rail roughness profiles in the vertical direction. Thus, the reduced-order system consists of a reduced number of modes encapsulated in $\tilde{\mathbf{\Lambda}}$ and $\tilde{\mathbf{\Psi}}$, representing the matrices of the selected eigenvalues and eigenvectors, respectively. Accordingly, the state-space matrices of Eq. (11.12) are transformed into $\tilde{\mathbf{A}}_m$, $\tilde{\mathbf{B}}_m$, and $\tilde{\mathbf{C}}_m$. The feedforward matrix $\mathbf{D}_m$ remains unchanged as it is not associated with the truncated state vector.

To discretize the reduced state-space matrices of the system, a sampling rate $F_s = 1/T_s$ is adopted, thus the discretization time is $t = kT_s$. The stochastic discrete-time state-space representation of Eqs. (11.10) and (11.11) is written as

$$\tilde{\boldsymbol{\zeta}}_{k+1} = \tilde{\mathbf{A}}_d \tilde{\boldsymbol{\zeta}}_k + \tilde{\mathbf{B}}_d \mathbf{p}_k + \mathbf{w}_k \tag{11.13}$$

$$\mathbf{y}_k = \tilde{\mathbf{C}}_d \tilde{\boldsymbol{\zeta}}_k + \mathbf{D}_d \mathbf{p}_k + \mathbf{r}_k \tag{11.14}$$

where $\mathbf{w}_k$ and $\mathbf{r}_k$ denote the discrete-time process and measurement noise terms, with known covariance matrices $\mathbf{Q}^w = \mathbb{E}\left[\mathbf{w}_k \mathbf{w}_k^T\right]$ and $\mathbf{Q}^r = \mathbb{E}\left[\mathbf{r}_k \mathbf{r}_k^T\right]$, respectively. Note that, in this case, for the discretization of the pertinent matrices $\tilde{\mathbf{A}}_d$, $\tilde{\mathbf{B}}_d$, $\tilde{\mathbf{C}}_d$, and $\mathbf{D}_d$ a first-order hold assumption is made. A first-order hold discretization assumes the input as piece-wise linear, different from the commonly adopted zero-order hold assumption that assumes piece-wise constant input. In other words, the relationship between the continuous and digital signal counterparts is based on a linear interpolation between samples as

$$\mathbf{p}(t) = \mathbf{p}_k + \frac{t - kT_s}{T}(\mathbf{p}_{k+1} - \mathbf{p}_k), \quad kT_s \le t \le (k+1)T_s \tag{11.15}$$

This assumption emerges from our effort to accurately reconstruct the input vector, which herein comprises the roughness profile (i.e., a random time-series), as indicated by Eq. (11.2), in a reliable and efficient manner. Eventually, the discretized system matrices are given as

$$\tilde{\mathbf{A}}_d = e^{T_s \tilde{\mathbf{A}}_m}, \quad \tilde{\mathbf{B}}_d = \frac{1}{T_s}\tilde{\mathbf{A}}_m^{-2}\left(\tilde{\mathbf{A}}_d - \mathbf{I}\right)^2 \tilde{\mathbf{B}}_m, \quad \tilde{\mathbf{C}}_d = \tilde{\mathbf{C}}_m, \quad \tilde{\mathbf{D}}_d = \tilde{\mathbf{D}}_m + \tilde{\mathbf{C}}_m\left[\frac{1}{T_s}\tilde{\mathbf{A}}_m^{-2}\left(\tilde{\mathbf{A}}_d - \mathbf{I}\right) - \tilde{\mathbf{A}}_m^{-1}\right]\tilde{\mathbf{B}}_m \tag{11.16}$$

11.4 Rail Roughness Profile Identification

For the identification of the roughness profile, the study follows a dual Kalman filter (DKF) approach for joint input-state estimation, as introduced in the study of Eftekhar Azam et al. [9]. This approach performs a sequential identification that first updates the input and then the state based on the updated input value, with the Kalman filter applied to both stages. To this end, for the temporal evolution of the unknown input a random walk model is adopted as follows:

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{v}_k \tag{11.17}$$

where $\mathbf{v}_k$ is the discrete-time input noise term, with covariance matrix $\mathbf{Q}^v = \mathbb{E}\left[\mathbf{v}_k \mathbf{v}_k^T\right]$. The DKF is described in detail in the study of Eftekhar Azam et al. [9], which shows that the accuracy of the identification task relies on three parameters: the process noise covariance $\mathbf{Q}^w$, the input noise covariance $\mathbf{Q}^v$, and the measurement noise covariance $\mathbf{Q}^r$. The process noise covariance matrix $\mathbf{Q}^w$ represents the accuracy of the adopted physical model, $\mathbf{Q}^v$ depends on the noise of the input, and $\mathbf{Q}^r$ depends on the accuracy of the measurement instruments used. $\mathbf{Q}^w$ and $\mathbf{Q}^r$ can be specified for a given model and measurement device. On the other hand, $\mathbf{Q}^v$ is typically adopted as the tuning parameter. Tuning is performed via an L-curve, which assumes input noise as a regularization parameter, which has to be tuned for optimizing the performance of the adopted filter [9].

11.5 Numerical Application

To validate the efficacy of the proposed approach to recover rail roughness profiles, this chapter adopts a 3D train model (Fig. 11.1) traversing a track system of known roughness. The adopted model represents a practical train with 35 DOFs in total whose properties are derived from the work of Zeng et al. [11]. The track is modeled via rigid beams with constant stiffness and the rail roughness is simulated as a stationary stochastic process with the spectral representation method [12]. The quality of roughness is defined according to the German spectra for high-speed railways [10]. Note that different roughness profiles are assumed for the left and right rails.

Simulated acceleration data are generated by employing the entire train model running on straight portions of track at a constant speed of 150 km/h. The sampling frequency of measured acceleration data is set to $f_s = 1000$ Hz. For the
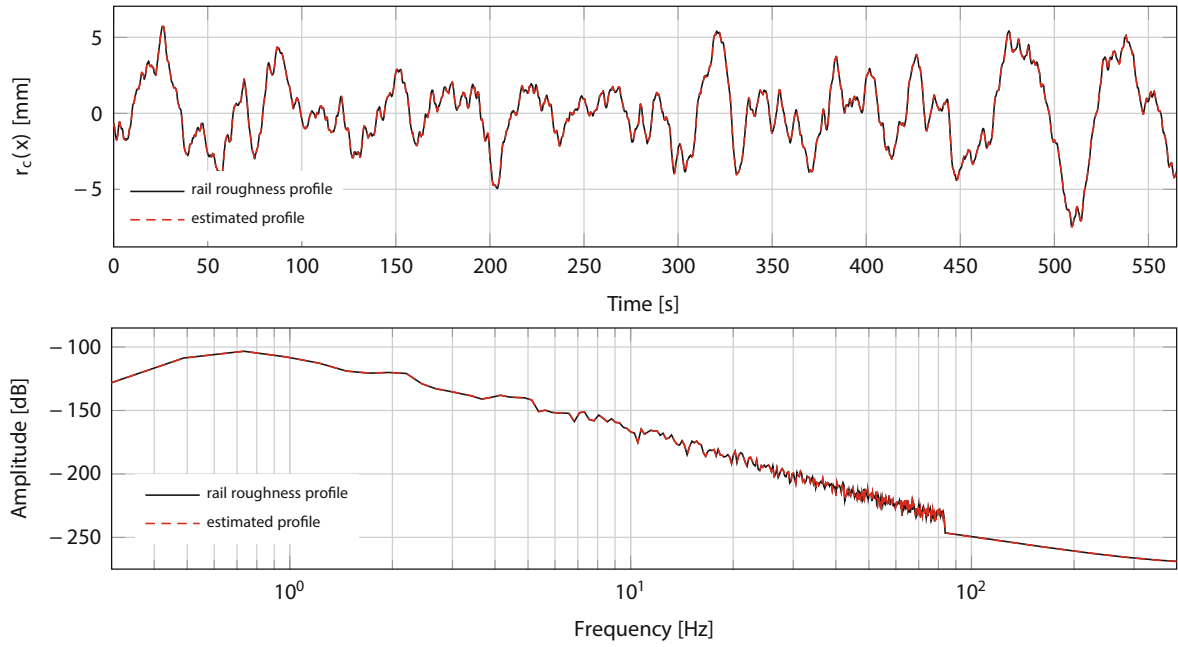
Fig. 11.2 Time-histories (top) and one-sided power spectral densities (bottom) of the true (black) and estimated via DKF (red) roughness profiles of the left rail using the complete vehicle model
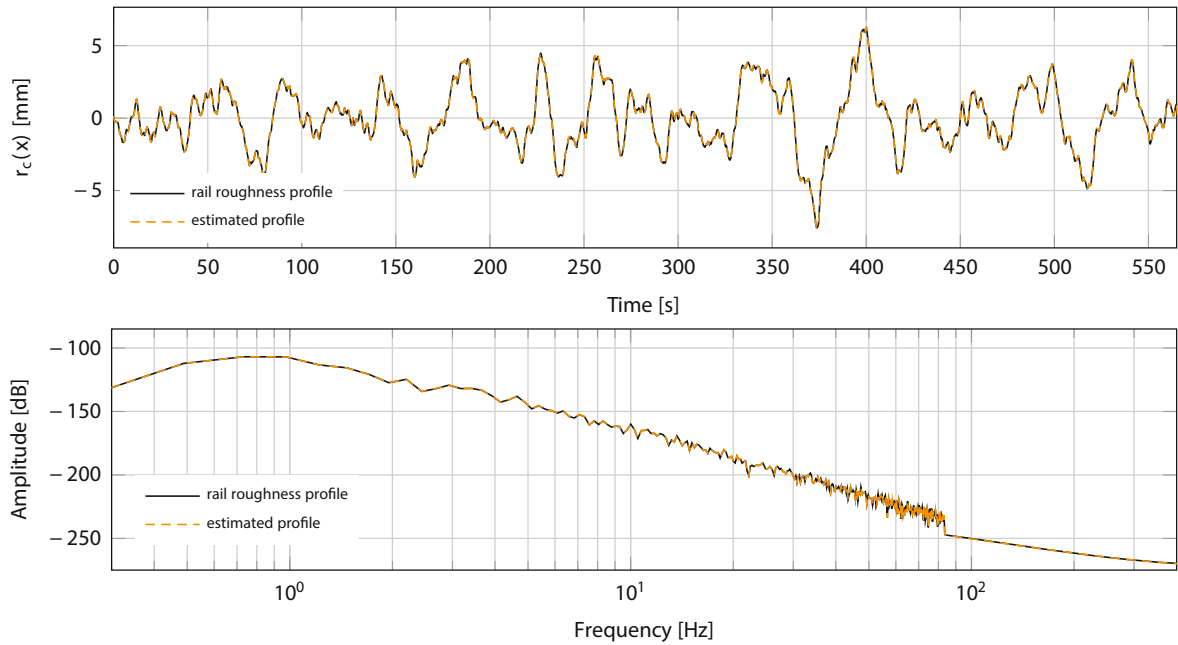
Fig. 11.3 Time-histories (top) and one-sided power spectral densities (bottom) of the true (black) and estimated via DKF (orange) roughness profiles of the right rail using the complete vehicle model

identification task, measurements of the axle boxes (above the wheels of each wheelset) are considered to be available as displacement data are usually more difficult to measure.

For the identification of rail roughness, the numerical application first considers the entire vehicle model (35 DOFs). The discretization of state-space matrices follows a first-order hold, as in Eq. (11.16), with sampling period $T_s = 10^{-3}$ s. The various noise parameters are defined next. Assuming an accurate vehicle model, the covariance matrix of the process noise is set to $\mathbf{Q}^w = 10^{-10} \cdot \mathbf{I}_1$, where $\mathbf{I}_1$ is an identity matrix with dimension equal to the number of the system states. The covariance matrix of the measurement noise of the obtained acceleration data is considered to be $\mathbf{Q}^r = 10^{-2} \cdot \mathbf{I}_2$, where $\mathbf{I}_2$
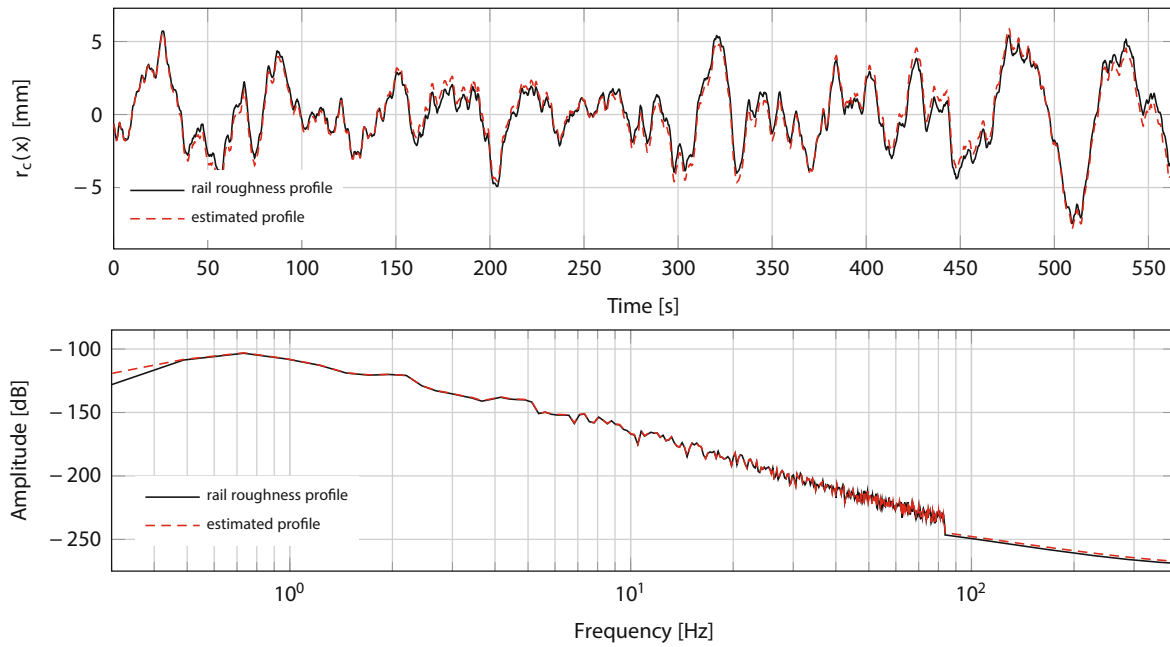
Fig. 11.4 Time-histories (top) and one-sided power spectral densities (bottom) of the true (black) and estimated via DKF (red) roughness profiles of the left rail using the reduced vehicle model
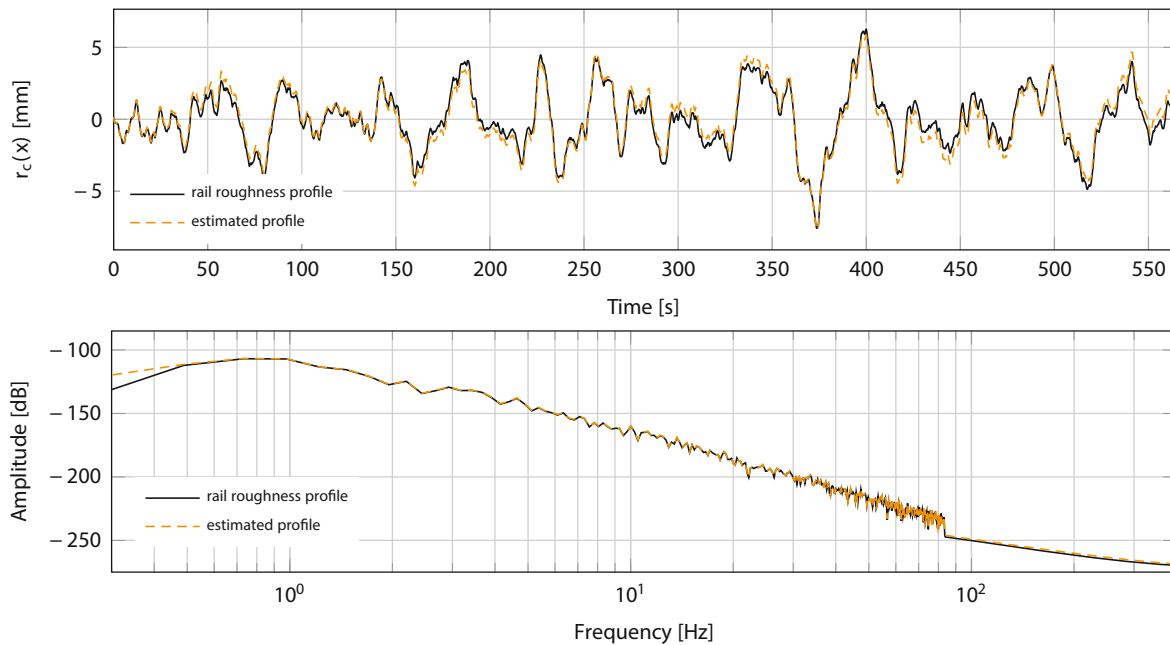
Fig. 11.5 Time-histories (top) and one-sided Power Spectral Densities (bottom) of the true (black) and estimated via DKF (orange) roughness profiles of the right rail using the reduced vehicle model

is an identity matrix with dimension equal to the number of measurements. Lastly, the input noise $\mathbf{Q}^v$ is estimated according to an L-curve. The estimated value is $\mathbf{Q}^v = 10^7 \cdot \mathbf{I}_3$, where $\mathbf{I}_3$ is an identity matrix with dimension equal to the number of inputs.

Figures 11.2 and 11.3 illustrate the true and estimated time histories and one-sided power spectral densities (PSDs) of the left and right rail roughness profiles considering the entire vehicle model for the application of DKF. There is perfect agreement between the true and estimated rail roughness profiles as no error is added to the measured data and no reduction of the model is performed. Also, according to the PSD graphs of Figs. 11.2 and 11.3 (bottom) all frequencies of interest are identified.

Following, the vehicle model is reduced via modal analysis, and the final number of modes is 16. Since the roughness of the left and right rails are different, both vertical and rolling modes need to be retained to properly capture the dynamics of the train. The number of modes can be reduced even more to further enhance computational efficiency, but the accuracy of the identification task is expected to drop. In addition, in the case that the same roughness profile for both rails is considered, the number of modes can be further decreased (to, e.g., 8 modes) as only vertical modes are necessary. Figures 11.4 and 11.5 demonstrate the true and identified rail roughness profiles of the left and right rails, respectively, in the case of the reduced vehicle model. As the selected modes capture the dynamics of the system in both vertical and rolling directions, the identified roughness profiles agree very well with the true roughness profiles of the rails. Some small discrepancies appear as the dynamics in other directions (lateral, yawing, and pitching) is ignored.

11.6 Conclusion

This study proposes a rail roughness identification scheme that adopts reduced, yet realistic, train–track models and Bayesian filtering. The adoption of realistic train models is motivated by the need for continuous monitoring of railway infrastructure that could rely on the use of in-service trains. To this end, detailed train models can enhance the accuracy and reliability of such schemes. The coupling between train and track is considered via a Hertzian contact model, while continuous contact at all times is assumed.

As practical train models may consist of hundreds of DOFs, this study additionally suggests a model reduction based on modal analysis. This decreases the number of DOFs increasing the computational efficiency of the identification task. The identification of rail roughness relies on the use of a dual Kalman filter for joint input and state estimation, which considers the roughness profile as the input of the estimation problem.

A numerical application on a 3D train traversing practical rail profiles reveals that the proposed method can accurately reconstruct the applied input, that is, the traversed rail profiles. When the entire vehicle model is adopted, there is perfect agreement between the real and identified rail profiles. The proposed approach still returns very good results in the case of a reduced vehicle model. The accuracy of the identification task relies on the number of modes finally retained.

References

1. Weston, P., Roberts, C., Yeo, G., Stewart, E.: Perspectives on railway track geometry condition monitoring from in-service railway vehicles. Veh. Syst. Dyn. **53**, 1063–1091 (2015)
2. Urda, P., Aceituno, J.F., Mu noz, S., Escalona, J.L.: Measurement of railroad track irregularities using an automated recording vehicle. Measurement **183**, 109765 (2021)
3. Dertimanis, V.K., Zimmermann, M., Corman, F., Chatzi, E.N.: On-board monitoring of rail roughness via axle box accelerations of revenue trains with uncertain dynamics. In: Model Validation and Uncertainty Quantification, vol. 3. Conference Proceedings of the Society for Experimental Mechanics Series, pp. 167–171 (2019)
4. Hoelzl, C., Dertimanis, V.K., Landgraf, M., Ancu, L., Zurkirchen, M., Chatzi, E.N.: On-board monitoring for smart assessment of railway infrastructure: a systematic review. In: The Rise of Smart Cities. Advanced Structural Sensing and Monitoring Systems (2022), pp. 223–259
5. Bocciolone, M., Caprioli, A., Cigada, A., Collina, A.: A measurement system for quick rail inspection and effective track maintenance strategy. Mech. Syst. Signal Process. **3**, 1242–1254 (2007)
6. Lee, J.S., Choi, S., Kim, S.S., Park, C., Kim, Y.G.: A mixed filtering approach for track condition monitoring using accelerometers on the axle box and bogie. IEEE Trans. Instrum. Meas. **61**, 749–758 (2012)
7. Stoura, C.D., Dimitrakopoulos, E.G.: Additional damping effect on bridges because of vehicle-bridge interaction. J. Sound Vib. **476**, 115294 (2020)
8. Tatsis, K., Dertimanis, V., Papadimitriou, C., Lourens, E., Chatzi, E.: A general substructure-based framework for input-state estimation using limited output-only measurements. Mech. Syst. Signal Process. **150**, 107223 (2021)
9. Azam, S.E., Chatzi, E.N., Papadimitriou, C.: A dual Kalman filter approach for state estimation via output-only acceleration measurements. Mech. Syst. Signal Process. **60–61**, 866–886 (2015)
10. Gao, W.W., Xia, H., De Roeck, G., Liu, K.: Integral model for train-track-bridge interaction on the Sesia viaduct: dynamic simulation and critical assessment. Comput. Struct. **112–113**, 205–216 (2012)
11. Zeng, Q., Stoura, C.D., Dimitrakopoulos, E.G.: A localized lagrange multipliers approach for the problem of vehicle-bridge interaction. Eng. Struct. **168**, 82–92 (2018)
12. Dimitrakopoulos, E.G., Zeng, Q.: A three-dimensional dynamic analysis scheme for the interaction between trains and curved railway bridges. Comput. Struct. **149**, 43–60 (2015)

Chapter 12
Optimal Sensor Placement for Developing Reliable Digital Twins of Structures

Tulay Ercan and Costas Papadimitriou

Abstract Sensor networks are mounted on structures to collect information for addressing a number of important but competing tasks involved in building a reliable digital twin from the collected data. These monitoring tasks include (1) modal identification under low vibration measurements assuming that the system can behave linearly, (2) physics-based model selection and model parameter estimation under various vibration levels activating nonlinear mechanisms at subsystem levels, (3) virtual sensing and response reconstruction over the whole body of the structure using the information from the limited number of sensors, and finally (4) structural health monitoring and damage identification (location and severity). Optimal sensor configuration (OSC) designs (type, number and location of sensors) have been developed in the past to address individual tasks, making assumptions about the loads, models and environmental conditions. However, the sensor network should be designed to collect data that are informative for all tasks simultaneously. In addition, the OSC design should be made robust to modelling, loading and environmental uncertainties. Cost issues related to budget availability for implementing and maintaining a sensor configuration should also be considered in the sensor network design. In this work, a multi-objective OSC framework based on utility functions that are built from information theoretic measures and cost considerations is presented for accounting simultaneously for the aforementioned tasks and thus using cost-effective information extracted from the physical sensing system for developing reliable digital twins. The Kullback-Liebler divergence is used to quantify the information gain from a sensor network, and heuristic algorithms to solve the multi-objective optimization problem are proposed.

Keywords Bayesian inference · Optimal experimental design · Information gain · Virtual sensing · Parameter estimation · Nonlinear structural dynamics · Multi-objective optimization

12.1 Introduction

The objective of an optimal sensor configuration (OSC) design is to maximize the quality of the data collected from a monitoring system. The instrumentation should be designed to collect data that are most informative for different and competing monitoring tasks, including model selection; model updating and parameter estimation; identification of location and magnitude of damage; as well as response reconstruction or virtual sensing of important quantities of interest that are deemed useful to evaluate the condition of structures, detect damages, and make decisions regarding structural health, safety, and performance. Realizing that an OSC for one monitoring task can be suboptimal for another task, a trade-off between the information gained from multiple tasks is needed when designing a monitoring system to be cost-effective and optimal for all the monitoring tasks.

Let $\underline{\delta}$ be a sensor configuration involving the type, location and number of sensors in a structure. Information theoretic measures [1] such as mutual information, Kullback-Leibler divergence, information entropy, joint information and value of information can be used to measure the information content in the data obtained from a sensor network installed on a structure. For a monitoring task i the information gained by the sensor configuration $\underline{\delta}$ is denoted by $\overline{U}_i\left(\underline{\delta}, \varphi\right)$. The information contained in the data depends on the type, location and number of sensors in $\underline{\delta}$. Uncertainties included in the parameter set φ arise from modelling and measurement errors, as well as environmental and operational variabilities. Cost

T. Ercan · C. Papadimitriou (✉)
Department of Mechanical Engineering, University of Thessaly, Volos, Greece
e-mail: costasp@uth.gr

of the monitoring system is also very important in the design of a sensor configuration in order to limit the overall lifetime cost associated with installing and maintaining the sensor network.

The present study uses previous developments on the information gain indices $\overline{U}_i\left(\underline{\delta}, \underline{\varphi}\right)$ for measuring the information contained in a sensor configuration for each monitoring task and presents a multi-objective methodology for designing an OSC for building a reliable digital twin of an engineering system for the purpose of monitoring its state, performance, reliability and safety. In particular, the information gain for various monitoring tasks has been studied in the literature. Information gain measures have been developed for modal identification [2], for model updating [3] and parameter estimation [1, 4] of nonlinear models of structures, for damage identification [5], and for virtual sensing using modal expansion techniques [6, 7] and sequential Bayesian techniques [1] for linear systems. The sensor network is designed in this work to be optimal for several monitoring tasks simultaneously, cost-effectively trading off the information gained for each of the aforementioned monitoring tasks.

12.2 Information Gain Accounting for Uncertainties

The parameter set $\underline{\varphi}$ introduced in $\overline{U}_i\left(\underline{\delta}, \underline{\varphi}\right)$ accounts for uncertainties in the model parameters such as stiffness and mass properties of the finite element model of the structure, model error uncertainties, as well as operational and environmental (e.g. input) uncertainties [1]. Probability distributions are used to quantify the uncertainty in the values of these model parameters. For this, the uncertain parameter vector $\underline{\varphi}$ is modelled by a prior probability distribution $\pi\left(\underline{\varphi}\right)$. Then the information gain $\overline{U}_i\left(\underline{\delta}, \underline{\varphi}\right)$ is extended to account for the uncertainty in the parameter vector $\underline{\varphi}$ so that the optimal design is robust to uncertainties involved in $\underline{\varphi}$. For this, the information contained in a sensor configuration for the i-th monitoring task is defined to be the expected information gain given by

$$U_i\left(\underline{\delta}\right) = \int \overline{U}_i\left(\underline{\delta}, \underline{\varphi}\right) \pi\left(\underline{\varphi}\right) d\underline{\varphi} \tag{12.1}$$

over all possible values of the parameter set $\underline{\varphi}$. The sources of uncertainties in the parameter set $\underline{\varphi}$ vary from excitation uncertainties to structural model and prediction error model uncertainties. The integral in (12.1) can be computed using Monte Carlo techniques.

12.3 Cost-Effective OSP for Multiple Monitoring Tasks

The OSC design $\underline{\delta}_{\mathrm{opt}}$ has to trade-off information provided for different monitoring tasks such as modal identification, structural identification, structural health monitoring, and response reconstruction (virtual sensing). It can be obtained by maximizing the normalized information gain values for each monitoring task. Herein, the OSC design is formulating as a multi-objective optimization problem of finding the optimal type and location of sensors that simultaneously maximizes the objectives

$$\underline{u}\left(\underline{\delta}\right) = \left\{u_1\left(\underline{\delta}\right), u_2\left(\underline{\delta}\right), \ldots, u_n\left(\underline{\delta}\right)\right\} \tag{12.2}$$

over all possible sensor configurations $\underline{\delta}$, where $u_i\left(\underline{\delta}\right) = U_i\left(\underline{\delta}\right) / U_{i,\mathrm{max}}$ is the normalized robust information gain and $U_{i,\,\mathrm{max}}$ is the maximum information gain that could be achieved for the monitoring task i by placing sensors at all possible sensor locations. The normalized information gain for each monitoring task guarantees that each term $u_i\left(\underline{\delta}\right)$ in Eq. (12.2) varies from 0 (no information gain) to 1 (maximum information gain). The different information gains can also be combined into a single measure of the total information gain for all tasks, as follows:

$$u\left(\underline{\delta}\right) = \sum_{i=1}^{n} w_i u_i\left(\underline{\delta}\right) \tag{12.3}$$

where the weights w_i, $i = 1, \ldots, n$, sum to 1 and they measure the contribution of each monitoring task on the OSC. Heuristic algorithms such as forward and backward sequential sensor placement (FSSP/BSSP) algorithms [8] can be employed to solve the optimization problem.

Analytical derivations [3] have shown that the information gain for each individual task increases as the number of sensors increases. As a result, the optimal number of sensors cannot be found by information gain considerations only, although after a number of sensors is optimally placed in the structure, the additional information gain is insignificant as one keeps adding sensors in the structure and this could be used as a criterion to select the optimal number of sensors. In practical applications, cost issues should be considered to select the optimal number of sensors. Specifically, the optimal number of sensors should be a trade-off between the information gain from the data and the lifetime cost of instrumentation and maintenance of the sensor system. Let $c\left(\underline{\delta}\right)$ be the total cost of a sensor configuration $\underline{\delta}$, including the cost of sensors, the installation cost and the maintenance cost over the lifetime of the sensor system. The optimal number and location of sensors is obtained as the one that maximizes the information gain vector $\underline{u}\left(\delta\right)$ and minimizes the cost $c\left(\delta\right)$. Using the total information gain in Eq. (12.3), the selection of the OSC (type, location and number of sensors) can be setup as a two-objective optimization problem of minimizing the weighted sum $u\left(\underline{\delta}\right)$ and maximizing $c\left(\underline{\delta}\right)$. The problem can be readily solved (e.g. [7]) to find the Pareto optimal solutions. Alternatively, given cost constrains (a fixed budget c_t available for designing a monitoring system), the optimal sensor configurations can be reformulated as a constrained optimization problem of maximizing the objectives $\underline{u}\left(\underline{\delta}\right)$ or the single objective $u\left(\underline{\delta}\right)$ subject to the cost constrain $c\left(\underline{\delta}\right) \le c_t$. A special case of cost consideration in optimal sensor placement (OSP) can be found in [7] for a single monitoring task.

12.4   Conclusions

The conceptual design of a cost-effective OSC is formulated as a multi-objective optimization problem that trades off the information gained for each monitoring task and the installation and maintenance cost of instrumentation. The formulation presented can account for a variety of monitoring tasks provided that an information gain index is built for each monitoring task. The tasks may include modal identification, model selection, model updating, parameter estimation, damage identification and virtual sensing. The proposed methodology can accommodate the environmental and operational uncertainties, including input as well as modelling uncertainties manifested in building information gain indices. Monte Carlo techniques can estimate the resulting probability integrals, and a number of optimization strategies are available to use for solving the resulting multi-objective optimization problem and estimate the Pareto optimal sensor configurations.

References

1. Ercan, T.: Bayesian optimal experimental design tools. Doctoral Dissertation, University of Thessaly (2022)
2. Argyris, C., Papadimitriou, C., Panetsos, P.: Bayesian optimal sensor placement for modal identification of civil infrastructures. J. Smart Cities. **2**(2), 69–86 (2016)
3. Papadimitriou, C., Beck, J.L., Au, S.K.: Entropy-based optimal sensor location for structural model updating. J. Vib. Control. **6**(5), 781–800 (2000)
4. Udwadia, F.E.: Methodology for optimum sensor locations for parameter identification in dynamic systems. J. Eng. Mech. **120**(2), 368–390 (1994)
5. Capellari, G., Chatzi, E., Mariani, S.: Structural health monitoring sensor network optimization through Bayesian experimental design. ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A: Civil Eng. **4**(2), 04018016 (2018)
6. Ercan, T., Papadimitriou, C.: Optimal sensor placement for reliable virtual sensing using modal expansion and information theory. Sensors. **21**(10), 3400 (2021)
7. Mehrjoo, A., Song, M., Moaveni, B., Papadimitriou, C., Hines, E.: Optimal sensor placement for parameter estimation and virtual sensing of strains on an offshore wind turbine considering sensor installation cost. Mech. Syst. Signal Process. **169**, 108787 (2022)
8. Papadimitriou, C.: Optimal sensor placement methodology for parametric identification of structural systems. J. Sound Vib. **278**(4–5), 923–947 (2004)

Chapter 13
DataSEA: Mature, Modern Data Management Enabling Sustainable Data Strategy

Justin Wu and Stephen C. Jackson

Abstract We present DataSEA, a mature data management platform that emphasizes the role of metadata in contextualizing and finding data through an extended evidentiary cycle and complements existing report-oriented documentation practices to produce sustainable, long-lived data strategy.

Keywords Data documentation · Data management · Data strategy · Data fusion · Sustainable data

13.1 Introduction

Vast amounts of data are generated as a by-product of the product lifecycle, reflecting the scope of engineering effort and design decisions that must be made to take a product from conception to production and maintenance. It can be said that the extent to which an organization is able to consistently make informed decisions is a function of its ability to make effective use of its data. When properly leveraged, this data provides critical insight into the processes of design and development and an opportunity to engage in future development cycles more efficiently and effectively [1].

Unfortunately, advances in technology allowing for ever-more-rigorous testing and extensive data capture can be a double-edged sword. That is to say, the greater the scope of data generated and archived during the development process, the larger the task in ensuring that data is effectively organized and documented [2]. When handled poorly, wide-scale data capture, far from serving to improve and streamline existing processes, instead becomes an obstacle to development and evaluation efforts, compelling engineers to devote significant time and energy to data management tasks—locating and wrangling data files—instead of to design and analysis efforts.

Simultaneously, in contrast with the promise offered by modern data technologies and processes, owing to historical technological limitations and consequently immature data practices, documentation practices have not evolved much in recent years. Instead, human-curated technical reports have served as the dominant medium for documenting testing performed during the development process. We assert, however, that these outmoded documentation practices that understate the importance of long-lived evidentiary data relative to report-based documentation affect fragility in the face of unforeseen scenarios.

More specifically, reports effectively summarize critical data points needed to inform and drive decision-making. However, the tailored, summative nature of report-based documentation forces a focus on the performance of designed systems within specific design parameters. When post hoc expansions of system scope broaden the range of environments in which a system or its successors are expected to perform, the potential of report-based documentation methods—and their heightened focus on previously forecast scenarios—to evaluate the applicability of existing results to new scenarios is limited.

There is a need to supplement existing report-based documentation practices with modern data practices that provide direct access to both generated data and the metadata that contextualizes them [3]. In particular, because both anomalies and scope expansion can occur years after testing is first performed, steps must be taken to ensure that data not only persists, but remains useful—that is, effectively informs decision-making—over a long lifetime.

For the above reasons, we argue that employing sustainable data practices is critical to ensuring that we get the most out of the vast amounts of data generated by modern engineering practices. To that end, we present DataSEA, a mature

J. Wu (✉) · S. C. Jackson
Sandia National Laboratories, Albuquerque, NM, USA
e-mail: juswu@sandia.gov; sjacks@sandia.gov

data management platform that emphasizes the role of metadata in contextualizing and finding data through an extended evidentiary cycle and complements existing report-oriented documentation practices to produce sustainable, long-lived data strategy.

13.2 DataSEA

DataSEA (Data for Systems Engineering Applications) is a data management platform built from the ground up to ensure the continued viability of test data through an extended lifecycle. It emphasizes organizing data into flexible, purposeful architectures, encouraging users to focus on who needs the data in question and how they can best structure data capture to achieve goals of findability and reproducibility.

Relevantly, DataSEA targets three pillars of data management as a motivating philosophy:

1. **Data capture should follow a "metadata-first" approach.** This ensures data is always contextualized by key metadata, facilitating both findability and reproducibility, which are necessary for long-term viability.
2. **All data must meet standards and follow protocols for configuration control.** This means that provenance for all data stored must be tracked and changes properly recorded.
3. **There must be full traceability for all data products.** This begins with connecting raw data products to analysis results and extends to data collected after production and into the maintenance phase.

13.2.1 Flexible, Purposeful Data Architecture

DataSEA acts as an authoritative, central repository for data collected during testing, standardizing the answers of who owns data, where it lives, and how it is meant to be consumed.

Entities in DataSEA can be divided into three categories: schemas, records, and data frames, depicted in Fig. 13.1.

Schemas are the core building block of DataSEA, and the key to its data organization architecture, laying out the strategy for how data and metadata will be collocated. Schemas are user-defined collections of metadata that annotate and contextualize raw data capture and characterize the structure that records must follow at instantiation. Schema provenance is tracked, meaning users have access to the full change history of a schema.

Records represent discretized instances of data capture and follow the structure outlined by their parent schema in their usage. As with schemas, record provenance is fully tracked.

Data frames enforce standardized access to raw data in an accessible format. DataSEA uses converters to transcode data stored in proprietary formats to formats that can be used by popular tools such as MATLAB or Python.

13.2.2 A "Self-Documenting" Data Approach

We achieve a "self-documenting" data approach with a multifaceted strategy with the following components:

- DataSEA utilizes a data hierarchy centered on metadata that spurs users to methodologically organize their data in terms of what will make it the easiest to find and understand.
- DataSEA traces the evolution of data by tracking changes and why they occurred. Each revision is stored, accessible, and interactable. This allows users to freely experiment with data capture until they are able to settle on an approach that works best for them, documenting the reasoning behind such changes all the while.
- DataSEA provides a mechanism for explicitly linking associated data. This facilitates traceability from raw data to analysis and beyond by allowing data products to be explicitly linked to related constructs, such as calibration factors, hardware assemblies, and analysis results. This enables both reproduction efforts as well as anomaly-finding activities.

Fig. 13.1 Left: a diagram of the structure and relationship of schemas, records, and data frames in DataSEA. Right: an example of the structure instantiated in a particular capture instance

13.2.3 Ensuring the Data Remains at the Forefront in Decision-Making

DataSEA makes directly engaging with data easier for users at many points in the product development lifecycle. It provides a robust browser interface and API to allow analysts to connect directly with the data. DataSEA's provenance history ensures that the data used for analysis is traceable to the source and can be used without fear. This enables interactive data reviews via tools like MATLAB Live Scripts or Jupyter notebooks. It also allows multiple data sources to be connected together for direct comparison.

13.3 Conclusion

Advances in product development technology and data capture have provided opportunities to shorten development time and improve product delivery. Taking advantage of these opportunities, however, is reliant on effective data management practices that account for the difficulties of data wrangling and the limitations of existing documentation practices.

   We present DataSEA, a data management platform designed from the ground up to enable better, more-sustainable data strategy. It utilizes a novel approach for storing discrete evidence generated during product development. This ensures data is properly organized and remains long-term-viable without causing undue burden on data producers and consumers.

References

1. Feng, Y., Zhao, Y., Zheng, H., Li, Z., Tan, J.: Data-driven product design toward intelligent manufacturing: a review. Int. J. Adv. Robot. Syst. **17**(2), 1729881420911257 (2020)
2. Gray, J., Liu, D.T., Nieto-Santisteban, M., Szalay, A., DeWitt, D.J., Heber, G.: Scientific data management in the coming decade. ACM Sigmod Record **34**(4), 34–41 (2005)
3. Beagrie, N.: Digital Curation for Science, Digital Libraries, and Individuals (2006)

Chapter 14
Optimal Sensor Placement Considering Operational Sensor Failures for Structural Health Monitoring Applications

Mayank Chadha, Yichao Yang, Zhen Hu, and Michael D. Todd

Abstract A structural health monitoring (SHM) system acquires sensor measurements from which a structural state can be inferred. An updated understanding of the structural state is crucial in making appropriate maintenance decisions over the life cycle of the structure. However, the inferred structural state may be incorrect if the sensing system that initiates the SHM workflow is unreliable. The operational and environmental conditions that these sensors can face, in addition to normal manufacturing defects, result in varying functionality at different monitoring locations, at different times. Therefore, it is important to account for sensor reliability in the optimal sensor design process for the SHM system at the outset. In this chapter, we propose an optimal sensor design framework that accounts for the time-dependent reliability of the sensor network over the life cycle of the structure. The targeted objective function (Bayes risk) must consider the consequence of unreliable measurements over time, uncertainties in loading, sensor readings, and bias. This makes the Bayes risk a multidimensional integral with a non-linear integrand. The algorithm deploys the Bayesian optimization technique in tandem with univariate dimensional reduction and Gaussian-Hermite numerical approximation of the Bayes risk that catalyzes efficient numerical implementation of an otherwise computationally exhaustive process. We consider monitoring of a miter gate as the demonstration example and focus on the inference of an unknown and uncertain state-parameter(s) (i.e., damage from the loss of contact between the gate and wall, the "gap") from the acquired sensor data.

Keywords Bayesian optimization · Sensor reliability · Structural health monitoring · Miter gate · Uncertainty quantification

14.1 Introduction

This chapter describes a sensor optimization framework with consideration of spatial and time-dependent sensor reliability. That is, the framework considers the possibility of sensors malfunctioning over time. The goal here is to design a reliable sensor network such that the measurements lead to a reliable inference of the damage state over the life cycle of the structure even in a situation where some sensors in the network have malfunctioned. This research is built on the Bayesian optimization-based sensor placement framework that we had developed in our previous works (see Yang et al. [1, 2]). A well-designed data acquisition system leads to an improvement in the Value of Information (see Chadha et al. [3]) and is crucial for reliable decision-making (see Chadha et al. [4]). The sensor optimization algorithm is applied to a complex real-world miter gate structure where different parts of the structure are exposed to different environmental conditions that dynamically change over time.

M. Chadha · Y. Yang · M. D. Todd (✉)
Department of Structural Engineering, University of California San Diego, La Jolla, CA, USA
e-mail: mdtodd@ucsd.edu

Z. Hu
Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, MI, USA

14.2 Problem Definition

We consider three scenarios of sensor reliability. In scenario 1, the sensor is perfectly functional (hence yields reliable readings) but has standard measurement noise. In scenario 2, the sensor is partially malfunctioning. Therefore, along with the standard measurement noise, the sensor readings also suffer from reliability bias. In scenario 3, the sensor has completely failed. In most cases, the third scenario can be easily recognized and rectified by an on-site repair/replacement. We focus on the second scenario of accounting for sensor reliability over the life cycle of the structure in an optimal sensor placement framework.

Consider a sensor design $e \in \Omega_E$ consisting of $N_{sg}(e)$ number of sensors with the measurement $x_e(t) \in \Omega_{X_e(t)}$. Over its life cycle, a structure is subjected to uncertain loading, denoted by the random vector $H(t)$ with a realization $h(t) \in \Omega_{H(t)}$. Let $\Theta(t)$ denote a random variable representing the state/damage parameter vector at any time t with a realization denoted by $\theta(t)$. Let the random vector $\zeta_e(t)$ represent the measurement noise (composed of standard observation noise and reliability bias), with a realization denoted by $\varepsilon_e(t) = (\varepsilon_{e1}(t), \varepsilon_{e2}(t), \cdots, \varepsilon_{eN_{sg}(e)}(t)) \in \Omega_{\zeta_e(t)}$. We consider a general case where the structure is divided into various reliability zones depending on the parts of structures exposed to different environmental conditions [modeled by the loading term denoted by $h(t)$]. Let $\Omega_S = \{s_{unrel}, s_{rel}\}$ denote the set of functional states of the sensor, where s_{unrel} denotes that a sensor is malfunctional and unreliable (scenario 2) and s_{rel} represents a fully functional and reliable sensor (scenario 1). The probability mass functions $P_{S_{ei}(t)|H(t)}(s_{unrel}|h(t))$ and $P_{S_{ei}(t)|H(t)}(s_{rel}|h(t)) = 1 - P_{S_{ei}(t)|H(t)}(s_{unrel}|h(t))$ denote the probability of the i-th sensor malfunction or fully functional at time t conditioned upon the loading zone [defined by the load vector $h(t)$ and the sensor location of the i-th sensor in sensor array design e]. Let $f_{\zeta_{ei}(t)|H(t)}(\varepsilon_{ei}(t)|h(t))$ denote the distribution of observation noise in the i-th sensor conditioned upon the loading zone at which the sensor is installed (which is assumed to be defined in terms of various loading situations). The uncertainty in the measurements is contributed by two effects: (a) the standard observation noise of scenario 1 and (b) the reliability bias as a consequence of the sensor's partial malfunction. That is,

$$\varepsilon_{ei}(t) = \begin{cases} \overline{\varepsilon}_{ei}(t) & \text{for sensor functional state } s_{rel}; \\ \overline{\varepsilon}_{ei}(t) + n_{ei}(t) & \text{for sensor functional state } s_{unrel}. \end{cases} \qquad (14.1)$$

Here, $n_{ei}(t)$ is a realization of the random variable η_{ei} (with the mean $\mu_{n_{ei}(t)}$ and standard deviation $\sigma_{n_{ei}(t)}$) that models the sensor reliability bias in the i-th sensor of the design e. $\overline{\varepsilon}_{ei}(t)$ is the standard observation noise for scenario 1 and it has a standard deviation of $\sigma_{\overline{\varepsilon}_{ei}(t)}$. The distribution of the measurement noise is then given by:

$$f_{\zeta_{ei}(t)|H(t)}(\varepsilon_{ei}(t)|h(t)) = P_{S_{ei}(t)|H(t)}(s_{rel}|h(t)) \frac{1}{\sigma_{\overline{\varepsilon}_{ei}(t)}} \phi\left(\frac{\varepsilon_{ei}(t)}{\sigma_{\overline{\varepsilon}_{ei}(t)}}\right) +$$
$$P_{S_{ei}(t)|H(t)}(s_{unrel}|h(t)) \frac{1}{\sqrt{\sigma_{\overline{\varepsilon}_{ei}(t)}^2 + \sigma_{n_{ei}(t)}^2}} \phi\left(\frac{\varepsilon_{ei}(t) - \mu_{n_{ei}(t)}}{\sqrt{\sigma_{\overline{\varepsilon}_{ei}(t)}^2 + \sigma_{n_{ei}(t)}^2}}\right). \qquad (14.2)$$

We obtain the observed sensor readings $x_e(t)$ by adding the noise vector $\varepsilon_e(t)$ to the exact/ground-truth value of the sensor measurement obtained using the FEM model $g_e(\theta_{true}(t), h(t); t)$. That is,

$$x_e(t) = g_e(\theta_{true}(t), h(t); t) + \varepsilon_e(t). \qquad (14.3)$$

Obtaining optimal sensor design using Bayesian optimization requires accessing the posterior distribution of the damage parameter $\theta(t)$ for a given measurement reading $x_e(t)$ numerous times. The posterior distribution is obtained using Bayes theorem as $f_{\Theta(t)|X_e(t),H(t)}(\theta(t)|x_e, h) \propto f_{X_e(t)|\Theta(t)}(x_e|\theta) \cdot f_{\Theta(t)}(\theta(t))$. The likelihood is obtained using the measurement model for x_e defined in Eq. (14.1) and the observation noise structure defined in Eq. (14.2) as:

$$f_{X_e(t)|\Theta(t)}(x_e|\theta) = \prod_{i=1}^{N_{sg}(e)} \left(P_{S_{ei}(t)|H(t)}(s_{rel}|h(t)) \frac{1}{\sigma_{\overline{\varepsilon}_{ei}(t)}} \phi\left(\frac{\varepsilon_{ei}(t)}{\sigma_{\overline{\varepsilon}_{ei}(t)}}\right) \right.$$

$$\left. + P_{S_{ei}(t)|H(t)}(s_{unrel}|h(t)) \frac{1}{\sqrt{\sigma_{\overline{\varepsilon}_{ei}(t)}^2 + \sigma_{n_{ei}(t)}^2}} \phi\left(\frac{\varepsilon_{ei}(t) - \mu_{n_{ei}(t)}}{\sqrt{\sigma_{\overline{\varepsilon}_{ei}(t)}^2 + \sigma_{n_{ei}(t)}^2}}\right) \right). \qquad (14.4)$$

We use particle filtering technique for Bayesian inference of damage parameter $\theta(t)$ given the measurements $x_e(t)$.

14.3 Objective Function Focused on Reliable Sensor Measurements

We define the risk of sensor bias at time t as the absolute deviation between the true state and the mean of the posterior:

$$\mathcal{L}\left(\theta_{\text{true}}(t), h(t), \varepsilon_e(t); t\right) = \left|\mu_{\Theta(t)|X_e(t), H(t)} - \theta_{\text{true}}(t)\right|. \tag{14.5}$$

The expected risk of sensor bias at time t is defined as follows:

$$\mathfrak{E}(e; t) = \iiint f_{\Theta(t)}\left(\theta(t)\right) f_{H(t)}\left(h(t)\right) f_{\zeta_{ei}(t)|H(t)}\left(\varepsilon_{ei}(t)|h(t)\right) \mathcal{L}\left(\theta_{\text{true}}(t), h(t), \varepsilon_e(t); t\right) \, d\theta(t) dh(t) d\varepsilon_e(t). \tag{14.6}$$

Finally, the aggregate expected risk of sensor bias over the life cycle is given as:

$$\mathfrak{E}_{\text{LC}}(e) = \int_{\Omega_T} \mathfrak{E}\left(e; t\right) \, dt = \sum_{k=1}^{t_k \in \Omega_T} \mathfrak{E}(e; t_k). \tag{14.7}$$

We obtain the optimal sensor design e^* using $\mathfrak{E}_{\text{LC}}(e)$ as the objective functional by deploying the Bayesian optimization algorithm described in Yang et al. [1, 2], such that:

$$e^* = \underset{e}{\text{argmin}} \mathfrak{E}_{\text{LC}}(e) \tag{14.8}$$

14.4 Results

Figure 14.1 shows the optimal sensor network design e^* obtained using Eq. (14.8) and the optimization algorithm delineated in Yang et al. [1, 2].

We observe that in the reliability-focused design there are more sensors above the mean downstream water head than the number of sensors below it. This is because the probability of sensors malfunctioning is higher when they are located below the mean downstream water head (higher likelihood of being in the submerged zone) than when they are installed above the mean water head (higher likelihood of being in the splash zone). The sensors are strategically placed in the gap's neighborhood allowing for a realistic inference of the gap length and at the same time, collectively, sensors spend a higher average time in the splash zone over the life span of the structure, such that if the submerged sensors malfunction, the sensors in the splash zone can carry the burden of performing acceptable inference over the life cycle of the miter gate.

Fig 1a: Rendered front view   Fig 1b: Rendered side view
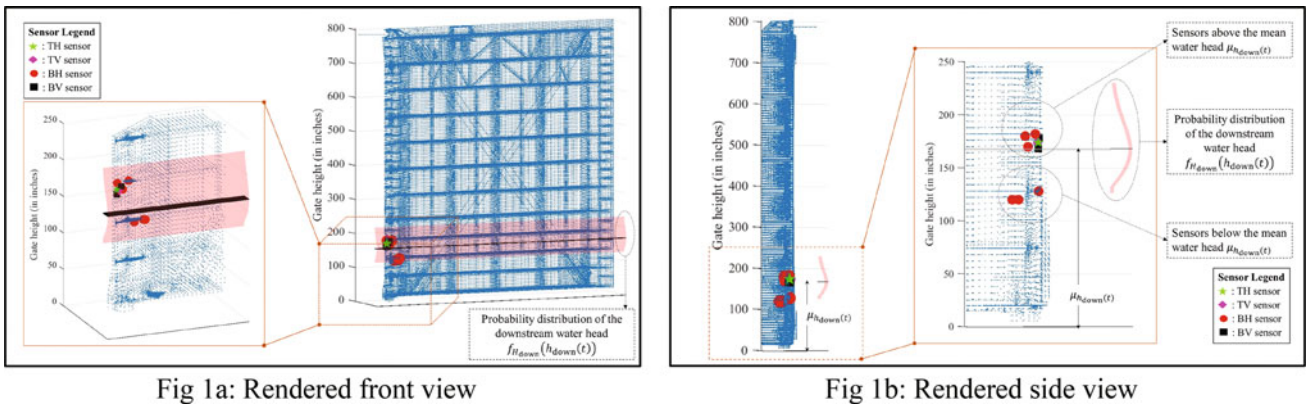
Fig. 14.1 Miter gate and the optimal sensor network design for $\mu_{n_{ei}(t)} = 10^{-4}$ and $\sigma_{n_{ei}(t)} = 2 \times 10^{-6}$. (**a**) Rendered front view. (**b**) Rendered side view

14.5   Conclusions

This chapter briefly details the mathematical formulation behind a sensor optimization framework with a dual target: (1) the design obtained should lead to damage inference to an acceptable degree of accuracy; (2) the framework must consider all the uncertainties that the system is subjected to and account for the possibility of sensors malfunctioning over time. The reliability-focused designs lead to inference results that are overall reliable, consistent, representative of true gap evolution over time, and hence lead to improved Value of Information relative to random design.

References

1. Yang, Y., Chadha, M., Hu, Z., Todd, M.D.: An optimal sensor placement design framework for structural health monitoring using Bayes risk. Mech. Syst. Signal Process. **168**, 108618 (2022)
2. Yang, Y., Chadha, M., Hu, Z., Parno, M., Todd, M.D.: A probabilistic sensor design approach for structural health monitoring using risk-weighted f-divergence, vol. 161, p. 107920. Mech. Syst. Signal Process. (2021)
3. Chadha, M., Hu, Z., Todd, M.D.: An alternative quantification of the value of information in structural health monitoring. Structural Health Monitoring: Value of Information Perspective, SAGE. **21**, 138 (2021)
4. Chadha, M., Ramancha, M.K., Vega, M.A., Conte, J.P., Todd, M.D.: The modeling of risk perception in the use of structural health monitoring information for optimal maintenance decisions. Reliab. Eng. Syst. Saf. **229**, 108845 (2022)

Chapter 15
Sequential Harmonic Component Tracking for Underdetermined Blind Source Separation in a Multitarget Tracking Framework

Romain Delabeye, Martin Ghienne, and Jean-Luc Dion

Abstract Smart factories are composed of heterogeneous cyber-physical systems. In light of their complexity and the lack of transparency in their design, monitoring the health of these machines in real time is made possible by the use of nonintrusive sensors. Such sensors produce mixed signals capturing component-specific signatures. Retrieving the activation statuses of the components (over the different operating modes of a machine) is essential for estimating their associated performance indicators. This is a special case of underdetermined blind source separation (UBSS), yet a sensor fusion perspective is adopted in this chapter. A harmonic component detector produces observations in the time-frequency (TF) domain, inherently entailing noise-induced false alarms. The main contribution of this chapter consists of a clutter-resilient multiharmonic component tracking algorithm, based on the sequential Monte Carlo probability hypothesis density (SMC-PHD) filter. Additionally, this chapter presents a track association algorithm adapting the results obtained in the multitarget tracking framework for unsupervised multilabel classification. The combination of the two algorithms mitigates typical difficulties encountered in traditional UBSS problems, such as nonstationary and partially coupled mode decomposition. The performance of the proposed technique is assessed on synthetic data.

Keywords Harmonic component tracking · Multitarget tracking · Sensor fusion · Underdetermined blind source separation · SMC-PHD filter

15.1 Introduction

Energy sustainability is one of the greatest challenges faced by the manufacturing industry. The manufacturing industry is energy-intensive by nature, making it worthwhile to put the emphasis on energy efficiency when aiming for substantial discounts in energy usage and associated carbon emissions. From a physical point of view, energy efficiency boils down to minimizing dissipated energy for a given production. General indicators such as the specific energy consumption (SEC), that is, the total energy consumption per unit of output, only allow for a shallow analysis of a system's energy efficiency. A key success factor in enhancing a production system's energy sustainability lies in the ability to allocate energy performance indicators (EnPI) to dedicated active components, actuators, and operating modes, designated as *components*, *actuators*, and *operations*, respectively. In this context, an actuator consists of a group of physical components always active simultaneously (e.g., a rotor and bearings), and an operation relates to the accomplishment of a task using a fixed group of actuators (e.g., drilling would use two motors to rotate and advance the drill). A machine thus obeys the same dynamics throughout an operation. This dynamics is more specifically made up of the actuators' dynamics and possible couplings between them. In practice, though it is common to monitor a manufacturing machine's total energy consumption (a mandatory requirement to compute EnPI), the activation sequences of (i) the actuators composing this machine and (ii) the different operations performed by the machine are seldom available. Hence, the times at which components, actuators, and operations are active need to be inferred from sensor data, without any prior regarding the studied system (no physical or process models). Furthermore, in order for such a process identification technique to scale, the use of nonintrusive sensors is preferred (e.g., accelerometers or current sensors). These sensors have the particularity of sensing much information from multiple remote sources, resulting in coupled dynamics from a sensor's point of view. Signal processing is thus required to uncouple these sources. This constitutes an underdetermined blind source separation (UBSS) problem, yet only activation statuses are sought

R. Delabeye (✉) · M. Ghienne · J.-L. Dion
Department of Vibrations, Acoustics and Structures, Quartz Laboratory, Saint-Ouen, France
e-mail: romain.delabeye@isae-supmeca.fr; martin.ghienne@isae-supmeca.fr; jean-luc.dion@isae-supmeca.fr

rather than mixed signal. By putting the emphasis on the activation statuses of the components, actuators, and operations rather than source signal recovery, the latter aspect can be performed using an independent physics-informed regression algorithm instead of a statistical decomposition. Only the former aspect is considered here. Moreover, data are represented in the time-frequency (TF) domain.

UBSS problems are traditionally tackled using decomposition algorithms either identifying a mixing matrix and source signals or learning a sparse representation from a dictionary of representative vectors built iteratively. Such processes (subspace methods in particular) are very efficient when data are piecewise stationary [1], yet this assumption is too restrictive in this context since this would not cover controlled systems.

This motivates the use of multitarget tracking (MTT) for estimating both the state and the number of active components in a signal. The main contribution of this chapter hence consists of an MTT formulation allowing for tracking harmonic components over time. We propose a simple peak-based harmonic component detector (stemming from the signal's power spectral density). A particularity of the problem on hand is the time-varying number of false alarms per scan, which depends on the level of noise associated with each source signal. We hence propose a feature-aided tracking (FAT) formulation, based on the spectral kurtosis, increasing the clutter resilience of our tracking filter.

Related works focusing on MTT and frequency tracking are presented in the first section. Problem formulation is detailed in the second section, together with background knowledge regarding the techniques used in this chapter. The enhancements made to the tracker in order to increase its resilience to clutter are presented in the third section. The fourth section details how this MTT formulation is adapted to solve an unsupervised multilabel classification problem. The verification of the developed technique is performed on synthetic data; the results are presented and discussed in the fifth section.

15.2   Related Works

In blind source separation, independent component analysis (ICA) and its extension to the underdetermined case [2] have received great attention over the years, yet this method cannot separate sub-Gaussian distributions and is not well suited to discrete event data. Diverse techniques exist to estimate the number of source signals [3], often based on subspace methods. For the problem on hand, the number of actuators cannot be recovered using these methods [4], but the number of distinct operations can be retrieved. A traditional approach to tackle UBSS problems consists of clustering time-series data into chunks over which the number of sources is assumed to be constant; the signal is then factorized into a mixing matrix and unmixed source signals [5–7]. Another formulation consists of factorizing data as a dictionary of atoms (representing relevant modes) and a representation (linear combination of atoms) [8]. For these linear combinations to truly represent the sought labels, the representation must remain binary as in the semi-binary non-negative matrix factorization [9, 10]. Alternatively, dynamic time warping (DTW) can be coupled with hidden Markov models when clustering signals with different shapes [11, 12].

In this chapter, we adopt a multitarget tracking framework in which the number and states of target harmonic components are sought. Multifrequency tracking has been extensively studied in the literature as a data assimilation problem using Kalman filters [13, 14] and particle filters [15, 16] in particular. In these applications, tracks are initiated heuristically, and tracking is then treated solely as a state estimation problem.

Two additional challenges arise in the class of applications considered in this chapter. First, the number of harmonic components to track is not known a priori, evolves over time, and is not necessarily detected at each time step. A second important aspect is the presence of a time-varying number of false alarms (clutter). Suitable real-time compatible data association filters (responsible for mitigating the effect of clutter on state estimation performance) include the joint probabilistic data association (JPDA) filter, probabilistically associating measurements to tracks, and probability hypothesis density (PHD) filters [17, 18], implicitly fusing all states with all measurements at each time step. In between these two types of filter, a set JPDA has been proposed in [19].

With a view to increasing state estimation performance, feature-aided tracking has been investigated in the literature. The association probability of the JPDA was refined using signal-to-noise ratio (SNR) [20], radar high-resolution range (HRR) [21], and wavelet-based spectral features [22]. Target Doppler and down-range extent were implemented in feature-aided PHD filters [23, 24]. In these applications, the ingenious integration of feature information—essentially in the data association part of the filter—resulted in better tracking performance and exhibited clutter-resilient behaviors.

15.3 Problem Formulation

Among existing frameworks, the MTT formulation has the potential to track time-varying spectral components while detecting when a new component appears or disappears, thus alleviating major limitations in subspace decomposition methods.

As a machine operates through its manufacturing process, an actuator a with status $\delta_t^a \in \{0, 1\}$ at a time step t can switch on (1) or off (0) components with status $\delta_t^c \in \{0, 1\}$, producing sudden changes in sensor data, and bringing the machine into a new operating mode with status $\delta_t^o \in \{0, 1\}$. From a set of H_t harmonic components (targets) with states $\{x_t^h\}_{h=1}^{H_t}$ at time step t in the TF domain, a set $\{z_t^m\}_{m=1}^{M_t}$ of M_t measurements is produced from a single sensor. Targets follow a Markov transition model $p(x_t^h|x_{t-1}^h)$.

A simple harmonic component detector is proposed. The discrete Fourier transform (DFT) is evaluated on successive overlapping windows with time index t. Then a detection $z_t^m = \left[a_t^m, b_t^m, A_t^m, \omega_t^m\right]^T + w_t$ is triggered every time the power spectral density reaches a peak above a threshold τ, where $w_t \sim \mathcal{N}\left(.; 0; diag([\sigma_a, \sigma_b, \sigma_A, \sigma_\omega])\right)$; A_t^m, ω_t^m, a_t^m, b_t^m are a harmonic component's amplitude, pulse, and complex coefficients, respectively, and σ_A, σ_ω, σ_a, σ_b denote their corresponding standard deviations. This measurement m corresponds to a complex yet undamped modulated sine wave supplemented with Gaussian noise $w_t^s \sim \mathcal{N}(.; 0; \sigma_s)$, i.e., $s_t^m = A_t^m \exp\left(i\,\omega_t^m t\right) + w_t^s = a_t^m + i\,b_t^m$. An immediate consequence of this detection technique is the time-varying nature of the number of false alarms per scan λ_{FA}. In this formulation, λ_{FA} is implicitly defined by the noise level σ_s. The underlying detection probability p_D is unknown, yet close to unity.

The trajectory $T_{t_i:t_f}^h$ of a target h is made of associated states x_t^h (defined by the same variables as the measurement vector) between times t_i and t_f. For simplicity, the same target index h is kept over time.

For interpretation, a trajectory represents the behavior of a physical component, unless its frequency can be expressed as a positive integer multiple of another trajectory's (i.e., as a harmonic of the fundamental frequency).

Formally, an occurrence of an operation o thus corresponds to a set of trajectories $\boldsymbol{\Omega}_{t_i^o:t_f^o}^o = \left\{T_{t_i^o:t_f^o}^h\right\}_{h=1}^{H^o}$ made of H^o components between times t_i^o and t_f^o. Similarly, each time an actuator a is activated, it induces a set of trajectories $\boldsymbol{\Omega}_{t_i^a:t_f^a}^a = \left\{T_{t_i^a:t_f^a}^h\right\}_{h=1}^{H^a}$ composed of H^a components between times t_i^a and t_f^a.

15.4 Feature-Aided SMC-PHD for Harmonic Component Tracking

The SMC-PHD filter is considered in this chapter due to its inherent clutter-resilience and computational efficiency (from a data association point of view) [17]. Indeed, in this MTT framework, target dynamics is nonlinear and much clutter (false alarms) is expected, yet the detection probability is high. The SMC-PHD filter is particularly well suited to such problems. This filter consists of five steps: (i) particle sampling, (ii) prediction, (iii) update, (iv) resampling, (v) clustering, and (vi) assignment. Particle sampling is responsible for exploring the state space and spotting new targets. Prediction, update, resampling, and clustering constitute the multitarget state estimation activities. Assignment binds tracks (states and associated covariances) to one another across time steps, resulting in trajectories. The SMC-PHD filter relies on two assumptions [24]:

Assumption 1 *The targets are independent of one another and generate at most one measurement per scan.*

Assumption 2 *Clutter and target birth distributions are Poisson and target-independent.*

We apply uniform particle sampling over the field of view (FOV), that is, the whole spectrum. This degrades the state estimation performance but allows locating any target appearing within the FOV.

State estimation relies on a transition model, here the amplitude and the frequency follow a generic random walk, whereas the complex coefficients rotate at the harmonic component's angular frequency. The transition equation is given by Stephan et al. [25]

$$\begin{bmatrix} a_{t+1}^h \\ b_{t+1}^h \\ A_{t+1}^h \\ \omega_{t+1}^h \end{bmatrix} = \begin{bmatrix} \cos(\omega_t^h \Delta t) & -\sin(\omega_t^h \Delta t) & 0 & 0 \\ \sin(\omega_t^h \Delta t) & \cos(\omega_t^h \Delta t) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_t^h \\ b_t^h \\ A_t^h \\ \omega_t^h \end{bmatrix} + v_t \tag{15.1}$$

where $\boldsymbol{v}_t$ is a zero-mean Gaussian noise and Δt is the duration of a time step.

Similarly to [24], we augment the likelihood (Eq. (15.3a)) as well as the PHD of the posterior (Eq. (15.3b)) using a feature likelihood, that is, the probability for a measurement to be target-originated, based on the spectral kurtosis. The *spectral kurtosis* κ_t^m is obtained by evaluating the kurtosis on a window t filtered using a band-pass filter (BPF) with a $\Delta \omega_{bp}$ bandwidth centered around ω_t^m; a more detailed definition is presented in [26, 27]. In order to interpret the spectral kurtosis and elicit a feature likelihood, the following assumption is made [27]:

Assumption 3 *Noise and excitations highlighting normal modes are assumed to be mesokurtic or leptokurtic.*

That is, the deterministic part of the signals of interest (in their bandwidth) must have a kurtosis strictly below 3 (kurtosis of a Gaussian distribution, i.e., noise in this context). This assumption also leads to a restriction on transient responses. Fast and spiky transients will be leptokurtic and difficult to distinguish from noise and exceptional events. Such transients will thus be treated as clutter. For this reason, we use the spectral kurtosis to correct the estimation in a probabilistic framework rather than triggering measurements solely based on this information.

With a view to lower the weight of clutter-influenced particles, a feature likelihood (assessing the extent to which a measurement was target- or clutter-originated) is elicited from the spectral kurtosis. To this end, we fit a gamma distribution such that the cumulative probability function (cdf) reaches 95% at $\kappa = 3$ (shape parameter $\alpha = 2.615$ and scale parameter $\theta = 0.525$), resulting in the spectral kurtosis likelihood:

$$p_f(\kappa_t^m) = \frac{1}{\Gamma(\alpha)\theta^\alpha}\big(\kappa_t^m\big)^{\alpha-1} \exp\left(-\frac{\kappa_t^m}{\theta}\right) \tag{15.2}$$

This feature likelihood is assumed to be independent of the kinematic likelihood $g_t(z_t^m|\boldsymbol{x}_t^h)$. It refines the weights $\{w_{t|t}^{(p)}\}_{p=1}^{L_t}$ of the L_t particles approximating the PHD of the posterior at time step t, expressed with respect to prior weights $\{w_{t|t-1}^{(p)}\}_{p=1}^{L_t}$:

$$g_t(z_t^m|\boldsymbol{x}_t^h) \leftarrow g_t(z_t^m|\boldsymbol{x}_t^h)p_f(\kappa_t^m) \tag{15.3a}$$

$$w_{t|t}^{(p)} = \left[1 - p_D + \sum_{m \in [\![1,M_t]\!]} \frac{p_D \; g_k(z_t^m|\boldsymbol{x}_t^{(p)})}{K + \sum_{p'=1}^{L_t} p_D \; g_t(z_t^m|\boldsymbol{x}_t^{(p')})w_{t|t-1}^{(p')}}\right] w_{t|t-1}^{(p)} \tag{15.3b}$$

where the detection probability p_D and the clutter spatial density K are assumed to be constant and uniformly distributed over the FOV. These weights are also further scaled up or down (and carefully re-normalized to their original mass) according to their probability $p_f(\kappa_t^{(p)})$ to represent a target, that is, $w_{t|t}^{(p)} \leftarrow w_{t|t}^{(p)} p_f(\kappa_t^{(p)})$. However, evaluating the spectral kurtosis of each particle at its estimated frequency would be computationally intractable. Instead, the spectral kurtosis of a particle is computed as a linear interpolation of the one calculated during preprocessing (for each frequency bin of the DFT). Alternatively, this step can be skipped for real-time applications.

This feature-aided SMC-PHD filter finally yields tracks (states and associated covariances) $\{\boldsymbol{T}^h\}_{h=1}^H$, where H is the total number of harmonic components detected in a dataset.

15.5 Multitarget Tracking to Unsupervised Multilabel Classification

The tracks obtained in the previous section at most represent the behavior of physical components. In this section, a technique is proposed to process and interpret these tracking results in order to recover the activation sequences of the actuators, and the operations they perform.

A first step consists of grouping harmonic components according to their trajectories' states. For simplicity, a descriptive vector is computed for each component, namely $\boldsymbol{\mu}^h = \left[A^h, \omega^h, \sigma_A^h, \sigma_\omega^h\right]^T$, corresponding to the average amplitude, angular frequency, and associated covariances over a trajectory $\boldsymbol{T}^h$. Pairwise distances are computed, allowing for harmonic components to be grouped with one another according to these descriptors. Euclidean distance upon standardized features was considered in this chapter. In more complex cases than those considered, other metrics can be used to associate the tracks between them. For instance, Fréchet distance [28] takes the shape of the trajectory into account, and distances based

on Gaussian processes (GP) [29] can take advantage of uncertainty information (i.e., state covariance along each trajectory) provided by the MTT framework. Furthermore, actual harmonics are grouped together. That is, given two components h and h', h is paired with h' as one of its harmonics if there exist $k \in \mathbb{N}$ such that $\omega_t^h \approx k\,\omega_t^{h'}$ throughout the tracks' lifespans. This step results in sets of trajectories $\boldsymbol{\Omega}^c$ for each component c. The activation statuses $\boldsymbol{\delta}^c \in \{0, 1\}^{1 \times T}$ over T time steps are immediately deducted from these groups.

In a second step, components are grouped together according to their activation sequences. Pairwise similarities are computed to identify components that are always simultaneously active. The *Jaccard index* [30] is considered here, yet other clustering metrics can be used to compare label sequences with each other. This step results in the activation statuses $\boldsymbol{\delta}^a$ of the actuators. At last, each operation can be characterized by a set of actuators simultaneously active. Operation activation statuses $\boldsymbol{\delta}^o$ hence immediately stem from the actuators' statuses.

Interestingly, the transition from multitarget tracking to unsupervised multilabel classification follows a bottom-up approach (gradually building the operation activation statuses from atomic components), whereas traditional underdetermined blind source separation methods use top-down approaches (from operation clustering to their decomposition into atoms) [4, 5].

15.6 Numerical Simulations and Discussion

A study has been conducted to assess the ability of the proposed approach to identify a machine's production process. A representative synthetic scenario has been designed. A univariate signal was composed as the superimposition of actuator-originated signals, according to the pattern "$AC - AB - BC - ABC$", where A, B, C denote both actuators and atomic operations (i.e., originated by a single actuator).

These actuators produce a 50 Hz-triangle wave, a 700–800 Hz second order (with rise time $\tau_1^a = 0.3\,\text{s}$ and damping $\zeta = 0.3$), and a 400–500 Hz first order (with rise time $\tau_2^a = 4s$) modulated sine waves as source signals respectively, with amplitudes 1, 1 and 3 units; noise standard deviations $\sigma_{s,0}^a = 0.2$, $\sigma_{s,1}^a = 0.8$, $\sigma_{s,2}^a = 1.2$. The signal is sampled at $f_s = 6250\,\text{Hz}$, and windowed at $w = 0.3\,\text{s}$ with a 50% overlap for short-time Fourier transform (STFT) and spectral kurtosis computation. Measurements are generated according to a detection threshold $\tau = -5.5$ on the log power spectral density.

To compute the spectral kurtosis, a second-order Butterworth band-pass filter is used with a bandwidth $\Delta\omega_{bp} = 3\Delta f$, depending on the spectral resolution $\Delta f = 1/w$.

This scenario has several specificities. An actuator is never active alone to begin with. This case would typically be misinterpreted by traditional decomposition algorithms, in that the pairs AB, BC, and AC would form atomic unseparated modes. Furthermore, the second-order source signal is characterized by its fast rise time and important overshoot. This signal highlights the expected difficulties encountered by the filter when presented with such transients, as mentioned in the third section.

The SMC-PHD is parametered by a clutter rate $\hat{\lambda}_{FA} = 20$ false alarms per scan, a probability of detection $p_D = 99\%$, noise standard deviations $\sigma_A = 0.3$, $\sigma_\omega = 2 \cdot 2\pi\ rad \cdot s^{-1}$, and 1500 particles per expected target. As a birth model, particles are sampled uniformly over the FOV in order to spot targets as they appear. Due to the high-frequency resolution and the fast convergence of target states, we apply the roughening strategy proposed in [31] in order to limit the risk of sample impoverishment.

Despite obvious difficulties with fast-rising transients, we observe that smooth transients are correctly tracked. In comparison to other UBSS frameworks in which data is represented as successive vectors, the orthogonality between the dimensions would make the associated techniques fail (e.g., singular value decomposition [SVD] or sparse regression).

The results of the proposed approach are presented in Fig. 15.1.

Numerical experiments highlighted the little sensitivity the SMC-PHD filter has with respect to its estimated clutter rate. Furthermore, slightly overestimating $\hat{\lambda}_{FA}$ results in better estimation performance, yet at this stage, artifacts remained. This motivated the use of the spectral kurtosis feature to make the filter less dependent on the true (noise-induced) clutter rate across the different operations. Although this enhancement had a very positive effect on clutter resilience, the formulation proposed to elicit a feature likelihood out of the spectral kurtosis experimentally suffered the (theoretically expected) drawback of preventing the tracker to pick up on fast transient responses. Additionally, the birth model is corrected immediately after particle sampling using the proposed feature likelihood. This prevents erroneous tracks from being generated.

Another major advantage of the proposed method for UBSS problems is the ability to decompose a signal despite nonlinearly mixed signals. In practice, actuators emit component-specific signatures, that is, harmonic components that
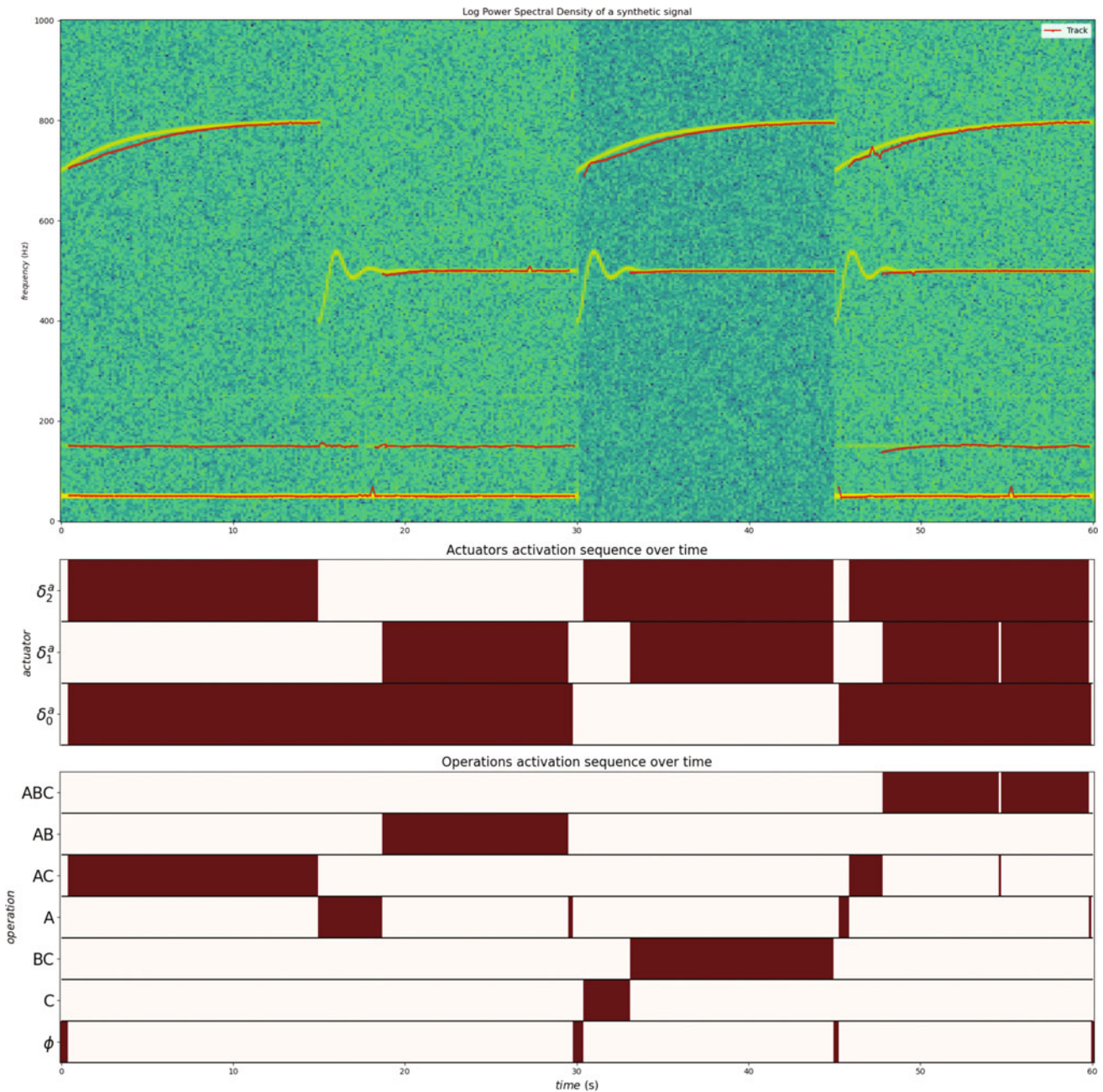
Fig. 15.1 Tracking and decomposition results on a synthetic use case ("$AB - BC - AC - ABC$" sequence of operations with actuators $A = \delta_0^a$, $B = \delta_1^a$, $C = \delta_2^a$; active operations and actuators are shown in red)

uniquely define them in a machine. Hence, by removing coupled harmonics (same frequencies, but different amplitudes), actuators can be well separated regardless of the way they were aggregated in the first place by the remote nonintrusive sensor.

15.7 Conclusions

In this chapter, a feature-aided SMC-PHD was proposed to track harmonic components using the spectral kurtosis to distinguish targets from noise-originated clutter. An algorithm was developed to convert MTT results (trajectories, states,

and covariances) into component, actuator, and operation activation sequences. This end-to-end unsupervised process identification approach was verified on synthetic data in a nontrivial scenario in which typical UBSS methods would underperform.

Future work will focus on increasing clutter resilience, handling stronger nonlinearities and transient responses, and validating the generalizability of the approach. Indeed, real-world signals often exhibit a variety of mixed behaviors, from stationary to nonstationary and nonlinear harmonic components. The use of heterogeneous models will be investigated. Moreover, uncertainty information provided by the MTT framework will be leveraged using statistical distances between tracked harmonic components for a better multilabel clustering performance.

References

1. Sadhu, A., Narasimhan, S., Antoni, J.: A review of output-only structural mode identification literature employing blind source separation methods. Mech. Syst. Signal Process. **94**, 415–431 (2017)
2. Kim, S.G., Yoo, C.D.: Underdetermined independent component analysis by data generation. In: International Conference on Independent Component Analysis and Signal Separation, pp. 445–452. Springer, Berlin (2004)
3. Antoni, J., Chauhan, S.: Second order blind source separation techniques (SO-BSS) and their relation to stochastic subspace identification (SSI) algorithm. In: Structural Dynamics, vol. 3, pp. 177–187. Springer, Berlin (2011)
4. Delabeye, R., Ghienne, M., Kosecki, A., Dion, J.-L.: Unsupervised Manufacturing Process Identification Using Non-intrusive Sensors (2022)
5. Wang, Q., Zhang, Y., Yin, S., Wang, Y., Wu, G.: A novel underdetermined blind source separation method based on optics and subspace projection. Symmetry **13**(9), 1677 (2021)
6. Loesch, B., Yang, B.: Source number estimation and clustering for underdetermined blind source separation. In: Proceedings of the IWAENC (2008)
7. Xie, Y., Xie, K., Wu, Z., Xie, S.: Underdetermined blind source separation of speech mixtures based on k-means clustering. In: 2019 Chinese Control Conference (CCC), pp. 42–46. IEEE, New York (2019)
8. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 689–696 (2009)
9. Wodecki, J., Zdunek, R., Wyłomańska, A., Zimroz, R.: Local fault detection of rolling element bearing components by spectrogram clustering with semi-binary NMF. Diagnostyka **18** (2017)
10. Matsumoto, M., Fujimoto, Y., Hayashi, Y.: Energy disaggregation based on semi-binary NMF. In: International Conference on Machine Learning and Data Mining in Pattern Recognition, pp. 401–414. Springer, Berlin (2016)
11. Yao, Y., Zhao, X., Wu, Y., Zhang, Y., Rong, J.: Clustering driver behavior using dynamic time warping and hidden markov model. J. Intell. Transp. Syst. **25**(3), 249–262 (2021)
12. He, K., Stankovic, V., Stankovic, L.: Building a graph signal processing model using dynamic time warping for load disaggregation. Sensors **20**(22), 6628 (2020)
13. Nie, X.: Detection of grid voltage fundamental and harmonic components using Kalman filter based on dynamic tracking model. IEEE Trans. Ind. Electron. **67**(2), 1191–1200 (2019)
14. Wu, C., Magaña, M.E., Cotilla-Sánchez, E.: Dynamic frequency and amplitude estimation for three-phase unbalanced power systems using the unscented kalman filter. IEEE Trans. Instrum. Meas. **68**(9), 3387–3395 (2018)
15. Climente-Alarcon, V., Antonino-Daviu, J.A., Haavisto, A., Arkkio, A.: Particle filter-based estimation of instantaneous frequency for the diagnosis of electrical asymmetries in induction machines. IEEE Trans. Instrum. Meas. **63**(10), 2454–2463 (2014)
16. Kim, S., Holmstrom, L., McNames, J.: Multiharmonic tracking using marginalized particle filters. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 29–33. IEEE, New York (2008)
17. Vo, B.-N., Singh, S., Doucet, A., et al.: Sequential monte carlo implementation of the PHD filter for multi-target tracking. In: Proceedings of the International Conference on Information Fusion, pp. 792–799 (2003)
18. Clark, D.E., Panta, K., Vo, B.-N.: The GM-PHD filter multiple target tracker. In: 2006 9th International Conference on Information Fusion, pp. 1–8. IEEE, New York (2006)
19. Svensson, L., Svensson, D., Guerriero, M., Willett, P.: Set JPDA filter for multitarget tracking. IEEE Trans. Signal Process. **59**(10), 4677–4691 (2011)
20. He, S., Shin, H.-S., Tsourdos, A.: Multi-sensor multi-target tracking using domain knowledge and clustering. IEEE Sensors J. **18**(19), 8074–8084 (2018)
21. Ruan, Y., Hong, L.: Feature-aided tracking with GMTI and HRR measurements via mixture density estimation. IEE Proceedings-Control Theory and Applications **153**(3), 342–356 (2006)
22. Pace, D.W., Mallick, M., Eldredge, W.: Spectral feature-aided multi-target multi-sensor passive sonar tracking. In: Oceans 2003. Celebrating the Past... Teaming Toward the Future (IEEE Cat. No. 03CH37492), vol. 4, pp. 2120–2126. IEEE, New York (2003)
23. Ying, C., Zhen, C., Shuliang, W.: Feature aided gaussian mixture probability hypothesis density filter with modified 2d assignment. In: Proceedings of 2011 IEEE CIE International Conference on Radar, vol. 1, pp. 800–803. IEEE, New York (2011)
24. Delabeye, R., Shin, H.-S., Inalhan, G.: Feature-aided SMC-PHD filter for nonlinear multi-target tracking in cluttered environments. In: International Conference on Robot Intelligence Technology and Applications, pp. 351–362. Springer, Berlin (2022)
25. Stephan, C., Festjens, H., Renaud, F., Dion, J.-L.: Poles tracking of weakly nonlinear structures using a Bayesian smoothing method. Mech. Syst. Signal Process. **84**, 136–151 (2017)
26. Antoni, J.: The spectral kurtosis: a useful tool for characterising non-stationary signals. Mech. Syst. Signal Process. **20**(2), 282–307 (2006)

27. Dion, J.-L., Tawfiq, I., Chevallier, G.: Harmonic component detection: Optimized spectral kurtosis for operational modal analysis. Mech. Syst. Signal Process. **26**, 24–33 (2012)
28. Devogele, T., Etienne, L., Esnault, M., Lardy, F.: Optimized discrete fréchet distance between trajectories. In: Proceedings of the 6th ACM SIGSPATIAL Workshop on Analytics for Big Geospatial Data, pp. 11–19 (2017)
29. Schweppe, F.C.: State space evaluation of the bhattacharyya distance between two gaussian processes. Inf. Control. **11**(3), 352–372 (1967)
30. da F Costa, L.: Further Generalizations of the Jaccard Index. arXiv preprint arXiv:2110.09619 (2021)
31. Li, T., Corchado, J.M., Sun, S., Fan, H.: Multi-eap: Extended eap for multi-estimate extraction for SMC-PHD filter. Chin. J. Aeronaut. **30**(1), 368–379 (2017)

Chapter 16
Physics-Based Corrosion Reliability Analysis of Miter Gates Using Multi-scale Simulations and Adaptive Surrogate Modeling

Check for updates

Guofeng Qian, Zhen Hu, and Michael D. Todd

Abstract Corrosion-induced crack initiation is primarily simulated in the meso-scale. Such physics-based simulation usually is computationally very expensive. It is computationally even more challenging, or impossible, if the meso-scale simulation model is coupled with macro-scale structural analysis for reliability analysis. This chapter breaks the computational barrier and makes it possible to perform physics-based corrosion reliability analysis of large structures using localized meso-scale simulations, by developing a novel adaptive surrogate modeling framework. A global surrogate model is first constructed at the macro-scale level to enable for the propagation of various input uncertainty sources, such as water levels and gap damage, to uncertainty of the stress response of the structure. After that, a local surrogate model is constructed to predict the local failure probability of any given location by accounting for uncertainty sources presented in both the macro- and meso-scale analysis models. To guarantee the accuracy of the local surrogate model and reduce the required number of meso-scale phase-field (PF) simulations for corrosion reliability analysis, an adaptive surrogate modeling method is proposed based on importance sampling (IS) and active learning to adaptively refine the surrogate model in critical regions. Corrosion reliability analysis of a miter gate structure is employed to demonstrate the efficacy of the proposed method. The result shows that the proposed framework can efficiently and accurately generate a failure probability map for a large structure like miter gate based on computationally expensive meso-scale PF simulations.

Keywords Corrosion-induced cracking · Phase-field simulation · Reliability analysis · Surrogate modeling · Importance sampling

16.1 Introduction

Stress corrosion cracking (SCC) is an important issue in large civil infrastructure. Because of the intrinsic complexity of the multi-physics phase-field (PF) corrosion simulation, it is impossible to perform probabilistic studies which require completing thousands of simulations. Instead, adaptive surrogate modeling combined with importance sampling (IS) makes it possible to perform such probability analysis without computation burden. Figure 16.1 shows the propagation of the uncertainties in multi-scale simulations and final SCC initiation time.

16.2 Surrogate Modeling of Macro-scale Structural Analysis Model

A macro-scale structural analysis model is required to predict the stress response of the miter gate under different load conditions. To this end, a high-fidelity finite element (FE) model of the Greenup miter gate has been developed as shown in Fig. 16.1. The stress response at a certain location $d = [d_x, d_y, d_z]$ predicted from the macro-scalestructural analysis model

G. Qian (✉) · M. D. Todd
Department of Structural Engineering, University of California San Diego, La Jolla, CA, USA
e-mail: g4qian@ucsd.edu

Z. Hu
Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, MI, USA
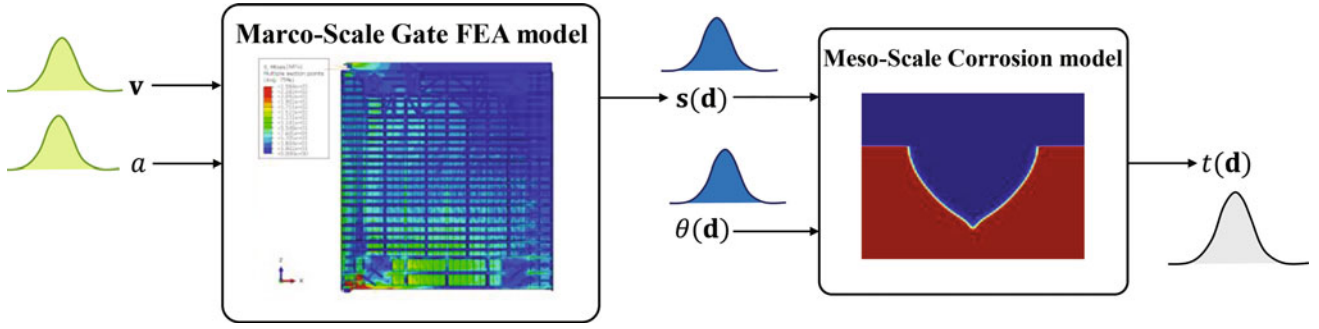
Fig. 16.1 A schematic representation of the proposed method
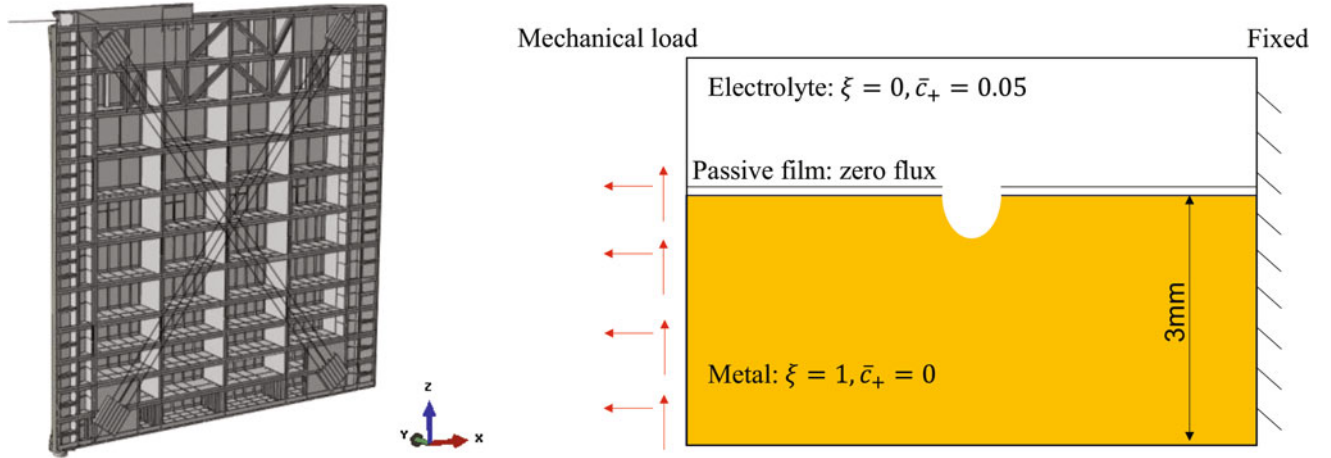
Fig. 16.2 The finite element method (FEM) model for Greenup miter gate (*left*) and the implementation of phase-field corrosion model (*right*)

can be mathematically represented as $[s_x, s_y, s_z] = F(a, \boldsymbol{v}, \boldsymbol{d})$, where $\boldsymbol{v} = [v_u, v_d]$; v_u and v_d are, respectively, the upstream and downstream hydraulic water heads; s_x, s_y, s_z are stress responses in three directions; F stands for the high-fidelity FE model; and a is the gap length which is loss of contact between the gate and the wall quoin block near the bottom of the gate. The gap length is one of the common damage modes that is related to corrosion in long term. After surrogate modeling at macro-scale, the stress response prediction of the whole gate for given $\boldsymbol{x} = [a, \boldsymbol{v}, \boldsymbol{d}]$ is given by $\hat{s}_i(\boldsymbol{x}) = \pi_i(\boldsymbol{x}) \sim \aleph\left(\mu_{\pi i}(\boldsymbol{x}), \sigma_{\pi i}(\boldsymbol{x})\right)$, where $\mu_{\pi i}(\boldsymbol{x})$ are the mean predictions and $\sigma_{\pi i}(\boldsymbol{x})$ is the covariance matrix of the gaussian process regression (GPR) surrogate model for the i-th component of the stress $\pi_i(\boldsymbol{x})$, which can be obtained by following the derivations given in [1].

16.3 Adaptive Surrogate Modeling for Corrosion Reliability Analysis at Meso-scale

For any given location on the gate, the local corrosion growth is simulated as a two-dimensional plate using the phase-field model as illustrated in Fig. 16.2. The mechanical stress load as well as the reaction constant and the diffusion coefficient from the environment are the major factors that affect the corrosion behavior [2]. Based on the SCC initiation criterion and the macro-scale structural analysis model, we define the SCC initiation time T_c at a given location $\boldsymbol{d}$ on the miter gate as $T_c(\boldsymbol{d}) = P_a(\boldsymbol{s}, \theta)$ where P_a represents the phase-field simulation model and θ is a vector of reaction-related parameters in the phase-field model including the reaction constant and the metal diffusion coefficient. The limit state function is defined as $G(\boldsymbol{s}, \boldsymbol{u}) = P_a(\boldsymbol{s}, \boldsymbol{u}) - T_c$ where $\boldsymbol{u}$ is a standard normal random variable transformed from θ and T_c is the time of interest. A GPR surrogate is constructed as $\Im = \hat{G}\left(\boldsymbol{s}, \boldsymbol{u} | \boldsymbol{s}_t, \boldsymbol{u}_t, \boldsymbol{y}_t\right) \sim \aleph\left(\mu_{\hat{G}}(\boldsymbol{s}, \boldsymbol{u}), \sigma_{\hat{G}}(\boldsymbol{s}, \boldsymbol{u})\right)$ where $\boldsymbol{s}_t, \boldsymbol{u}_t, \boldsymbol{y}_t$ are the training points from the simulations.

The failure probability of given stress can be calculated using the surrogate model as

$$p_{\mathrm{f}}(s|\mathfrak{I}) = \int I\left(\mu_{\hat{G}}(s, \boldsymbol{u})\right) f_u(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} \approx \frac{1}{N_{\mathrm{MCS}}} \sum_{i=1}^{N_{\mathrm{MCS}}} I\left(\mu_{\hat{G}}\left(s, \boldsymbol{u}^i\right)\right), \tag{16.1}$$

where I is an indicator function given 0 for a positive value while 1 for a negative value, N_{MCS} is Monte Carlo simulation samples. The Adaptive Kriging Monte Carlo simulation (AK-MCS) method is applied to refine the surrogate model $\mathfrak{I}$ in critical area by defining the learning function $\Lambda(\boldsymbol{u}) = \frac{|\mu_{\hat{G}}(s,\boldsymbol{u})|}{\sigma_{\hat{G}}(s,\boldsymbol{u})}$ and adding the sample with smallest value to the current training samples.

The IS method with the Gaussian mixture model (GMM) is integrated for adaptive surrogate modeling. In IS, the expectation of the indicator function given in Eq. (16.1) is computed based on a different statistical distribution which favors the failure boundary, instead of the original standard normal distribution [3]. An essential element of the IS method is the instrumental density $h(\boldsymbol{u})$. For any given instrumental density function $h(\boldsymbol{u})$, the probability of failure defined in Eq. (16.1) can be rewritten as

$$p_{\mathrm{f}}(s|\mathfrak{I}) = \int I\left(\mu_{\hat{G}}(s, \boldsymbol{u})\right) \frac{f_u(\boldsymbol{u})}{h(\boldsymbol{u})} h(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} \approx \frac{1}{N_{\mathrm{IS}}} \sum_{i=1}^{N_{\mathrm{IS}}} I\left(\mu_{\hat{G}}\left(s, \boldsymbol{u}_{\mathrm{IS}}^i\right)\right) \frac{f_u\left(\boldsymbol{u}_{\mathrm{IS}}^i\right)}{h\left(\boldsymbol{u}_{\mathrm{IS}}^i\right)}, \tag{16.2}$$

where N_{IS} is the importance samples generated from the instrumental density $h(\boldsymbol{u})$. The surrogate model-based optimal instrumental density $h^*(\boldsymbol{u})$ proposed by [4] is employed. Using the surrogate model $\mathfrak{I}$ the approximate optimal instrumental density function $\hat{h}^*(\boldsymbol{u})$ is given by

$$\hat{h}^*(\boldsymbol{u}|\mathfrak{I}) = \frac{\varphi(\boldsymbol{u}|\mathfrak{I}) f_u(\boldsymbol{u})}{\int \varphi(\boldsymbol{u}|\mathfrak{I}) f_u(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}} \propto \varphi(\boldsymbol{u}|\mathfrak{I}) f_u(\boldsymbol{u}), \tag{16.3}$$

where $\varphi(\boldsymbol{u}|\mathfrak{I}) = \Phi\left(\frac{0-\mu_{\hat{G}}(s,\boldsymbol{u})}{\sigma_{\hat{G}}(s,\boldsymbol{u})}\right)$, Φ is the cumulative distribution function (CDF) of a standard normal distribution. Particle filtering is used to solve Eq. (16.3). The weight of each particle is computed first, and the posterior samples are obtained by resampling according to the weight. Since the probability density function (PDF) of the instrumental density function is needed for the calculation of the probability of failure based on IS, we then approximate $\hat{h}^*(\boldsymbol{u})$ as a GMM based on the posterior samples. The GMM is then used to generate the importance samples $\boldsymbol{u}_{\mathrm{IS}}^i$ to calculate Eq. (16.2). The algorithm stops when the maximum error is less than the requirement (i.e., 5% in this paper). If the maximum error requirement is not satisfied, a new training point u_{new} is identified as $u_{\mathrm{new}} = \mathrm{argmin}\Lambda(\boldsymbol{u})$ for $\boldsymbol{u} \in \boldsymbol{u}_{\mathrm{IS}}$. Together with the fixed values of s_i as the input, a new training sample of the inputs are obtained as $[s_i, u_{\mathrm{new}}]$. The corresponding response is obtained as $y_{\mathrm{new}} = G(s_i, u_{\mathrm{new}})$. After that the training dataset is updated with a new sample. With the updated training dataset, the surrogate model $\mathfrak{I}$ is updated. The above process is repeated until the accuracy requirement of the maximum error is satisfied. Once the accuracy requirement is satisfied, we obtain the estimate of $p_{\mathrm{f}}(s|\mathfrak{I})$. Figure 16.3 summarizes the overall procedure of computing the probability of failure for given stress s_i using the proposed adaptive surrogate modeling method. Combined with the macro-FE model $F(\boldsymbol{a}, \boldsymbol{v}, \boldsymbol{d})$ of the miter gate, the failure probability of any location $p_{\mathrm{f}}(\boldsymbol{d})$ can be calculated.

16.4 Results

Figure 16.4 shows the probability of failure map for SCC initiation for the entire miter gate generated using the proposed method. This risk of failure map can be used to provide valuable information for risk-informed maintenance decision-making of the miter gates.
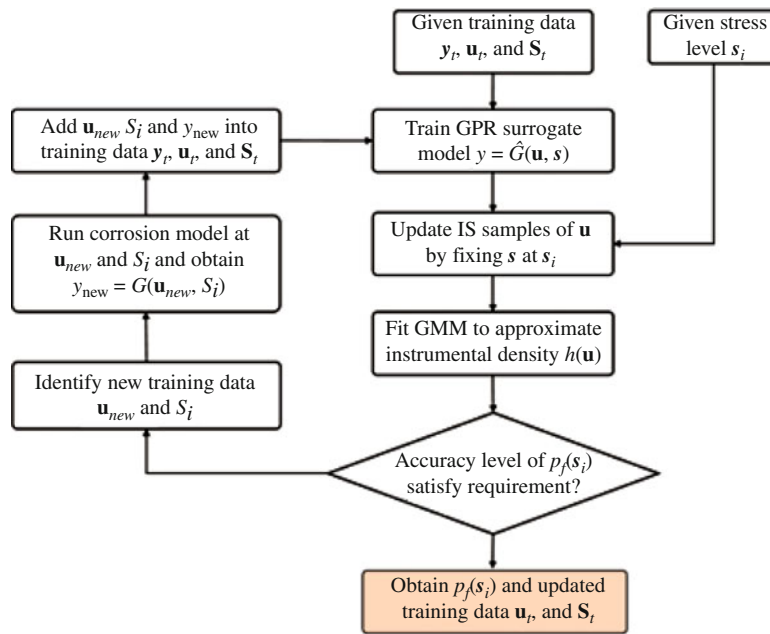
Fig. 16.3 Compute $p_f(s_i)$ of given stress s_i
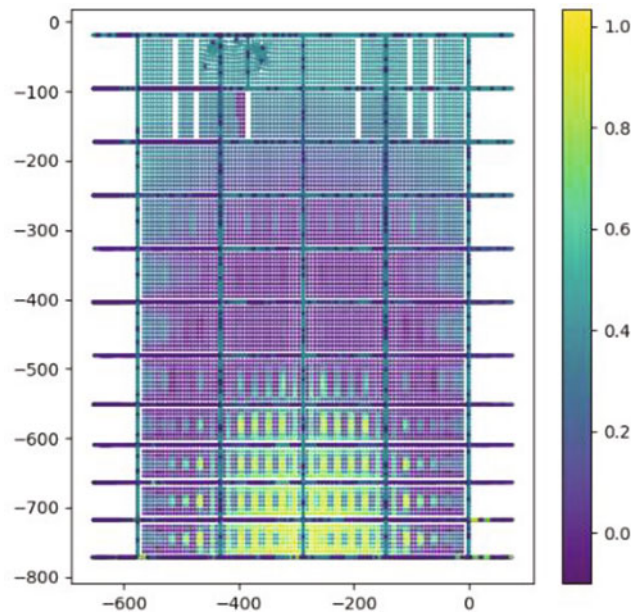
Fig. 16.4 The risk map for SCC initiation on miter gate

References

1. Vega, M.A., Zhen, H., Todd, M.D.: Optimal maintenance decisions for deteriorating quoin blocks in miter gates subject to uncertainty in the condition rating protocol. Reliab. Eng. Syst. Saf. **204**, 107147 (2020)
2. Qian, G., Tantratian, K., Chen, L., Zhen, H., Todd, M.D.: A probabilistic computational framework for the prediction of corrosion-induced cracking in large structures. Sci. Rep. **12**(1), 20898 (2022)
3. Glynn, P.W., Iglehart, D.L.: Importance sampling for stochastic simulations. Manag. Sci. **35**(11), 1367–1392 (1989)
4. Dubourg, V., Sudret, B., Deheeger, F.: Metamodel-based importance sampling for structural reliability analysis. Probab. Eng. Mech. **33**, 47–57 (2013)

Chapter 17
Adaptive Randomized Sketching for Dynamic Nonsmooth Optimization

Robert J. Baraldi, Evelyn Herberg, Drew P. Kouri, and Harbir Antil

Abstract Dynamic optimization problems arise in many applications, including optimal flow control, full waveform inversion, and medical imaging, where they are plagued by significant computational challenges. For example, memory is often a limiting factor on the size of problems one can solve since the evaluation of derivatives requires the entire state trajectory. Additionally, many applications employ nonsmooth regularizers such as the L^1-norm or the total variation as well as auxiliary constraints on the optimization variables. In this chapter, we introduce a novel trust-region algorithm for minimizing the sum of a smooth, nonconvex function and a nonsmooth, convex function that addresses these two challenges. Our algorithm employs randomized sketching to store a compressed version of the state trajectory for use in derivative computations. By allowing the trust-region algorithm to adaptively learn the rank of the state sketch, we arrive at a provably convergent method with near optimal memory requirements. We demonstrate the efficacy of our method on a parabolic PDE-constrained optimization problem with measure-valued control variables.

Keywords Nonsmooth optimization · Optimal control · Randomized sketching · Dynamic optimization · Compression · PDE-constrained optimization

17.1 Introduction

We consider the discrete-time dynamic optimization problem

$$\min_{u_n \in \mathbb{R}^M, \, z_n \in \mathbb{R}^m} \sum_{n=1}^{N} f_n(u_{n-1}, u_n, z_n) + \phi_n(z_n) \qquad \text{subject to} \qquad c_n(u_{n-1}, u_n, z_n) = 0 \quad \text{for} \quad n = 1, \ldots, N, \qquad (17.1)$$

where $z_n \in \mathbb{R}^m$ is the control variable, $u_n \in \mathbb{R}^M$ is the state variable at the nth time step for $n = 1, \ldots, N$, and $u_0 \in \mathbb{R}^M$ is the prescribed initial system state. Additionally, $f_n : \mathbb{R}^M \times \mathbb{R}^M \times \mathbb{R}^m \to \mathbb{R}$ is the objective function associated with the nth control and state, $\phi_n : \mathbb{R}^m \to (-\infty, +\infty]$ is a potentially nonsmooth control penalty function, and $c_n : \mathbb{R}^M \times \mathbb{R}^M \times \mathbb{R}^m \to \mathbb{R}^M$ is the dynamic constraint function, which advances the state from u_{n-1} to u_n. Dynamic optimization problems of the form (17.1) arise in many applications, including turbulent flow control [1], energy system operations [2], vortex control in nuclear reactors and superconductors [3], optimal tomography [4, 5], full waveform inversion [6–8], and airflows in closed environments [9–11]. In addition, nonsmooth penalties are often used to enforce constraints [12–14] or to ensure sparsity in optimal control, parameter estimation and learning [15–23].

R. J. Baraldi (✉) · D. P. Kouri
Optimization and Uncertainty Quantification, Sandia National Laboratories, Albuquerque, NM, USA
e-mail: rjbaral@sandia.gov; dpkouri@sandia.gov

E. Herberg
Interdisciplinary Center for Scientific Computing, Heidelberg University, Heidelberg, Germany
e-mail: eherberg@gmu.edu

H. Antil
Department of Mathematical Sciences and Center for Mathematics and Artificial Intelligence, George Mason University, Fairfax, VA, USA
e-mail: hantil@gmu.edu

The memory required to store the state trajectory $\{u_1, \ldots, u_N\}$ and auxiliary information like Lagrange multipliers presents a significant challenge when solving (17.1). For example, sequential quadratic programming (SQP) methods require the storage of $N(2M+m)$ floating point numbers. In full waveform inversion, the spatial discretization size often is $M \approx 10^{10}$ and the temporal discretization size is $N \approx 10^5$, requiring the storage of $\mathcal{O}(10^{50})$ floating point numbers [24]. In contrast, if $c_n(u_{n-1}, u_n, z_n) = 0$ is uniquely solvable for u_n with fixed u_{n-1} and z_n for each n, then one can reformulate (17.1) as a minimization problem only in $\{z_1, \ldots, z_N\}$. On the surface, this approach reduces the memory requirement to Nm. However, when solving the reduced problem using derivative-based optimization, the gradient calculation requires the entire state trajectory, again leading to $\mathcal{O}(N(M + m))$ storage. Reducing these storage requirements typically comes at the cost of model fidelity by using, for example, reduced-order models (ROMs) or low-order discretizations [25–27]. The quality of a fixed ROM can degrade as the optimization routine progresses, leading to adaptive ROM generation [28, 29]. Unfortunately, ROMs are generally limited to specific classes of dynamical systems and can be difficult to implement in legacy codes. On the other hand, for the reduced problem one can reduce the memory burden using checkpointing [30–33], which stores judiciously chosen snapshots of the state trajectory for use when computing the gradient. Although this procedure has lower memory requirements, it drastically increases the cost of computing the gradient.

In this chapter, we employ adaptive randomized sketching to compress the state trajectory as in [34] to reduce the memory requirement for solving (17.1). In particular, we generate low-rank approximations of the state trajectory that we use to compute an inexact gradient. In contrast to checkpoint, this approach does not increase the computational burden. We control the gradient error using the trust-region algorithm introduced in [35], resulting in a provably convergent, low-memory algorithm for solving (17.1). We demonstrate our algorithm's performance on a discretized parabolic PDE-constrained optimization problem with measure-valued controls.

17.2 Dynamic Optimization Problem Formulation

We consider the reduced form of (17.1) where u_n is replaced by the unique solution to $c(u_{n-1}, u_n, z_n) = 0$ for fixed u_{n-1} and z_n. To formulate the reduced problem, we collect the controls and states into stacked column vectors, denoted by

$$\mathbf{z} = [z_1^\top, \ldots, z_N^\top]^\top \in \mathscr{Z} := \mathbb{R}^{Nm} \qquad \text{and} \qquad \mathbf{u} = [u_1^\top, \ldots, u_N^\top]^\top \in \mathscr{U} := \mathbb{R}^{NM}.$$

We employ the notation $U \in \mathbb{R}^{M \times N}$ to denote the matrix with nth column u_n for $n = 1, \ldots, N$. Using this notation, we can represent the dynamic constraint and objective functions as

$$c(\mathbf{u}, \mathbf{z}) := \begin{bmatrix} c_1(u_0, u_1, z_1)) \\ \vdots \\ c_N(u_{N-1}, u_N, z_N) \end{bmatrix}, \qquad f(\mathbf{u}, \mathbf{z}) := \sum_{n=1}^N f_n(u_n, z_n), \qquad \text{and} \qquad \phi(\mathbf{z}) := \sum_{n=1}^N \phi_n(z_n),$$

enabling us to rewrite (17.1) as

$$\min_{\mathbf{u} \in \mathscr{U}, \, \mathbf{z} \in \mathscr{Z}} f(\mathbf{u}, \mathbf{z}) + \phi(\mathbf{z}) \qquad \text{subject to} \qquad c(\mathbf{u}, \mathbf{z}) = 0. \tag{17.2}$$

We assume that f and c are continuously differentiable on $\mathscr{U} \times \mathscr{Z}$ and that there exists a control-to-state map $\mathbf{z} \mapsto S(\mathbf{z}) : \mathscr{Z} \to \mathscr{U}$, where $S(\mathbf{z})$ is the unique state trajectory satisfying $c(S(\mathbf{z}), \mathbf{z}) = 0$ for each $\mathbf{z} \in \mathscr{Z}$. In addition, we require that the state Jacobian of c, denoted $d_{\mathbf{u}} c(\mathbf{u}, \mathbf{z})$, has a bounded inverse for all controls $\mathbf{z} \in \mathscr{Z}$. Analogously, we denote the control Jacobian by $d_{\mathbf{z}} c(\mathbf{u}, \mathbf{z})$ and the partial derivatives of f by $d_{\mathbf{u}} f(\mathbf{u}, \mathbf{z})$ and $d_{\mathbf{z}} f(\mathbf{u}, \mathbf{z})$. The control-to-state map has the form

$$S(\mathbf{z}) := \begin{bmatrix} S_1(u_0, z_1) \\ S_2(S_1(u_0, z_1), z_2) \\ \vdots \\ S_N(S_{N-1}(\ldots, z_{N-1}), z_N) \end{bmatrix},$$

where the implicit function theorem [36, Th. 1.41] ensures that S_n and S are continuously differentiable. We can thereby reformulate (17.2) as the reduced dynamic optimization problem

$$\min_{\mathbf{z} \in \mathscr{Z}} \{F(\mathbf{z}) := j(\mathbf{z}) + \phi(\mathbf{z})\}, \tag{17.3}$$

where $j(\mathbf{z}) := f(S(\mathbf{z}), \mathbf{z})$ is the reduced objective function. Under the stated assumptions, j is continuously differentiable and its gradient is given by

$$\nabla j(\mathbf{z}) = \mathsf{d}_{\mathbf{z}} f(S(\mathbf{z}), \mathbf{z}) + (\mathsf{d}_{\mathbf{z}} c(S(\mathbf{z}), \mathbf{z}))^{\top} \boldsymbol{\lambda}, \tag{17.4}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{MN}$ solves the adjoint equation

$$\mathsf{d}_{\mathbf{u}} c(S(\mathbf{z}), \mathbf{z})^{\top} \boldsymbol{\lambda} = -\mathsf{d}_{\mathbf{u}} f(S(\mathbf{z}), \mathbf{z}). \tag{17.5}$$

Recall that the adjoint equation (17.5) is solved backward in time, starting at $n = N$ and requires the entire state trajectory.

17.3 Low-Memory Matrix Approximation

For many real-world applications, the state trajectory can be so large as to prohibit storage in working memory. To overcome this challenge, we utilize low-rank matrix sketching to compress the state, which collects *sketched* information about the matrix U from which it can be accurately reconstructed on a fixed storage budget. There are many randomized sketching approaches available (cf. [34] and the references therein) that can be interchanged with the method described below.

We produce a sketch of the state matrix $U \in \mathbb{R}^{M \times N}$ with target rank r, denoted U^r, that requires $\mathcal{O}(r(M + N))$ storage [37]. Let the sketch parameters be $s \geq k \geq r$. A common choice for these parameters is $k = 2r + 1$ and $s = 2k + 1$. The sketch is defined by fixing four random linear dimension reduction maps (DRMs) with i.i.d. standard normal entries:

$$\Upsilon \in \mathbb{R}^{k \times M}, \qquad \Omega \in \mathbb{R}^{k \times N}, \qquad \Phi \in \mathbb{R}^{s \times M}, \qquad \text{and} \qquad \Psi \in \mathbb{R}^{s \times N}.$$

The sketch of U consists of the co-range sketch X, the range sketch Y, and the core sketch Z given by

$$X := \Upsilon U \in \mathbb{R}^{k \times N}, \qquad Y := U \Omega^{\top} \in \mathbb{R}^{M \times k}, \qquad \text{and} \qquad Z := \Phi U \Psi^{\top} \in \mathbb{R}^{s \times s}.$$

The range sketch captures the row space (top-left singular vectors), the co-range sketch captures the column space (top-right singular vectors), and the core sketch captures their interactions (singular values). Linearity of the sketch allows for the online computation of U^r without storing the full state. Since the columns of the state matrix, U, are computed sequentially, we can update sketch components X, Y, and Z in a streaming fashion. For example, the co-range sketch $X = X^{(N)}$ is computed as

$$X^{(0)} = 0 \qquad \text{and} \qquad X^{(n)} = X^{(n-1)} + \Upsilon u_n e_n^{\top} \qquad \text{for} \qquad n = 1, \ldots, N,$$

where e_n is the nth unit vector. Analogous schemes are used to update Y and Z. The sketching matrices require storing $k(M + N) + s^2$ floating point numbers, and hence for target rank r, the memory requirement is $\mathcal{O}(r(M + N) + r^2)$.

To recover the state trajectory from the sketching matrices X, Y, and Z, we first compute QR factorizations of $X^{\top}$ and Y [34]

$$X^{\top} = P R_1 \qquad \text{and} \qquad Y = Q R_2,$$

where $P \in \mathbb{R}^{N \times k}$ and $Q \in \mathbb{R}^{M \times k}$. We then solve two small least-squares problems to form the matrix

$$C = (\Phi Q)^{\dagger} Z ((\Psi P)^{\dagger})^{\top} \in \mathbb{R}^{k \times k}.$$

The rank-k approximation of U is then given by

$$U \approx Q C P^{\top}.$$

This is truncated to rank r by replacing C with its best rank-r approximation. While solving the dynamic optimization problem (17.3), we overwrite X and Y with Q and $W := CP^\top$. For more information, see [34, Sect. 3] and the references therein.

17.4 Sketched Trust-Region Algorithm

We utilize a trust-region method to solve (17.3), while leveraging inexact gradient computations resulting from sketching. As mentioned, our algorithm is an instance of the trust-region method introduced in [35]. Although the method in [35] is provably convergent in Hilbert space, we restrict our developments to $\mathscr{Z} = \mathbb{R}^{mN}$. Following standard convex analysis notation, we denote the subdifferential of a proper, closed and convex function $\psi : \mathscr{Z} \to (-\infty, \infty]$ at an arbitrary vector $z \in \mathscr{Z}$ by

$$\partial \psi(z) := \{\eta \in \mathscr{Z} \mid \psi(y) \geq \psi(z) + \langle \eta, y - z \rangle \ \forall y \in \mathscr{Z}\}$$

and the effective domain of ψ and $\partial \psi$ by $\operatorname{dom} \psi := \{z \in \mathscr{Z} \mid \psi(z) < +\infty\}$ and $\operatorname{dom} \partial \psi := \{z \in \mathscr{Z} \mid \partial \psi(z) \neq \emptyset\}$, respectively. Furthermore, the proximal mapping of ψ for fixed $t > 0$ is

$$\operatorname{Prox}_{t\psi}(y) := \arg\min_{z \in \mathscr{Z}} \left\{ \psi(z) + \tfrac{1}{2t} \|z - y\|^2 \right\}. \tag{17.6}$$

Recall that if $\psi = \iota_{\mathscr{C}}$ is the indicator function of a nonempty, closed and convex set $\mathscr{C} \subset \mathscr{Z}$ (i.e., $\iota_{\mathscr{C}}(z) = 0$ if $z \in \mathscr{C}$ and $+\infty$ otherwise), then $\operatorname{Prox}_{t\psi}$ is the metric projection onto $\mathscr{C}$.

To develop our convergence theory, we make the following assumptions on the components of the objective function F in (17.3).

Assumption 1 (Problem Data)

1. *The function $\phi : \mathscr{Z} \to (-\infty, +\infty]$ is proper, closed, and convex.*
2. *The function $j : \mathscr{Z} \to \mathbb{R}$ is L-smooth on $\operatorname{dom} \phi$. That is, j is Fréchet differentiable and its gradient ∇j is Lipschitz continuous with modulus $L > 0$ on an open set $\bar{\mathscr{Z}} \subseteq \mathscr{Z}$ containing $\operatorname{dom} \phi$.*
3. *The objective function $F := j + \phi$ is bounded below on $\operatorname{dom} \phi$.*

At each iteration of our algorithm, we compute a trial iterate $\mathbf{z}_k^+$ that approximately solves the trust-region subproblem

$$\min_{\mathbf{z} \in \mathscr{Z}} \{m_k(\mathbf{z}) := j_k(\mathbf{z}) + \phi(\mathbf{z})\} \qquad \text{subject to} \qquad \|\mathbf{z} - \mathbf{z}_k\| \leq \Delta_k, \tag{17.7}$$

where $\mathbf{z}_k \in \operatorname{dom} \phi$ is the current iterate, j_k is a smooth local model of j around $\mathbf{z}_k$, and $\Delta_k > 0$ is the trust-region radius. We restrict our attention to quadratic models, j_k, with the form

$$j_k(\mathbf{z}) = \tfrac{1}{2}(\mathbf{z} - \mathbf{z}_k)^\top B_k(\mathbf{z} - \mathbf{z}_k) + g_k^\top(\mathbf{z} - \mathbf{z}_k),$$

where $B_k = B_k^\top \in \mathbb{R}^{mN \times mN}$ approximates the Hessian of j at $\mathbf{z}_k$ and g_k approximates the gradient (e.g., via sketching). For example, we employ the sketched Hessian application described in Algorithms A.5 and A.6 in [34].

To ensure convergence of our trust-region algorithm, we require that the trial iterate $\mathbf{z}_k^+$ satisfies the trust-region constraint and the fraction of Cauchy decrease (FCD) condition:

$$\|\mathbf{z}_k^+ - \mathbf{z}_k\| \leq \kappa_{\mathrm{rad}} \Delta_k \qquad \text{and} \qquad m_k(\mathbf{z}_k) - m_k(\mathbf{z}_k^+) \geq \kappa_{\mathrm{fcd}} h_k \min \left\{ \frac{h_k}{1 + \|B_k\|}, \Delta_k \right\}, \tag{17.8}$$

where $\kappa_{\mathrm{rad}}, \kappa_{\mathrm{fcd}} > 0$ are independent of k and

$$h_k := t^{-1} \left\| \operatorname{Prox}_{t\phi}(\mathbf{z}_k - t g_k) - \mathbf{z}_k \right\|$$

for a fixed positive constant $t > 0$. Commonly, one has $\kappa_{\mathrm{rad}} = 1$. Note that (17.8) ensures that $\mathbf{z}_k^+ \in \mathrm{dom}\,\phi$ since the left-hand side of the second inequality would be $-\infty$ otherwise. Given a trial iterate $\mathbf{z}_k^+$ that satisfies (17.8), the trust-region algorithm decides whether or not to accept $\mathbf{z}_k^+$ based on the ratio of actual and predicted reduction

$$\rho_k := \frac{\mathrm{ared}_k}{\mathrm{pred}_k} = \frac{F(\mathbf{z}_k) - F(\mathbf{z}_k^+)}{m_k(\mathbf{z}_k) - m_k(\mathbf{z}_k^+)}. \tag{17.9}$$

Here, ared_k is the reduction of the objective function F achieved by $\mathbf{z}_k^+$ relative to $\mathbf{z}_k$ and pred_k is the reduction of the model m_k. In particular, if $\rho_k \geq \eta_1$ for $\eta_1 \in (0, 1)$, we accept $\mathbf{z}_{k+1} = \mathbf{z}_k^+$. Otherwise, we set $\mathbf{z}_{k+1} = \mathbf{z}_k$. The trust-region algorithm then increases the radius Δ_k if $\rho_k \geq \eta_2$ for $\eta_2 \in (\eta_1, 1)$ and reduces Δ_k if $\rho_k < \eta_1$. The algorithmic parameters $0 < \eta_1 < \eta_2 < 1$ are user-specified with common values $\eta_1 = 10^{-4}$ and $\eta_2 = 0.75$.

The computation of the gradient of j requires the solution of the backward-in-time adjoint equation (17.5), which depends on the state trajectory $S(\mathbf{z})$. Instead of storing the entire state trajectory, we compress $S(\mathbf{z})$ using sketching and then recover each u_n as needed. This procedure introduces errors in the adjoint and hence gradient. Fortunately, trust-region algorithms are able to rigorously handle inexact gradients, while guaranteeing global convergence [38–41]. The following assumption describes the required gradient accuracy and is adapted from [42]. Moreover, this condition is related to the classical conditions used in [43–45].

Assumption 2 (Inexact Gradient) *There exists a constant $\kappa_{\mathrm{grad}} \geq 0$, independent of k, such that the gradient g_k satisfies*

$$\|g_k - \nabla j(\mathbf{z}_k)\| \leq \kappa_{\mathrm{grad}} \min\{h_k, \Delta_k\} \quad \forall k. \tag{17.10}$$

We provide implementation details for the inexactness conditions (17.10) in Algorithm 2 in the following section. Algorithm 2 is a combination of Algorithm 4 in [35] and the adaptive rank procedure described in Algorithm 4.4 of [34]. We list the nonsmooth trust-region algorithm in Algorithm 1. This algorithm is closely related to the inexact trust-region algorithm described in [46] for smooth unconstrained problems and in [39] for convex-constrained problems.

Algorithm 1 Sketched nonsmooth trust-region algorithm

Require: Initial guess $\mathbf{z}_1 \in \mathrm{dom}\,\phi$, initial rank parameter r_1, initial radius $\Delta_1 > 0$, $0 < \eta_1 < \eta_2 < 1$, and $0 < \gamma_1 \leq \gamma_2 < 1 \leq \gamma_3$
1: **for** $k = 1, 2, \ldots$ **do**
2:     **Model Selection:** Use Algorithm 2 with rank r_k to compute g_k and choose B_k
3:     **Step Computation:** Compute $\mathbf{z}_k^+ \in \mathscr{Z}$ that satisfies (17.8)
4:     **Step Acceptance and Radius Update:** Compute ρ_k as in (17.9)
5:     **if** $\rho_k < \eta_1$ **then**
6:         $\mathbf{z}_{k+1} \leftarrow \mathbf{z}_k$
7:         $\Delta_{k+1} \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k]$
8:     **else**
9:         $\mathbf{z}_{k+1} \leftarrow \mathbf{z}_k^+$
10:         **if** $\rho_k \in [\eta_1, \eta_2)$ **then**
11:             $\Delta_{k+1} \in [\gamma_2 \Delta_k, \Delta_k]$
12:         **else**
13:             $\Delta_{k+1} \in [\Delta_k, \gamma_3 \Delta_k]$
14:         **end if**
15:     **end if**
16: **end for**

17.5   Inexact Gradient Computation via Sketched State

In order to describe the adaptive gradient approximation procedure, we first define the adjoint equation residual $G : \mathscr{U} \times \mathscr{U} \times \mathscr{Z} \to \mathscr{U}$ by $G(\lambda, \mathbf{u}, \mathbf{z}) := \mathsf{d}_{\mathbf{u}} f(\mathbf{u}, \mathbf{z}) + (\mathsf{d}_{\mathbf{u}} c(\mathbf{u}, \mathbf{z}))^* \lambda$ and denote by $\Lambda(\mathbf{u}, \mathbf{z}) \in \mathscr{U}$ the solution to the adjoint equation $G(\Lambda(\mathbf{u}, \mathbf{z}), \mathbf{u}, \mathbf{z}) = 0$ for the fixed state $\mathbf{u}$ and control $\mathbf{z}$. We further define the map

$$g(\lambda, \mathbf{u}, \mathbf{z}) := \mathsf{d}_{\mathbf{z}} f(\mathbf{u}, \mathbf{z}) + (\mathsf{d}_{\mathbf{z}} c(\mathbf{u}, \mathbf{z}))^* \lambda.$$

When evaluated at $\mathbf{u} = S(\mathbf{z})$ and $\boldsymbol{\lambda} = \Lambda(\mathbf{u}, \mathbf{z})$, $g(\boldsymbol{\lambda}, \mathbf{u}, \mathbf{z})$ is the gradient of the reduced objective function j as in (17.4). By evaluating $g(\boldsymbol{\lambda}, \mathbf{u}, \mathbf{z})$ at the sketch state $\mathbf{u}^r = \text{vec}(U^r)$[1] instead of the full state trajectory $\mathbf{u} = S(\mathbf{z})$, we reduce the memory burden for gradient computation. However, the computed value $g^r(\mathbf{z}) = g(\Lambda(\mathbf{u}^r, \mathbf{z}), \mathbf{u}^r, \mathbf{z})$ is only an approximation of true gradient $g(\Lambda(S(\mathbf{z}), \mathbf{z}), S(\mathbf{z}), \mathbf{z})$. Algorithm 2 describes an adaptive procedure for approximating the gradient using the sketched state $\mathbf{u}^r$.

Algorithm 2 Inexact gradient computation with adaptive rank

Require: Control iterate $\mathbf{z}_k \in \mathbb{R}^{mN}$, initial rank parameter r, sketch object for state $\mathbf{u}_k^r$, trust-region radius $\Delta_k > 0$, positive constant $\kappa_{\text{scale}} > 0$, and tolerance $\mu_{\text{grad}} > 1$.

1: Set $\tau_k^- \leftarrow \kappa_{\text{scale}} \Delta_k$
2: Compute $g_k \leftarrow g(\Lambda(\mathbf{u}_k^r, \mathbf{z}_k), \mathbf{u}_k^r, \mathbf{z}_k)$ and $h_k \leftarrow t^{-1} \left\| \text{Prox}_{r\phi}(\mathbf{z}_k - tg_k) - \mathbf{z}_k \right\|$
3: Set $\tau_k^+ \leftarrow \kappa_{\text{scale}} \min\{h_k, \Delta_k\}$
4: **while** $\tau_k^- > \mu_{\text{grad}} \tau_k^+$ **do**
5:     **while** $r < \min\{M, N\}$ **do**
6:         Compute norm of the constraint residual $\texttt{rnorm} \leftarrow \left\| c(\mathbf{u}_k^r, \mathbf{z}_k) \right\|$
7:         **if** $\texttt{rnorm} < \tau_k^+$ **then**
8:             Compute gradient $g_k^r \leftarrow g(\Lambda(\mathbf{u}_k^r, \mathbf{z}_k), \mathbf{u}_k^r, \mathbf{z}_k)$
9:             **break**
10:         **end if**
11:         Increase Rank parameter $r \leftarrow 2r$
12:         Solve the state equation at $\mathbf{z}_k$ and resketch to produce $\mathbf{u}_k^r$
13:     **end while**
14:     Set $g_k \leftarrow g_k^r$ and compute $h_k \leftarrow t^{-1} \left\| \text{Prox}_{r\phi}(\mathbf{z}_k - tg_k) - \mathbf{z}_k \right\|$
15:     Set $\tau_k^- \leftarrow \tau_k^+$ and $\tau_k^+ \leftarrow \kappa_{\text{scale}} \min\{h_k, \Delta_k\}$
16: **end while**
17: **return** Approximate gradient $g_k \approx \nabla f(\mathbf{z}_k)$ using $\mathcal{O}(r(M + N) + mN)$ storage for $r \leq \min\{M, N\}$.

To ensure that Algorithm 2 satisfies the required accuracy (17.10) with finitely many rank updates, we make the following regularity assumptions on the problem data in (17.1).

Assumption 3 (Regularity Properties for (17.1)) *The following conditions hold for the data in (17.1):*

1. *The set of states corresponding to controls in any open and bounded set $\mathcal{Z}_0 \subseteq \mathcal{Z}$ is bounded: there exists $\mathcal{U}_0 \subset \mathcal{U}$ open and bounded such that $\{\mathbf{u} \in \mathcal{U} \mid \exists \mathbf{z} \in \mathcal{Z}_0, c(\mathbf{u}, \mathbf{z}) = 0\} \subseteq \mathcal{U}_0$.*
2. *There exist singular value thresholds $0 < \sigma_0 \leq \sigma_1 < +\infty$ such that for any $\mathbf{u} \in \mathcal{U}_0$ and $\mathbf{z} \in \mathcal{Z}_0$, the state Jacobian matrix $d_{\mathbf{u}} c(\mathbf{u}, \mathbf{z})$ satisfies $\sigma_0 \leq \sigma_{\min}(d_{\mathbf{u}} c(\mathbf{u}, \mathbf{z})) \leq \sigma_{\max}(d_{\mathbf{u}} c(\mathbf{u}, \mathbf{z})) \leq \sigma_1$.*
3. *The following functions are Lipschitz continuous on $\mathcal{U}_0 \times \mathcal{Z}_0$ with respect to their first arguments, and their respective Lipschitz moduli are independent of $\mathbf{z} \in \mathcal{Z}_0$:*

   (a) The state Jacobian of the constraint $d_{\mathbf{u}} c(\mathbf{u}, \mathbf{z})$
   (b) The control Jacobian of the constraint $d_{\mathbf{u}} c(\mathbf{u}, \mathbf{z})$
   (c) The state gradient of the smooth objective term $d_{\mathbf{u}} f(\mathbf{u}, \mathbf{z})$
   (d) The control gradient of the smooth objective term $d_{\mathbf{z}} f(\mathbf{u}, \mathbf{z})$

Using Assumption 3, we can bound the state, adjoint, and gradient errors as in [34, Prop. 4.1] and ultimately show that Algorithm 2 produces a gradient approximation that satisfies (17.10).

Proposition 1 (Proposition 4.1 in [34]) *Suppose Assumption 3 holds for a bounded control set $\mathcal{Z}_0$. Then there exists $\kappa_0, \kappa_1 > 0$ such that the error in the state satisfies*

$$\kappa_0 \|\mathbf{u} - S(\mathbf{z})\| \leq \|c(\mathbf{u}, \mathbf{z})\| \leq \kappa_1 \|\mathbf{u} - S(\mathbf{z})\|, \ \forall \mathbf{u} \in \mathcal{U}_0, \ \mathbf{z} \in \mathcal{Z}_0$$

where $\mathcal{U}_0 \subseteq \mathcal{U}$ is defined in condition 1 of Assumption 3. Additionally, the error in the adjoint is controlled by the adjoint residual together with the state residual: for some $\kappa_2, \kappa_3 > 0$

$$\|\boldsymbol{\lambda} - \Lambda(S(\mathbf{z}), \mathbf{z})\| \leq \kappa_2 \|c(\mathbf{u}, \mathbf{z})\| + \kappa_3 \|G(\boldsymbol{\lambda}, \mathbf{u}, \mathbf{z})\|, \ \forall \mathbf{u}, \boldsymbol{\lambda} \in \mathcal{U}_0, \ \forall \mathbf{z} \in \mathcal{Z}_0.$$

---

[1] The notation $\text{vec}(U)$ denotes the vector obtained by stacking the columns of U.

Therefore, the error in the gradient approximation $g(\lambda, \mathbf{u}, \mathbf{z})$ is controlled by the adjoint and state residuals: for some $\kappa_4, \kappa_5 > 0$

$$\|g(\lambda, \mathbf{u}, \mathbf{z}) - g(\Lambda(S(\mathbf{z}), \mathbf{z}), S(\mathbf{z}), \mathbf{z})\| = \|g(\lambda, \mathbf{u}, \mathbf{z}) - \nabla f(\mathbf{z})\| \leq \kappa_4 \|c(\mathbf{u}, \mathbf{z})\| + \kappa_5 \|G(\lambda, \mathbf{u}, \mathbf{z})\|.$$

Recall that both the state and adjoint are intermediate variables used to compute the gradient $\nabla j(\mathbf{z})$ and require MN storage each. The control $\mathbf{z}$ only requires mN storage, which is often much smaller in practical applications where $m \ll M$. All constants $\kappa_i > 0$ for $i = 0, \ldots, 5$ in Proposition 1 depend only on the finite quantities defined in Assumption 3. We can now prove that Algorithm 2 produces an approximate gradient that satisfies (17.10) in finitely many iterations.

Lemma 1 (Adaptive Rank Gradient Approximation) *If Assumption 3 holds, then Algorithm 2 produces a gradient approximation $g_k = g(\Lambda(\mathbf{u}_k^r, \mathbf{z}_k), \mathbf{u}_k^r, \mathbf{z}_k)$, in finitely many iterations, that satisfies the gradient error bound Assumption 2 with $\kappa_{\mathrm{grad}} = \kappa_4 \kappa_{\mathrm{scale}} \mu_{\mathrm{grad}}$.*

One can prove Lemma 1 using [34, Th. 4.4] and the discussion in Appendix B in [35]. A consequence of Lemma 1 is that Algorithm 1 is guaranteed to converge as demonstrated in the following result.

Theorem 1 (Convergence of Algorithm 1) *Let $\{\mathbf{z}_k\}$ be the sequence of iterates generated by Algorithm 1 and assume that Assumptions 1 and 3 hold. In addition, suppose that there exists an open bounded set $\mathscr{Z}_0 \subset \mathscr{Z}$ with $\{z_k\} \subseteq \mathscr{Z}_0$ and that the model Hessians B_k satisfy*

$$\sum_{k=1}^{\infty} \frac{1}{b_k} = +\infty, \qquad where \qquad b_k := 1 + \max_{i=1,\ldots,k} \|B_i\|.$$

Then

$$\liminf_{k \to +\infty} h_k = 0.$$

Proof The problem data satisfies Assumption 3 and therefore Lemma 1 ensures that Assumption 2 holds. The result then follows from [35, Th. 3]. □

17.6 Numerical Results

In this section, we apply Algorithm 1 to a discretization of the parabolic PDE-constrained optimization problem

$$\min_{z, u} \frac{1}{2} \|u - u_d\|_{L^2(Q)}^2 + \iota_C(z)$$

$$\text{subject to} \quad \begin{cases} \partial_t u - \Delta u = 0 & \text{in } Q := \Omega \times (0, T) \\ \nabla u \cdot n = 0 & \text{on } \Sigma := \partial\Omega \times (0, T) . \\ u(0) = z & \text{in } \Omega \end{cases} \tag{17.11}$$

Here, $\Omega = (0, 1)^2$ and $u_d(x) = |(\sin(2\pi x_1) \sin(2\pi x_2))|^{10}$ for $x = (x_1, x_2) \in \Omega$ and all $t \in [0, T]$. In this application, the control variable z is a nonnegative, regular Borel measure representing the initial state and we enforce the constraints

$$C := \{z \in M(\Omega) \mid \|z\|_{M(\Omega)} \leq \alpha, \quad z(B) \geq 0 \quad \forall \text{ Borel subsets } B \subseteq \Omega\},$$

where $\alpha = 0.1$ and $M(\Omega)$ denotes the Banach space of regular Borel measures on Ω endowed with the total variation norm. We discretize the state variable u in space using continuous piecewise linear finite elements on a uniform triangular mesh ($M = 4225$) and employ a variational discretization for the controls [47]. We further discretize in time using implicit Euler with $N = 501$ timesteps for $T = 2$ to arrive at a problem with the form (17.1). After discretization, the control is represented as a linear combination of point masses located at the mesh vertices and the nonsmooth term ϕ is the indicator function of the feasible set

Table 17.1 Algorithmic performance summary using fixed rank, adaptive rank, and full storage. The table displays the final function value (`objective`), the number of iterations (`niter`), the number of smooth objective evaluations (`nobj`), the number of gradient evaluations (`ngrad`), the number of hessian evaluations (`nhess`), the number of nonsmooth objective evaluations (`nobjn`), the number of proximal operator evaluations (`nprox`), and the compression factor (ζ)

Rank	Objective	niter	nobjs	ngrad	nhess	nobjn	nprox	ζ
*1	2.680962e-02	16	17	9	521	1036	1726	148.78
2	2.680946e-02	37	38	38	1948	4597	3873	89.12
3	2.680946e-02	31	32	32	1635	3759	3248	63.55
4	2.680946e-02	22	23	23	1163	2654	2308	49.35
5	2.680946e-02	22	23	23	1160	2640	2305	40.31
Adaptive	2.680946e-02	22	23	25	1162	2530	2309	25.95
Full	2.680946e-02	23	24	24	1212	2793	2409	---

$$\mathscr{C} = \left\{ \mathbf{z} = (z_1, \ldots, z_m)^\top \in \mathbb{R}^m \;\middle|\; \sum_{i=1}^m z_i \leq \alpha, \quad z_i \geq 0 \quad \text{for} \quad i = 1, \ldots, m \right\},$$

where $m = M$ is the number of mesh vertices. Although the control is time independent, Algorithm 1 is still applicable. We quantify the memory savings of Algorithm 1 using the compression ratio

$$\zeta := \frac{\text{full storage}}{\text{reduced storage}} = \frac{4225 \times 501}{k(4225 + 501) + s^2}.$$

We solved the discretized problem using Algorithm 1 with $\kappa_{\text{scale}} = 10^{-4}$. For comparison, we also solved it using Algorithm 1 in [35] with fixed-rank sketching and with no sketching. To compute the trial step in line 3 of Algorithm 1, we use the spectral proximal gradient method described in [35, Alg. 5] with a maximum of 50 iterations. We further set the maximum number of trust-region iterations to 100. We terminate Algorithm 1 if either

$$h_k \leq 10^{-4} h_1 \qquad \text{or} \qquad \left\| \mathbf{z}_k^+ - \mathbf{z}_k \right\| \leq 10^{-6} h_1.$$

Table 17.1 compares the performance of Algorithm 1 with the fixed-rank ($r \in \{1, \ldots, 5\}$) and full-storage approaches. When using rank-1 sketching, the algorithm stopped because the norm of the trial step size was smaller than the prescribed tolerance. Overall, we see a decreasing trend in the number of iterations, resulting from fewer rejected steps as the fixed rank increases. In comparison, Algorithm 1 finished with the final rank of $r = 8$. The performance of the full-storage, adaptive sketching, and fixed-rank with $r = 4, 5$ approaches are comparable, suggesting that Algorithm 1 is a memory-efficient, application-agnostic approach to solving dynamic optimization problems with the form (17.3).

17.7 Conclusion

In this work, we describe a low-memory, application-agnostic approach for solving a class of nonsmooth dynamic optimization problems without the need to store or recompute the entire state trajectory. Our method uses randomized matrix sketching to compress the state trajectory for use when solving the adjoint equation and inexactly evaluating the gradient. We employ a trust-region algorithm to control the gradient approximation by adaptively learning the state sketch rank $r \ll \min\{M, N\}$, where N is the number of time steps and M is the size of each state. In contrast to traditional approaches that require $\mathscr{O}(N(M + m))$ memory (m being the control dimension) or significant recomputation of the state trajectory, our approach greatly reduces the storage to $\mathscr{O}(r(M + N) + mN)$ with no additional computational cost, enabling the solution of large-scale dynamic problems.

References

1. Gunzberger, M.D.: Perspectives in Flow Control and Optimization. SIAM, Philadelphia (2003)
2. Dentcheva, D., Römisch, W.: Optimal power generation under uncertainty via stochastic programming. In: Marti, K., Kall, P. (eds.) Stochastic Programming Methods and Technical Applications, pp. 22–56. Springer, Berlin (1998)
3. Harada, K., Matsuda, T., Bonevich, J., Igarashi, M., Kondo, S., Pozzi, G., Kawabe, U., Tonomura, A.: Real-time observation of vortex lattices in a superconductor by electron microscope. Nature **360**(6399), 51 (1992)
4. Arridge, S.R., Schotland, J.C.: Optical tomography: forward and inverse problems. Inverse Problems **25**, 123010 (2009)
5. Klose, A.D., Hielscher, A.H.: Optical tomography using the time-independent equation of radiative transfer-part 2: inverse model. J. Quantit. Spectr. Radiat. Trans. **72**, 715–732 (2002)
6. Krebs, J.R., Anderson, J.E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., Lacasse, M.-D.: Fast full-wavefield seismic inversion using encoded sources. Geophysics **74**, WCC177–WCC188 (2009)
7. Tarantola, A.: Linearized inversion of seismic reflection data. Geophys. Prospect. **32**, 998–1015 (1984)
8. Warner, M., Gausch, L.: Adaptive waveform inversion: theory. Geophysics **81**, R429–R445 (2016)
9. Löhner, R., Antil, H.: High fidelity modeling of aerosol pathogen propagation in built environments with moving pedestrians. Int. J. Numer. Methods Biomed. Eng. **37**(3), 3428–3434 (2021)
10. Löhner, R., Antil, H., Idelsohn, S., Oñate, E.: Detailed simulation of viral propagation in the built environment. Comput. Mech. **66**(5), 1093–1107 (2020)
11. Löhner, R., Antil, H., Srinivasan, A., Idelsohn, S., Oñate, E.: High-fidelity simulation of pathogen propagation, transmission, and mitigation in the built environment. Arch. Comput. Methods Eng. **28**, 1–26 (2021)
12. Kouri, D.P.: A matrix-free trust-region Newton algorithm for convex-constrained optimization. Optim. Lett. **16**, 1–15 (2021)
13. Schmidt, M., Berg, E., Friedlander, M., Murphy, K.: Optimizing costly functions with simple constraints: A limited-memory projected quasi-Newton algorithm. In: van Dyk, D., Welling, M. (eds.) Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 5, pp. 456–463, Hilton Clearwater Beach Resort, Clearwater Beach, Florida (2009). PMLR
14. Toint, P.L.: Global convergence of a class of trust-region methods for nonconvex minimization in Hilbert space. IMA J. Numer. Analy. **8**(2), 231–252 (1988)
15. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
16. van den Berg, E., Friedlander, M.P.: Probing the Pareto frontier for basis pursuit solutions. SIAM J. Sci. Comput. **31**(2), 890–912 (2008)
17. Combettes, P.L., Pesquet, J.-C.: Proximal splitting methods in signal processing. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer, New York (2011)
18. Herzog, R., Obermeier, J., Wachsmuth, G.: Annular and sectorial sparsity in optimal control of elliptic equations. Comput. Optim. Appl. **62**(1), 157–180 (2015)
19. Herzog, R., Stadler, G., Wachsmuth, G.: Directional sparsity in optimal control of partial differential equations. SIAM J. Control Optim. **50**(2), 943–963 (2012)
20. Porcelli, M., Simoncini, V., Stoll, M.: Preconditioning PDE-constrained optimization with l1-sparsity and control constraints. Comput. Math. Appl. **74**(5), 1059–1075 (2017). SI: SDS2016 – Methods for PDEs
21. Stadler, G.: Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. Comput. Optim. Appl. **44**(2), 159 (2009)
22. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B **58**(1), 267–288 (1996)
23. Yun, J., Zheng, P., Yang, E., Lozano, A., Aravkin, A.: Trimming the L1 regularizer: Statistical analysis, optimization, and applications to deep learning. In: International Conference on Machine Learning, pp. 7242–7251. PMLR (2019)
24. Antil, H., Kouri, D.P., Lacasse, M.-D., Ridzal, D.: Frontiers in PDE-Constrained Optimization, vol. 163. Springer, New York (2018)
25. Jalali, A.A., Sims, C.S., Famouri, P.: Reduced order systems. In: Lecture Notes in Control and Information Sciences, vol. 343. Springer, Berlin (2006)
26. Antoulas, A.C.: Approximation of Large-Scale Dynamical Systems. Society for Industrial and Applied Mathematics, Philadelphia (2005)
27. Dedé, L.: Reduced basis method and a psoteriori error estimation for parameterized linear-quadratic optimal control problems. SIAM J. Sci. Comput. **32**, 997–1019 (2010)
28. Zahr, M.J., Carlberg, K.T., Kouri, D.P.: An efficient, globally convergent method for optimization under uncertainty using adaptive model reduction and sparse grids. SIAM/ASA J. Uncert. Quantif. **7**(3), 877–912 (2019)
29. Fahl, M., Sachs, E.: Reduced order modelling approaches to PDE-constrained optimization based on proper orthogonal decomposition. In: Beigler, L.T., Ghattas, O., Heinkenschloss, M., van Bloemen Waanders, B. (eds.) Large-Scale PDE-Constrained Optimization. Lecture Notes in Computational Science and Engineering, vol. 30. Springer, Berlin (2003)
30. Griewank, A., Walther, A.: Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation. ACM Trans. Math. Softw. **26**, 19–45 (2000)
31. Aupy, G., Herrmann, J., Hovland, P., Robert, Y.: Optimal multistage algorithm for adjoint computation. SIAM J. Sci. Comput. **38**, C232-255 (2016)
32. Stumm, P., Walther, A.: New algorithms for optimal online checkpointing. SIAM J. Sci. Comput. **32**, 836–854 (2010)

33. Wang, Q., Moin, P., Iaccarino, G.: Minimal repetition dynamic checkpointing algorithm for unsteady adjoint calculation. SIAM J. Sci. Comput. **31**, 2549–2567 (2009)
34. Muthukumar, R., Kouri, D.P., Udell, M.: Randomized sketching algorithms for low-memory dynamic optimization. SIAM J. Optim. **31**(2), 1242–1275 (2021)
35. Baraldi, R.J., Kouri, D.P.: A proximal trust-region method for nonsmooth optimization with inexact function and gradient evaluations. Math. Programm. (2022). Submitteds
36. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE Constraints, vol. 23. Springer, New York (2009)
37. Tropp, J.A., Yurtsever, A., Udell, M., Cevher, V.: Streaming low-rank matrix approximation with an application to scientific simulations. SIAM J. Sci. Comput. **41**, A2430–A2463 (2019)
38. Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust Region Methods. SIAM, Philadelphia (2000)
39. Garreis, S., Ulbrich, M.: An inexact trust-region algorithm for constrained problems in Hilbert space and its application to the adaptive solution of optimal control problems with PDEs. Preprint, submitted, Technical University of Munich (2019)
40. Kouri, D.P., Heinkenschloss, M., Ridzal, D., van Bloemen Waanders, B.G.: A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty. SIAM J. Sci. Comput. **35**(4), A1847–A1879 (2013)
41. Kouri, D.P., Ridzal, D.: Inexact trust-region methods for PDE-constrained optimization. In: Frontiers in PDE-Constrained Optimization, pp. 83–121. Springer, New York (2018)
42. Heinkenschloss, M., Vicente, L.N.: Analysis of inexact trust-region SQP algorithms. SIAM J. Optim. **12**, 283–302 (2001)
43. Carter, R.G.: Numerical optimization in Hilbert space using inexact function and gradient evaluations. Technical Report 89-45, ICASE, Langley, VA (1989)
44. Carter, R.G.: On the global convergence of trust region algorithms using inexact gradient information. SIAM J. Numer. Anal. **28**, 251–265 (1991)
45. Carter, R.G.: Numerical experience with a class of algorithms for nonlinear optimization using inexact function and gradient information. SIAM J. Sci. Comput. **14**(2), 368–388 (1993)
46. Kouri, D.P., Heinkenschloss, M., Ridzal, D., van Bloemen Waanders, B.G.: Inexact objective function evaluations in a trust-region algorithm for PDE-constrained optimization under uncertainty. SIAM J. Sci. Comput. **36**(6), A3011–A3029 (2014)
47. Herberg, E., Hinze, M.: Variational discretization approach applied to an optimal control problem with bounded measure controls. Optim. Control Partial Differ. Equ.: Uncert. Quantif. Open Closed-Loop Control Shape Optim. **29**, 113 (2022)

Chapter 18
Predicting Nonlinear Structural Dynamic Response of ODE Systems Using Constrained Gaussian Process Regression

Yishuang Wang, Yang Yu, Xinyue Xu, and Sez Atamturktur

Abstract Identification and characterization of a nonlinear structural dynamic system often involve inferring unknown parameters from experimental data. Compared to its linear counterparts, nonlinear systems include additional parameters related to the restoring force, making identifiability more challenging. In this chapter, we propose to augment linear structural dynamic models with empirically inferred state-dependent parameters to predict the responses of nonlinear structural dynamic models. Specifically, we represent the state-dependent parameter by a constrained Gaussian process regression (cGP). In addition to computational efficiencies, the use of cGP to constrain the model from uncertainty by incorporating prior knowledge is intended to enhance extrapolation performance. To demonstrate the feasibility and effectiveness of the proposed approach, we focus on a simple ordinary differential equations (ODEs) case study and impose the monotonicity constraint on a state-dependent parameter to highlight the impact of prior knowledge on predictive performance.

Keywords Nonlinear structural dynamics · Bayesian inference · Constrained Gaussian process · Uncertainty quantification · State-dependent parameters

18.1 Introduction

Computer models are widely used to simulate complex structural dynamics problems, such as bridge vibrations under different loads [12], rotating machines in mechanical systems [1], or offshore structure displacement facing wave and wind forces [13]. These simulations often entail poorly known model input that needs to be calibrated against experimental measurements and the models themselves can be computationally demanding and time-consuming, especially when modeling highly nonlinear structural dynamic behavior [16]. For example, the complicated displacement of structural elements during an earthquake can be simplified by the Bouc-Wen model; however, many unknown parameters in the model combined with limited prior knowledge of these parameters make it challenging to calibrate the model from experiments. Embedding surrogate models into ordinary differential equations (ODEs) of dynamical systems could offer a computationally feasible way of both calibrating the poorly known model parameters and achieving computational efficiencies. One popular method of surrogate modeling is a Gaussian Process (GP) – a non-parametric model widely used in nonlinear regression and classification [11].

In this chapter, we leverage GPs to statistically infer the poorly known parameters in the ODEs of nonlinear dynamical systems, such as those with a Duffing oscillator and the Bouc-Wen model. We refer to these terms as state-dependent (or functional) parameters as they depend on system inputs or state variables. We learn these state-dependent parameters from experimental data and then use the inferred parameters to predict the system behavior in unobservable settings. A key benefit of using a GP to represent the state-aware parameters is the ability to consider all relevant forms of uncertainty while empirically capturing the hidden relationship between the state-dependent variables and the state inputs into functions. Here, the term "state-dependent" means that this parameter has a functional dependency on one or more of the state inputs of the system. The goal therefore is to empirically capture this parameter by training a Gaussian process by solving the inverse problem.

The remainder of the chapter is outlined as follows: Sect. 18.2 reviews the constrained GP (cGP) regression as well as the past research on applications governed by ODEs. Section 18.3 details the proposed methodology of the cGP by incorporating

Y. Wang (✉) · Y. Yu · X. Xu · S. Atamturktur
Department of Architectural Engineering, The Pennsylvania State University, University Park, PA, USA
e-mail: YMW5459@PSU.EDU

the derivatives of existing data points in the covariance matrix of the original GP. Section 18.4 provides simulation results of a proof-of-concept case study based on an ODE problem with a monotonic input-output response imposed as a constraint. Finally, conclusions and limitations are summarized along with further work in Sect. 18.5.

18.2 Background

Studies on nonlinear structural dynamics involve the identification and characterization of equations of motion representing many physical phenomena, including but not limited to amplitude-dependent natural frequencies, contact and friction dynamics, amplitude-dependent response, self-excited oscillating behavior, and non-repeatable responses. As an example, a Duffing oscillator and the Bouc-Wen model both contain nonlinear terms in the ODE equation that treat the restoring force as amplitude dependent, which improves fidelity to real-life behavior all while requiring the challenging task of inferring the nonlinear model parameters from experiments. This problem of characterizing nonlinear systems to predict the response at new sampling points has received considerable attention in the literature [15]. Many statistical approaches were developed to solve the nonlinear term with the aid of fast running computer models [2–5].

A popular approach to determining the imperfect, biased nature of a numerical model is that which was proposed by Kennedy and O'Hagan (KOH) [2]. Bayesian treatment of the calibration problem permits putting prior distribution on poorly known input parameters to constrain them to elicited, plausible values. A Gaussian process (GP) prior is typically used to model the bias function reflecting a systematic shortcoming of the computer model that leads to the model predictions disagreeing with the observed response, regardless of the value of model input parameters and regardless of measurement precision. The calibration problem is then solved via optimization to find the maximum a posteriori estimator of model parameters while simultaneously and explicitly estimating model bias. To date, almost all KOH-style calibration methods proposed in the literature largely assumed that the calibrated parameter values are fixed throughout the input domain [2]. In certain applications, however, the calibration parameters vary over the input domain [11] so that they are not constant but function over the state space. In some cases, this variation is due to inconsistencies in the underlying processes so that calibration parameters vary randomly over experimental conditions [8, 9]. In many other cases, though, the calibration parameters behave as smoothly varying functions over the input domain. Such state-dependent behavior is also prevalent in nonlinear dynamical problems.

The earliest studies on the application of Gaussian processes in structural dynamics can be traced back to Shinozuka et al. [12, 13] where Monte Carlo sampling was used to characterize a nonlinear system and predict its dynamical response. The coefficients of a single degree of freedom system are modeled as a Gaussian process [13] with a constant mean, and then it is combined with a time-series model to predict the vibration response under different conditions of operations. Avendaño-Valencia et al. [15] implemented this framework to predict the vibration response of a bridge given acceleration signal and used maximum a posteriori estimation to infer the hyperparameters of the GP. Xia et al. [18] developed finite element models of simplified bladed disks from an industrial engine and predicted frequency responses using the GP under uncertainty. Their study demonstrated that GPs can be used as an effective data-driven tool to predict the response of linear equations of motion meanwhile reducing the computational time.

More recently, state-aware (or functional) calibration which treats the unknown parameters as a function of inputs became a topic of interest [10]. In Stevens et al. [4, 5], a missing dependency of a single model input on a single state variable was predicted in a functional form through Bayesian inference. The problem is formulated as a weak coupling of two different models, where the empirical constituent model, namely the feeder model, calculates the behaviors of interest and the physics-based model, namely the consumer model, receives the results from the feeder as its own inputs [6].

18.3 Methodology

For a given physics-based model, $\eta(\boldsymbol{x}, \boldsymbol{\theta})$, let $\boldsymbol{y} = (y(x_1), \ldots, y(x_N)))^{\mathrm{T}}$ denote the output of experimental data, where x represents the operational domain of the given system and $\boldsymbol{\theta}$ represents the calibration parameters of the model. Experiments are conducted at various settings within the domain such that $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$. Here, we assume that the physics-based model is uniformly accurate across the domain of applicability and hence, no discrepancy bias term is considered. Therefore, the experimental data can be represented as:

$$\boldsymbol{y} = \eta(\boldsymbol{x}, \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, 2, \ldots N \tag{18.1}$$

Suppose the error ε_i follows a normal distribution $N_N \left(\mathbf{0}, \lambda_y^{-1} \mathbf{I} \right)$ [10], where $\mathbf{I}$ is the identity matrix and N is the total number of experimental observations. In Bayesian calibration, the posterior distribution of the unknown parameter can be expressed as:

$$P \left(\boldsymbol{\theta}, \varepsilon_i | \mathbf{y}_i \right) \propto P \left(\boldsymbol{\theta}, \varepsilon_i \right) \cdot L \left(\mathbf{y}_i | \boldsymbol{\theta}, \varepsilon_i \right) \tag{18.2}$$

where $L(\mathbf{y}_i | \boldsymbol{\theta}, \varepsilon_i)$ is the likelihood function. $\boldsymbol{\theta}$ is state-aware and modeled as a GP with a mean μ_θ, hyperparameter λ_θ and $\mathbf{R}_{\rho_\theta}$, and the correlation function can then be represented as $\mathbf{R}_{\rho_\theta} \left(x, x' \right) = \rho_\theta^{4(x-x')^2}$. Therefore, the new proposed model becomes:

$$
\begin{aligned}
\mathbf{y} \mid \boldsymbol{\theta}^{(\mathbf{x})}, \lambda_y &\sim N_N \left(\boldsymbol{\eta} \left(\boldsymbol{\theta}^{\mathbf{x}} \right), \lambda_y^{-1} \mathbf{I} \right), \\
\lambda_y &\sim \text{Ga} \left(a_y, b_y \right), \quad a_y, b_y > 0, \\
\theta \left(\cdot \right) \mid \lambda_y, \rho_\theta &\sim \mathbf{GP} \left(\mu_\theta, \lambda_\theta^{-1} R_{\rho_\theta} \left(\cdot, \cdot \right) \right) \quad -\infty < \mu_\theta < \infty, \\
\theta_2 &\sim \text{Unif} \left(0, 1 \right), \\
\lambda_\theta &\sim \text{Ga} \left(a_\theta, b_\theta \right), a_\theta, b_\theta > 0, \\
\rho_\theta &\sim \text{Beta} \left(1, b_\theta \right), b_\theta > 0,
\end{aligned}
\tag{18.3}
$$

where $\boldsymbol{\theta}^{(\mathbf{x})} = (g(\theta(x_1)), \ldots, g(\theta(x_N)))^{\mathrm{T}}$. Therefore, the joint posterior distribution is

$$
\pi \left(\boldsymbol{\theta}^{(\mathbf{x})}, \rho_\theta, \lambda_\theta, \lambda_y | \mathbf{y} \right) \propto \lambda_y^{N/2 + a_y - 1} \exp \left\{ -\tfrac{\lambda_y}{2} \left(\mathbf{y} - \boldsymbol{\eta} \left(\boldsymbol{\theta}^{(\mathbf{x})} \right) \right)^{\mathrm{T}} \left(\mathbf{y} - \boldsymbol{\eta} \left(\boldsymbol{\theta}^{(\mathbf{x})} \right) \right) \right\}
$$
$$
\cdot \exp \left(-b_y \lambda_y \right) \lambda_\theta^{N/2 + a_y - 1} \left| \mathbf{R}_{\rho_\theta} \right|^{-1/2} \cdot \exp \left\{ -\tfrac{\lambda_\theta}{2} \left(\boldsymbol{\theta}^{(\mathbf{x})} - \mu_\theta \mathbf{1} \right)^{\mathrm{T}} \mathbf{R}_{\rho_\theta}^{-1} \left(\boldsymbol{\theta}^{(\mathbf{x})} - \mu_\theta \mathbf{1} \right) \right\} \cdot \exp \left(-b_\theta \lambda_\theta \right) \left(1 - \rho_\theta \right)^{b_\theta - 1}
\tag{18.4}
$$

The posterior for the calibration parameter and hyperparameter can be derived based on the joint posterior. As for the prior distribution, λ_y and λ_θ both follow $\text{Ga}(5, \ 5)$ and ρ_θ follows $\text{Beta}(1, 5)$.

18.4 Case Study

We provide a conceptual case study to demonstrate our proposed methodology. A cGP model is utilized to represent an ODE system combined with a monotonicity constraint on the state-aware parameter. Note that this analytical case study including both the parameter dependency and prior knowledge is identical to that of Stevens et al. [4] except for the form of $\eta(\mathbf{x}, \boldsymbol{\theta})$. A key difference between this earlier study and the present one is that our model is an ODE and generated through the following:

$$\frac{d\mathbf{y}}{d\mathbf{x}} = x + \boldsymbol{\theta} \left(x \right) \tag{18.5}$$

where $\boldsymbol{\theta} \left(x \right) = 2\sqrt{x}$. Eighty evenly spaced synthetic data points are generated from $x \in [0, 4]$ as the experimental data. The middle 16 points are held out for subsequent validation. The resulting system response is depicted in Fig. 18.1. As seen in Fig. 18.1, the system response is monotonically increasing.

We use the Metropolis algorithm with Gibbs sampler in MCMC to infer the posterior distribution of both the model parameters and GP hyperparameters. Three chains with different starting values are run, where 3000 samples are drawn as the burn-ins during the converging period and are followed by 1000 samples for the actual inferred parameters. The results are plotted in Fig. 18.2a with one-standard-deviation.

Suppose the monotonic nature of the system response depicted in Fig. 18.1 is known to the model developers (i.e., subject matter experts). Then, the monotonic behavior can be imposed as a constraint to the GP, which could substantially improve identifiability. To enforce the monotonicity, Riihimäki et al. [7] propose that virtual points θ', which represent the derivatives of samples in GP, are sampled together with the actual sample points θ'. At any sampling point $x^{(i)}$, the monotonicity constraint can be depicted as:
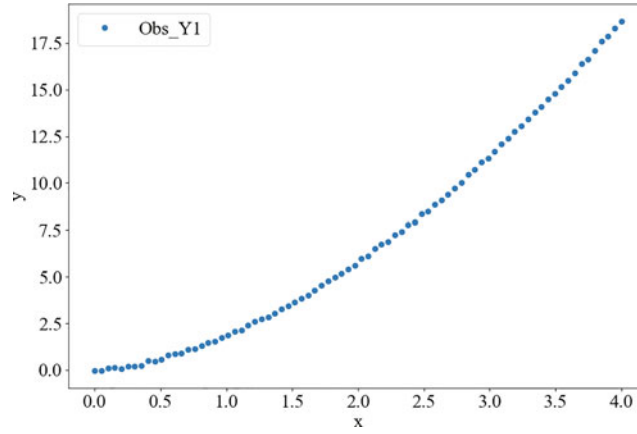
Fig. 18.1 The observation of the proposed nonlinear system
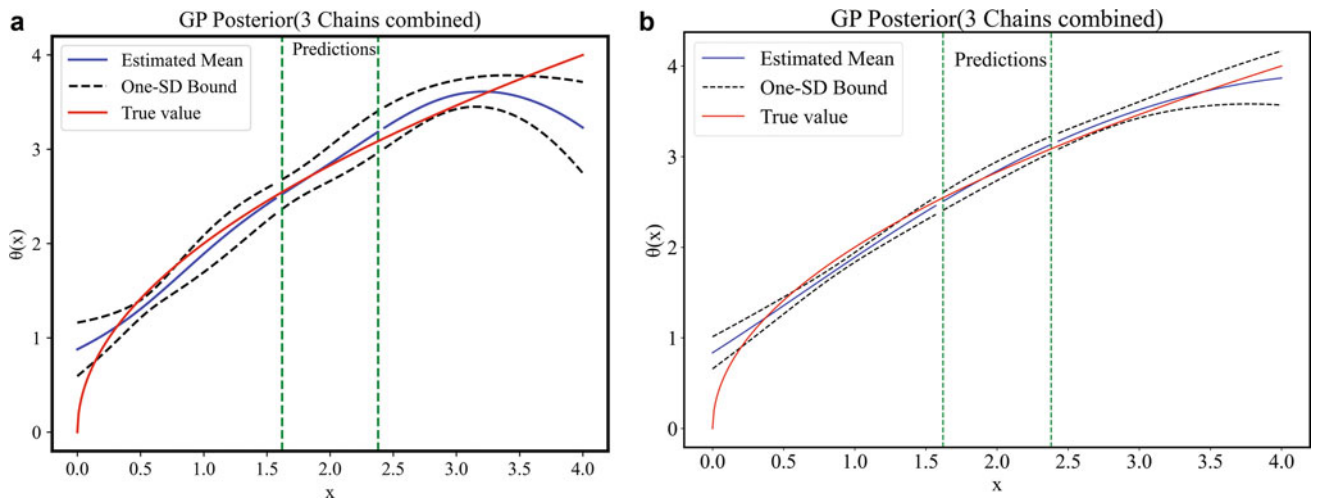
Fig. 18.2 (**a**) Posterior prediction distributions of the ODE model based on GPs without a monotonicity constraint; (**b**) Posterior prediction distributions of the ODE model based on GPs with a monotonicity constraint

$$\theta' = \frac{d\theta^{(i)}}{dx^{(i)}} > 0$$

Since the derivatives follow a GP, the new GP satisfies as follows:

$$p\left(\theta, \theta'|X, X_m\right) = N\left(\theta_{\text{joint}}|0, K_{\text{joint}}\right)$$

where

$$\theta_{\text{joint}} = \begin{bmatrix} \theta \\ \theta' \end{bmatrix}, \text{ and } K_{\text{joint}} = \begin{bmatrix} K_{\theta,\theta} & K_{\theta,\theta'} \\ K_{\theta',\theta} & K_{\theta',\theta'} \end{bmatrix}$$

where

$$K_{\theta',\theta} = \text{Cov}\left[\theta^{(i)} \ \theta^{(j)}\right] = \lambda_\theta^{-1} R_{\rho_\theta}\left(x^{(i)}, x^{(j)}\right)$$
$$K_{\theta',\theta} = \text{Cov}\left[\frac{\partial\theta^{(i)}}{\partial x^{(i)}} \ \theta^{(j)}\right] = \frac{\partial}{\partial x^{(i)}} \text{Cov}\left[\theta^{(i)} \ \theta^{(j)}\right]$$
$$K_{\theta',\theta'} = \text{Cov}\left[\frac{\partial\theta^{(i)}}{\partial x^{(i)}} \ \frac{\partial\theta^{(j)}}{\partial x^{(j)}}\right] = \frac{\partial}{\partial x^{(i)} x^{(j)}} \text{Cov}\left[\theta^{(i)} \ \theta^{(j)}\right]$$

Therefore, the new joint posterior distribution is expressed as:

$$\pi\left(\boldsymbol{\theta}^{(x)}, \frac{d\boldsymbol{\theta}^{(x)}}{x}, \theta_2, \rho_\theta, \lambda_\theta | y\right) \propto \lambda_y^{\frac{N}{2}} \exp\left\{-\frac{\lambda_y}{2}\left(y - \boldsymbol{\eta}\left(\boldsymbol{\theta}^{(x)}\right)\right)^{\mathrm{T}}\left(y - \boldsymbol{\eta}\left(\boldsymbol{\theta}^{(x)}\right)\right)\right\} \lambda_y^{a_y - 1} \exp\left(-b_y \lambda_y\right)$$

$$\times \lambda_\theta^{\frac{N}{2}} |\boldsymbol{R}_{\rho\theta}|^{-\frac{1}{2}} \exp - \frac{\lambda_\theta}{2}\left(\begin{bmatrix} \boldsymbol{\theta}^{(x)} \\ \frac{d\boldsymbol{\theta}^{(x)}}{x} \end{bmatrix} - \begin{bmatrix} \mu_\theta \\ \mu'_\theta \end{bmatrix} \mathbf{1}\right)^{\mathrm{T}} \boldsymbol{K}_{\mathrm{joint}}^{-1}\left(\begin{bmatrix} \boldsymbol{\theta}^{(x)} \\ \frac{d\boldsymbol{\theta}^{(x)}}{x} \end{bmatrix} - \begin{bmatrix} \mu_\theta \\ \mu'_\theta \end{bmatrix} \mathbf{1}\right)$$

$$\times \lambda_\theta^{a_\theta - 1} \exp\left(-b_\theta \lambda_\theta (1 - \rho_\theta)^{b_\theta - 1}\right)$$

For predicting systems response at new inputs, let θ^* denote the new state-dependent parameter at new input variables so the realization of the Gaussian process follows that:

$$\begin{bmatrix} \theta \\ \theta^* \end{bmatrix} \sim N\left(, \begin{bmatrix} \mathrm{Cov}\left[\theta\ \theta\right] & \mathrm{Cov}\left[\theta\ \theta^*\right] \\ \mathrm{Cov}\left[\theta^*\ \theta\right] & \mathrm{Cov}\left[\theta^*\ \theta^*\right] \end{bmatrix}\right)$$

Based on this new posterior distribution, three new chains were generated for the same model, and the results are presented in Fig. 18.2b.

Figure 18.2 compares the performance of posterior prediction between the original and the constrained GP which shows that the sampled parameters become monotonically increasing with inputs once the constraint is enforced. The range of the one-standard-deviation bound (one-SD) is substantially reduced because the monotonicity constraint imposed on the GP helps with identifiability and, thus, reduces the remaining uncertainty in the inferred parameters. Note that in both Fig. 18.2a, b, the estimated mean values notably deviate from the true values for small values of x. One explanation could be that the noise of the system begins to be nonnegligible at these low values of x hindering the inference.

18.5   Conclusion

The chapter adopts a GP-based method to infer a nonlinear parameter in an ODE model and provides a proof-of-concept case study in a structural dynamic system. It is extended from an integral-effect experiment where the inferred parameter was treated as a function of the input variable of the physics-based model and emulated as a GP. Based on this previous study, this chapter contributes to reduce the uncertainty by enforcing monotonicity on the state-dependent variable. The comparison between the GP method with monotonicity and without monotonicity demonstrates that prior knowledge can significantly improve the prediction performance and narrow the prediction bound.

There are a number of limitations of this study that will be further investigated in future studies. First, the application used for the demonstration purpose assumes a uniformly accurate numerical model (i.e., no discrepancy bias) and is based on synthetic data. It will be critical to evaluate the performance of the proposed approach on realistic engineering applications where model bias and experimental uncertainty are also present. The computational cost of the presented methodology may be prohibitive for complex problems with large dimensionality. The authors are currently studying ways to reduce the computational demand and are deploying cGP for the response for a Duffing oscillator model of a nonlinear structural dynamic system.

References

1. Xia, Z., Tang, J.: Characterization of dynamic response of structures with uncertainty by using Gaussian processes. J. Vib. Acoust. **135**, 5 (2013)
2. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B Stat Methodol. **63**(3), 425–464 (2001)
3. Higdon, D., et al.: Combining field data and computer simulations for calibration and prediction. SIAM J. Sci. Comput. **26**(2), 448–466 (2004)
4. Stevens, G.N., et al.: Statistical inference of empirical constituents in partitioned analysis from integral-effect experiments: an application in thermo-mechanical coupling. Eng. Comput. **35**, 672–691 (2018)
5. Atamturktur, S., et al.: A resource allocation framework for experiment-based validation of numerical models. Mech. Adv. Mater. Struct. **22**(8), 641–654 (2015)
6. Bernardo, J.M., Smith, A.F.M.: Bayesian Theory, vol. 405. Wiley (2009)
7. Riihimäki, J., Vehtari, A.: Gaussian processes with monotonicity information. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings (2010)

8. Ranjan, P., Haynes, R., Karsten, R.: A computationally stable approach to Gaussian process interpolation of deterministic computer simulation data. Technometrics. **53**(4), 366–378 (2011)

9. Higdon, D., et al.: Computer model calibration using high-dimensional output. J. Am. Stat. Assoc. **103**(482), 570–583 (2008)

10. Brown, D., Andrew, and Sez Atamturktur.: Nonparametric functional calibration of computer models. Stat. Sin., 721–742 (2018)

11. O'Hagan, A.: Curve fitting and optimal design for prediction. J. R. Stat. Soc. Ser. B Methodol. **40**(1), 1–24 (1978)

12. Shinozuka, M.: Simulation of multivariate and multidimensional random processes. J. Acoust. Soc. Am. **49**(1B), 357–368 (1971)

13. Shinozuka, M.: Monte Carlo solution of structural dynamics. Comput. Struct. **2**(5–6), 855–874 (1972)

14. Ratto, M., Pagano, A., Young, P.: State dependent parameter metamodelling and sensitivity analysis. Comput. Phys. Commun. **177**(11), 863–876 (2007)

15. Avendaño-Valencia, L.D., et al.: Gaussian process time-series models for structures under operational variability. Front. Built Environ. **3**, 69 (2017)

16. Soize, C.: Uncertainty Quantification. Springer International Publishing AG (2017)

Chapter 19
Probabilistic Model Updating for Structural Health Monitoring Using a Likelihood-Free Bayesian Inference Method

Jice Zeng, Michael D. Todd, and Zhen Hu

Abstract Bayesian inference has received considerable attention and is an accredited framework in structural health monitoring (SHM) to evaluate structural integrity. In Bayesian inference, structural parameters are estimated as probability density distributions (PDF) using measurements, and associated uncertainty is then naturally quantified. However, the likelihood function as a crucial component in Bayesian inference is usually analytically intractable due to model complexity. Furthermore, solving the likelihood function is computationally demanding. This study investigates a novel likelihood-free and computationally efficient Bayesian inference method, for probabilistic damage detection through model updating in SHM. The method is based on normalizing flow and conditional invertible neutral network (cINN) and is called BayesFlow. It consists of a training phase and an inference phase. In the training phase, a summary and a cINN are trained simultaneously given synthetic data. The summary network targets on automatically learning the most useful features from raw data for damage detection rather than handcrafted ones. The cINN aims to learn the posterior distribution of model parameters by sampling a Gaussian latent distribution and using the trained inverse function. Based on the summary network and cINN, Bayesian inference can be performed efficiently without evaluating any likelihood function in the inference phase. The performance of the BayesFlow is verified with a benchmark example, an 18-story steel frame. Results show that BayesFlow provides more accurate damage identification with less measurement data and lower uncertainties compared to traditional sampling-based Bayesian inference method.

Keywords Probabilistic damage detection · Likelihood-free · Invertible neural network · Bayesian model updating

19.1   Introduction

Bayesian model updating has been extensively studied in the context of probabilistic damage detection and system identification [1, 2]. According to the way that the likelihood function in Bayesian inference is used, Bayesian model updating can be categorized into *likelihood-based approaches* and *likelihood-free approaches*. The likelihood-based approach requires the evaluation of the likelihood function given in an analytical form. For some situations, however, the likelihood function is either computationally prohibitive or numerically intractable. Motivated by tackling the challenge on likelihood-based approaches for Bayesian inference, this work proposes a novel likelihood-free and computationally efficient Bayesian inference method, named BayesFlow developed by Radev et al. [3], to Bayesian model updating and damage detection in structural health monitoring (SHM). BayesFlow successfully realizes amortized inference, in which the entire parameter estimation is split into an upfront training phase that is computationally intensive and a subsequent inference phase that is very quick to execute. BayesFlow is a fully likelihood-free approach that directly estimates the posterior distribution without repeatedly evaluating the likelihood function. It encompasses two separate neural networks—a summary network and an inference network—to complete the task of parameter inference. The former is responsible for reducing data dimensionality from potentially large time-series datasets to a fixed-size vector. The latter predicts the posterior distribution efficiently by a conditional invertible neural network (cINN). These two networks are jointly trained and aligned well for parameter inference given synthetic data generated from a forward model. This work attempts to reveal this promising potential of BayesFlow

J. Zeng · Z. Hu (✉)
Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, MI, USA
e-mail: zhennhu@umich.edu

M. D. Todd
Department of Structural Engineering, University of California San Diego, La Jolla, CA, USA

and specifically adapt BayesFlow for the purpose of probabilistic damage detection in civil infrastructures. To the best of our knowledge, BayesFlow has not yet been applied in structural model updating and probabilistic damage detection in SHM applications. The remainder of it is organized as follows. Section 19.2 briefly introduces the fundamentals of BayesFlow and its application in damage detection. An 18-story steel shear frame benchmark example is utilized to demonstrate the capability of BayesFlow in damage detection in Sect. 19.3. Finally, conclusions are drawn in Sect. 19.4.

## 19.2 Damage Detection Using a New Likelihood-Free Bayesian Inference Method

BayesFlow is built upon normalizing flow-based theory and cINN. It is proposed by Radev and co-workers for neurocognitive and epidemiology models. The goal of BayesFlow is to approximate posterior distribution of model parameters $\boldsymbol{\theta}$ for any given observations $\mathbf{y}_{1:T}$ using cINN, where $\mathbf{y}_{1:T} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T)$ and $\mathbf{y}_i, \ \forall \ i = 1, \cdots, T$ is the i-th vector of observations. In addition to cINN, BayesFlow introduces and jointly trains a summary network along with cINN to deal with high-dimensional time-series data. Measured raw data (i.e., $\mathbf{y}_{1:T}$) is summarized or filtered using summary network to a fixed-size and low-dimensional vector. Mathematically, the summary network can be represented as $\tilde{\mathbf{y}} = \varphi_{\boldsymbol{\gamma}}(\mathbf{y}_{1:T})$, where $\varphi_{\boldsymbol{\gamma}}(\cdot)$ is the summary network with parameters $\boldsymbol{\gamma}$ and $\tilde{\mathbf{y}}$ is the summarized feature from the network which will be used as $\mathbf{y}_0$ in the inference network. BayesFlow estimates neural network model parameters $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$ by minimizing the expected Kullback-Leibler (KL) divergence between the target and the approximated posteriors given observations $\mathbf{y}_{1:T}$ as below [3].

$$
\begin{aligned}
\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\omega}} &= \arg\min_{\boldsymbol{\gamma}, \boldsymbol{\omega}} E_{f_{\mathbf{Y}_{1:T}}(\mathbf{y}_{1:T})} \left[\mathrm{KL}\left[f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}|\mathbf{y}_{1:T}\right) \middle\| \hat{f}_{\boldsymbol{\theta}, \boldsymbol{\omega}}\left(\boldsymbol{\theta}|\varphi_{\boldsymbol{\gamma}}\left(\mathbf{y}_{1:T}\right)\right) \right] \right], \\
&= \arg\min_{\boldsymbol{\gamma}, \boldsymbol{\omega}} E_{f_{\mathbf{Y}}(\mathbf{y}_{1:T})} \left[E_{f_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}|\mathbf{y}_{1:T})} \left[\log\left\{ f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}|\mathbf{y}_{1:T}\right)\right\} - \log\left\{ \hat{f}_{\boldsymbol{\theta}, \boldsymbol{\omega}}\left(\boldsymbol{\theta}|\varphi_{\boldsymbol{\gamma}}\left(\mathbf{y}_{1:T}\right)\right)\right\} \right] \right], \\
&= \arg\max_{\boldsymbol{\gamma}, \boldsymbol{\omega}} E_{f_{\mathbf{Y}}(\mathbf{y}_{1:T})} \left[E_{f_{\boldsymbol{\theta}|\mathbf{Y}}(\boldsymbol{\theta}|\mathbf{y}_{1:T})} \left[\log\left\{ \hat{f}_{\boldsymbol{\theta}, \boldsymbol{\omega}}\left(\boldsymbol{\theta}|\varphi_{\boldsymbol{\gamma}}\left(\mathbf{y}_{1:T}\right)\right)\right\} \right] \right], \\
&= \arg\max_{\boldsymbol{\gamma}, \boldsymbol{\omega}} \iint f_{\boldsymbol{\theta}, \mathbf{Y}}\left(\mathbf{y}_{1:T}, \boldsymbol{\theta}\right) \log\left\{ \hat{f}_{\boldsymbol{\theta}, \boldsymbol{\omega}}\left(\boldsymbol{\theta}|\varphi_{\boldsymbol{\gamma}}\left(\mathbf{y}_{1:T}\right)\right)\right\} \mathbf{d}\boldsymbol{\theta}\mathbf{d}\mathbf{y}_{1:T},
\end{aligned}
\tag{19.1}
$$

where $f_{\mathbf{y}_{1:T}}(\mathbf{y}_{1:T})$ is the PDF of $\mathbf{y}_{1:T}$, $E[\cdot]$ is the expectation operator, $\hat{f}_{\boldsymbol{\theta},\ \boldsymbol{\omega}}(\boldsymbol{\theta}|\varphi_{\boldsymbol{\gamma}}(\mathbf{y}_{1:T}))$ is the estimated posterior of $\boldsymbol{\theta}$ for the given parameters $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$ of the cINN and summary network, and KL$[\cdot]$ is the KL divergence function. According to the theory of normalizing flow, the optimization model given in Eq. (19.1) can then be approximated using Monte Carlo simulation (MCS) [3]. After the model parameters $\hat{\boldsymbol{\gamma}}, \ \hat{\boldsymbol{\omega}}$ are estimated, BayesFlow can be employed to efficiently obtain the posterior distribution $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ for given $\mathbf{y}_{1:T}$. The overall procedures of damage detection using BayesFlow consist of an offline training phase and an online detection phase. In the offline training phase, the optimal model parameters $\hat{\boldsymbol{\gamma}}, \ \hat{\boldsymbol{\omega}}$ of the summary network and inference network are estimated. In the online detection phase, measured data $\mathbf{y}_{o, 1:T}$ are collected from the field test, then passed through the trained summary and inference networks, and thus obtain samples of the posterior distributions of the uncertain model parameters. Finally, the posterior samples of model parameters are used for probabilistic damage detection.

19.3 Case Study

An 18-story shear frame is used as an example to validate the efficacy of damage detection using BayesFlow. A one-third scale shear frame specimen was built and tested at the E-Defense shaking table in Japan [4]. Figure 19.1a shows the front and side view of the structure. This numerical study simplifies the structure as a 9-DOF shear model as shown in Fig. 19.1b. In this study, the stiffness change parameters representing the relative change of stiffness at each floor are selected as updating parameters, denoted as $\theta_1 \sim \theta_9$, where $\theta_i = (E^{\mathrm{act}} - E^{\mathrm{nom}})/E^{\mathrm{nom}}, \forall i = 1, \cdots, 9$, E^{act} and E^{nom} are, respectively, actual and nominal elastic moduli. One damage scenario with multiple damage locations at different floors is studied. Damage severity is quantified by the percentage of stiffness change, for example, relative change in elastic modulus, such as 20% stiffness reduction at the first, fourth, and eighth floors and 10% stiffness reduction at the third and seventh floors. The performance of BayesFlow on damage detection is also compared with the Differential Evolutionary Adaptive Metropolis (DREAM) method [5]. Figure 19.2 presents the comparison of posterior distributions of updated parameters obtained by BayesFlow and DREAM using different number of datasets. The results show that BayesFlow has a more stable performance in damage detection than DREAM when the number of datasets varies. The posterior mean obtained from DREAM deviates from the ground-truth, especially for θ_2, θ_5, and θ_8. As the number of datasets increases, the performance of DREAM gets better and closer to that of BayesFlow.
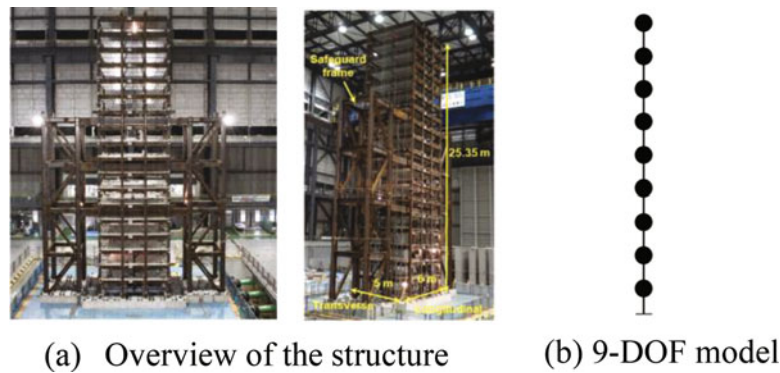
Fig. 19.1 An 18-story shear frame. (**a**) Overview of the structure. (**b**) 9-DOF model
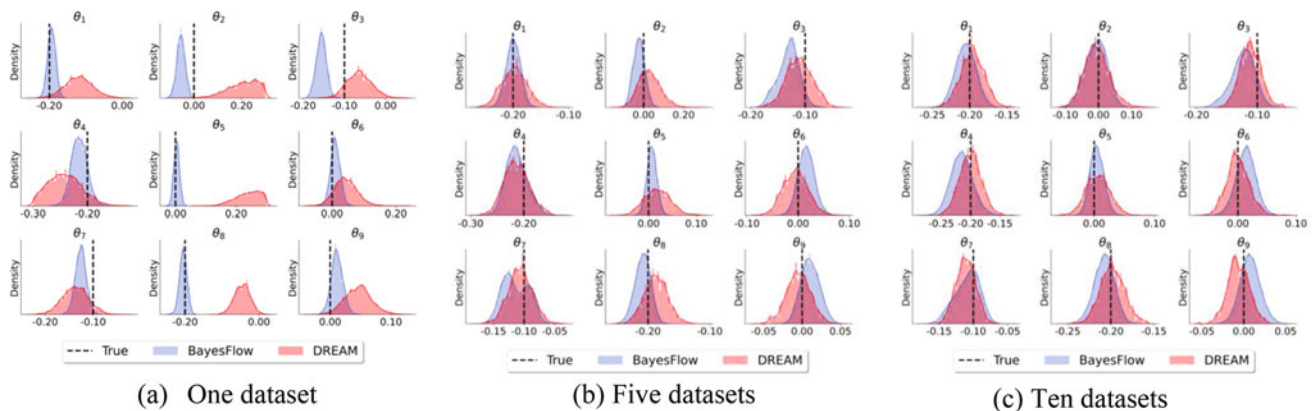

Fig. 19.2 Posterior distributions from BayesFlow and DREAM for different number of datasets. (**a**) One dataset. (**b**) Five datasets. (**c**) Ten datasets

19.4 Conclusion

In this study, we applied a novel likelihood-free Bayesian inference method, named BayesFlow, to probabilistic damage detection for an SHM application [6]. BayesFlow is fully likelihood-free, which directly approximates the posterior distribution without evaluating the likelihood function. The summary network in BayesFlow automatically learns the maximal information from data. The inference network enables nearly real-time parameter estimation once computational expensive training work is completed offline. In this study, BayesFlow is applied to an 18-story shear frame for damage detection. The results illustrate that BayesFlow exhibits superior accuracy and reliability compared to the traditional sampling method. The use of BayesFlow would provide new insights on structural damage detection and deliver a new solution for online SHM.

References

1. Beck, J.L., Katafygiotis, L.S.: Updating models and their uncertainties. I: Bayesian statistical framework. J. Eng. Mech. **124**(4), 455–461 (1998)
2. Zeng, J., Kim, Y.H.: Probabilistic damage detection and identification of coupled structural parameters using Bayesian model updating with added mass. J. Sound Vib. **539**, 117275 (2022)
3. Radev, S.T., Mertens, U.K., Voss, A., et al.: BayesFlow: learning complex stochastic models with invertible neural networks. IEEE Trans. Neural Netw. Learn. Syst. **33**(4), 1452–1466 (2020)
4. Suita, K., Suzuki, Y., Takahashi, M.: Collapse behavior of an 18-story steel moment frame during a shaking table test. Int. J. High-Rise Build. **4**(3), 171–180 (2015)
5. Vrugt, J.A.: Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation. Environ. Model Softw. **75**, 273–316 (2016)
6. Zeng, J., Todd, M.D., Hu, Z.: Probabilistic damage detection using a new likelihood-free Bayesian inference method. J. Civ. Struct. Heal. Monit., 1–23 (2022)

Chapter 20
Deep Learning for Image Segmentation and Subsurface Damage Detection Based on Full-Field Surface Strains

Ashish Pal, Wei Meng, and Satish Nagarajaiah

Abstract Damage detection plays a key role in estimating the health of the structure. Accurate damage detection allows judging the reduced capacity of the structure and further retrofitting the damage. Error in damage detection may have catastrophic consequences, especially when the damage is subsurface that may accumulate over time. In this study, a deep convolutional neural network (CNN) based on full-field strain measurements is developed to localize subsurface damage. The dataset is prepared artificially by finite element simulation of rectangular metal bars having subsurface damage of varied length, size, depth, and direction of propagation. For the trained network, the Intersection of Union score is found to be 0.72 for both training and testing set. This implies that the model can localize the subsurface damage and can be further explored for applications in nondestructive testing. For continuously generated strain maps, applications in dynamics systems to study damage initiation and propagation can be studied for dynamic loading.

Keywords Subsurface damage · Convolutional neural networks · Full-field strain map · Damage localization · Nondestructive testing

20.1 Introduction

Structures such as buildings and bridges, airplanes and ships, all are subjected to damage due to aging, environmental factors or extreme events such as storms and earthquakes. There has been a constant effort to detect those damages in time to maintain the functionality of the structure and prevent it from failing. The difficult type of damage to detect are the ones that are subsurface, which are also more fatal since they can grow beneath the surface for prolonged periods of time without detection. Therefore, timely detection of such subsurface damage is essential for the safety of the structure.

Manual inspection is the traditional method to carry out this task, which is slow, expensive, and expert knowledge is required to judge the presence of damage. It is also limited to sections of the structure that are physically reachable by a person. Subsurface damage is a major concern in a variety of structures [1–11], including wind turbine, aerospace components, reinforced concrete bridge, etc.

Several techniques are present in the literature, and a few of them are used in practice that can detect subsurface damage. These include ground-penetrating radar technique [12] and infrared thermography [5], among several others [2, 11, 13–15]. With the advancements in computer vision and machine learning methods, a few methods have also been developed to detect damage in concrete [16–21] and steel structures [22–24]. However, it is still scarce to find techniques for subsurface damage detection; the limited studies done are all based on infrared thermography [8, 25, 26].

In this study, a deep CNN based on surface strain data has been proposed to detect subsurface damage. The reason for choosing strain as an indicator is its close relation to stress concentration that directly leads to the damage information. The

A. Pal (✉) · W. Meng
Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA
e-mail: Ashish.Pal@rice.edu; wm14@rice.edu

S. Nagarajaiah
Department of Mechanical Engineering, Rice University, Houston, TX, USA

Smalley-Curl Institute, Rice University, Houston, TX, USA

Department of Material Science and NanoEngineering, Rice University, Houston, TX, USA
e-mail: satish.nagarajaiah@rice.edu

proposed CNN can perform pixel-level segmentation based on the resolution of the strain measurement. Since the input to the network are the strain maps, they can be used in static as well as dynamic scenarios. If the strain maps are continuously generated, then the damage behavior such as damage initiation and propagation can also be studied. The network architecture, data preparation using numerical simulation of aluminum bars with different damage, the training and testing process are shown in the following sections.

20.2 Deep Learning Architecture

The input to the CNN is the measured strain maps provided in the form of an image. The output from the CNN is a segmented image that shows the damage location by classifying each pixel in the image as damaged or undamaged. U-Net [27] CNN has been previously developed for the same task, but for biomedical images, naturally this architecture makes a good choice for this application as well. The U-Net architecture originally took 512×512 pixel input images, which in this study has been slightly modified to a size of 256×256 pixels. For each block, rectified linear unit (ReLU) nonlinear activation has been removed after the second convolution based on the slight increase in accuracy it was providing. In the expansion path, the final convolution operation converts the $256 \times 256 \times 64$ feature map to a single image that is then passed through the sigmoid activation layer to get the final output image that provides the localized damage information.

20.3 Dataset

For data generation, finite element numerical simulations were performed using the Ansys 2021 R1 software for rectangular bars with different damage. Except for the geometric properties of the damage, everything else was kept identical for each simulation. The material properties used are of aluminum and dimensions of each plate are $152.4\,\text{mm} \times 25.4\,\text{mm} \times 12.7\,\text{mm}$. The damage to the bars is given in the form of a cylindrical cavity of small radius. The depth of damage, size of damage, and the orientation of damage are randomly assigned to each plate. Axial load is applied along the longitudinal direction of the bar, while the strains are recorded on the surface of the plate that is closer to the damage. Figure 20.1 shows an example damage case and the recorded full-field strain data. The first image is the color-coded strain distribution on the surface of the bar due to the subsurface damage shown in the right image.

Several factors dictate the characteristic of the strain distribution observed on the surface of the bar due to the subsurface damage. The variance in the strain distribution is inversely related to the depth of damage, that is, when the damage gets deeper, the variance of the strain decrease. In simpler words, the difference between the maximum and minimum value of the strain decrease. The same can be said for the orientation of the subsurface damage. As the damage aligns with the direction of the force, the strain variations diminish as well. The most clear strain pattern is visible when the damage is perpendicular to the direction of the force. These factors put a limitation on the depth and orientation of the damage we can use for data preparation. The depth of the damage was restricted to half the thickness of the plate. The radius of the cylindrical cavity is varied between 0.76–1.52 mm with no limitation on the orientation of damage. A total of 290 simulations were performed by randomly selecting the radius of damage, length of damage, depth of damage, and orientation.
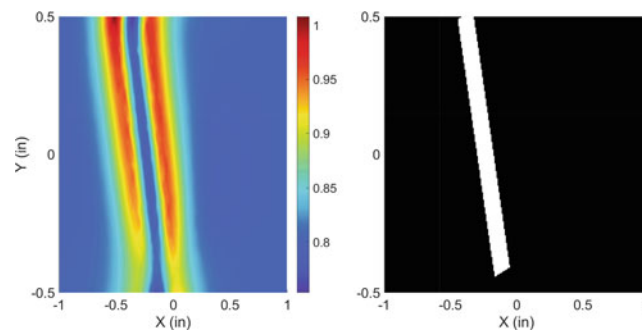

Fig. 20.1 Projection of the subsurface damage and its corresponding strain pattern

20.4 Data Augmentation

The U-Net network is huge in size with millions of parameters to find out. A set of only 290 examples is not enough to train the data. A large amount of data is generally needed to train such large networks. To increase the dataset using the available 290 example cases, several damage augmentation techniques have been adopted. Data augmentation is a well-known technique and more often than not is adopted to train deep networks such as U-Net. A few augmentation methods that were used in this study are listed below:

1. Random horizontal flip
2. Random vertical flip
3. Random rotation with 10°
4. Random resize cropping between scale 0.5 and 1

A standard practice during training in any deep learning network is to normalize the data before feeding them as an input to the CNN. Normalization is normally done by subtracting the ensemble mean and dividing each image by the ensemble standard deviation. However, given the nature of the data used in this study, the normalization was done by making each image zero mean and unit variance. This is because the magnitude of the strains is dependent on the force applied. This makes each strain map independent of the applied force, which might not be the case if we use ensemble mean and standard deviation for normalization.

20.5 Training and Hyperparameters

As mentioned earlier, the 256×256 input image is passed through an encoder and then a decoder that produces a feature map of size 256×256. This final feature map is then passed through a sigmoid activation layer, and the loss is calculated using the binary cross entropy loss. The expression of the loss term is written as follows:

$$l = -[y \cdot \log\sigma(x) + (1 - y) \cdot \log(1 - \sigma(x))] \tag{20.1}$$

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{20.2}$$

where y is the true class, x is the final feature map, and l is the loss of the associated pixel. The optimization algorithm used is the Adam optimization [28] with an adaptive learning rate to decrease the learning rate by 10% for an insignificant decrease (5%) in a loss for 5 epochs. The regularization is based on the L2 norm of the weights, and the initialization is assumed according to normal Glorot initialization [29]. The batch size for the training set is kept at 12 (the maximum size allowed by the NVIDIA RTX 2060 6GB graphics card), and the training is performed for 200 epochs.

20.6 Hyperparameter Tuning

While training a CNN network, there are several parameters whose values are not known and need to be found out using a hyperparameter tuning procedure. To do so, the dataset of 290 images is divided into three sets containing 145 training images, 87 validation images, and 58 testing images. Three hyperparameters are selected for tuning, and a complete grid search id performed in a space of $3 \times 4 \times 4$ points for learning rate (LR, 5e–4, 1e–4, and 5e–5), strength of regularization (SR, 0,1e–5,1e–4,1e–3), and positive weights assigned to the damaged pixels in the image (1, 2, 3, 5). A total of 48 models are trained and accuracy of each model is found on the validation set. The measure of the quality of the CNN is based on the Intersection over Union (IoU) score, represented as

$$IoU = \frac{TP}{TP + FP + FN} \tag{20.3}$$

where TP are true positive cases, FP are false positive, and FN are the false negative cases. The IoU score for the optimal model was found to be 0.72 and the hyperparameters turned out to be LR = 1e–4, SR = 1e–5, and weight = 2.

20.7 Predictions

The testing data is the set of images that are set aside during the training and validation procedure on which the final performance of the trained model is checked. The optimal model found during the training phase was tested on such unseen cases that were selected randomly prior to the training. Figure 20.2 shows the illustration of six example cases with different subsurface damage. In Fig. 20.2a, b, both the damages are almost vertical that has a characteristic strain pattern with two strips of high strain on either side of the damage and a single strip of low strain on top of the damage. The next three Fig. 20.2c, d, e show the damage oriented toward the direction of force. Now, the strain pattern shows a different pattern with a single strip of high strain with diminishing strains away from the damage. The spread of diminishing strains increases as the damage aligns with the direction of force. Figure 20.2f shows the case where the damage is aligned with the direction of force. The strain pattern is on the same lines as the previous three cases, just the spread is even more and variance has decreased, which makes it difficult to locate the damage. The strip of high strain is still above the damage, but it is not constant throughout. There is spatial variation along the direction of damage with peak strain at the ends of damage and lower strains toward the middle portion of the damage. In all cases, the damage is detected with excellent localization. The orientation of damage,
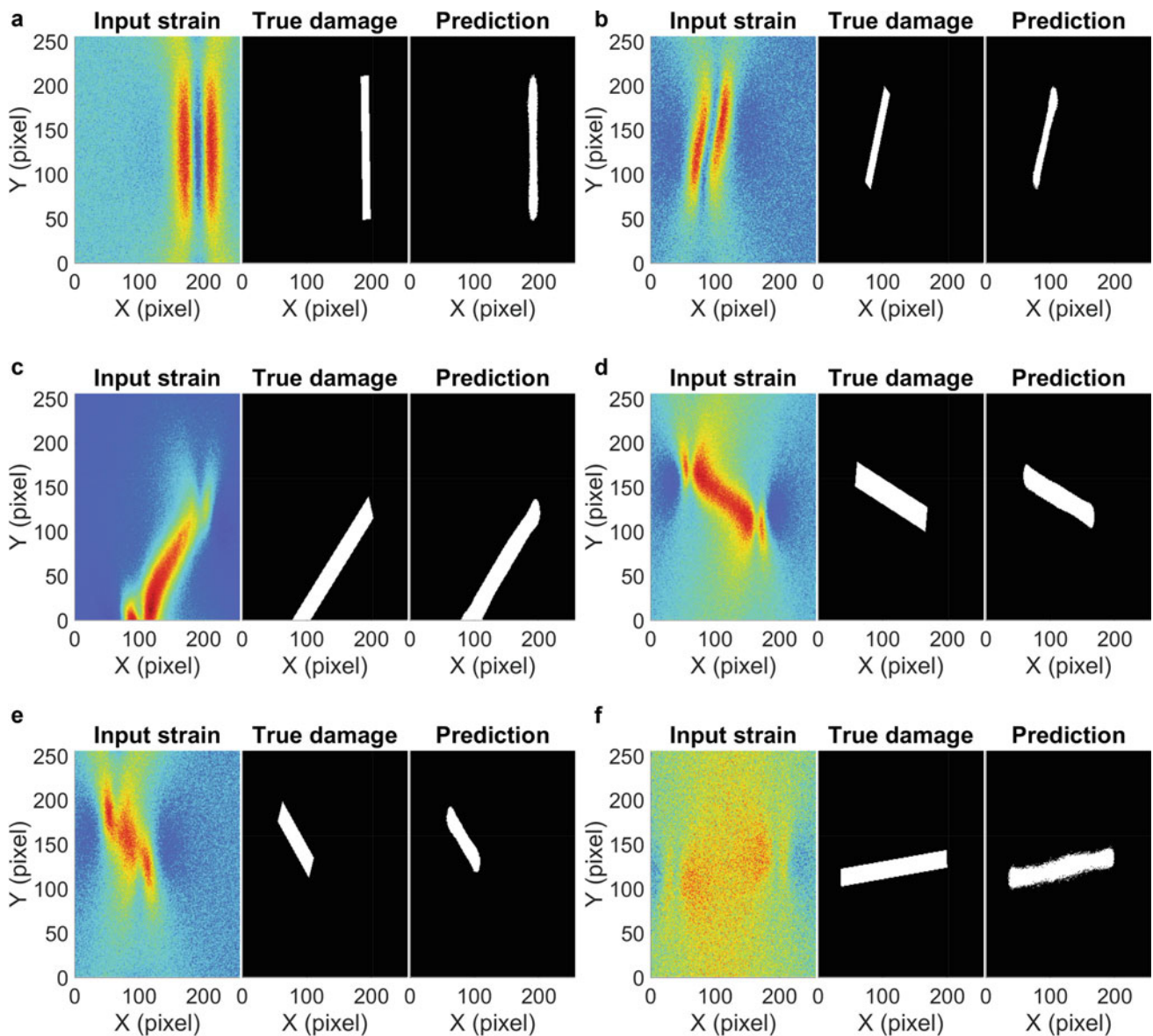

Fig. 20.2 The input strain map, true damage, and predicted damage from the trained model for selected aluminum example cases

thickness, and length is detected with good accuracy. Even when the strain pattern continuously changed from case (a)–(f), the network was able to observe the change and make accurate predictions. The boundary of the damage is also clean for all cases except at the ends or in case (f); however, the error is still low and it does not affect the quality of the prediction. Given the accuracy of the predictions made by the network, it can be used as a reliable predictor of subsurface damage using the full-field surface strain distribution.

20.8   Conclusion

Adopted from U-Net architecture, a deep learning CNN based on full-field surface strains has been proposed to localize the subsurface damage. The network is capable of pixel-level segmentation that classifies each pixel of the output image as damaged or undamaged. The data was prepared using finite element simulation of aluminum bars with different subsurface damage. Following a grid search on three hyperparameters, the optimum network was found to have an IoU score of 0.72 for the validation set. The predictions of the optimum model on test set showed that the proposed CNN is able to pick up the correct location of damage, orientation of the damage, and length of the damage. In a few cases, the shape of the boundary of the damage is not as clean as in other cases; still, the quality of predictions is reasonably good. The overall performance of the network is satisfactory, and it can potentially be used for subsurface damage localization as a nondestructive testing method.

References

1. Longching, C., Qing, C., Eryu, S: Study on initiation and propagation angles of subsurface cracks in GCr15 bearing steel under rolling contact. Wear **133**(2), 205–218 (1989)
2. Fuentes, R., Dwyer-Joyce, R.S., Marshall, M.B., Wheals, J., Cross, E.J.: Detection of sub-surface damage in wind turbine bearings using acoustic emissions and probabilistic modelling. Renew. Energy **147**, 776–797 (2020)
3. Mehnen, J., Tinsley, L., Roy, R.: Automated in-service damage identification. CIRP Ann. **63**(1), 33–36 (2014)
4. Gupta, S., Loh, K.J.: Characterization and localization of sub-surface structural features using non-contact tomography. In: Smart Materials, Adaptive Structures and Intelligent Systems, vol. 50480, p. V001T05A007. American Society of Mechanical Engineers, New York (2016)
5. Hiasa, S., Birgul, R., Necati Catbas, F.: Investigation of effective utilization of infrared thermography (IRT) through advanced finite element modeling. Constr. Build. Mater. **150**, 295–309 (2017)
6. Kim, Y.J., Jofre, L., De Flaviis, F., Feng, M.Q.: Microwave reflection tomographic array for damage detection of civil structures. IEEE Trans. Antennas Propag. **51**(11), 3022–3032 (2003)
7. Rathod, H., Gupta, R.: Sub-surface simulated damage detection using non-destructive testing techniques in reinforced-concrete slabs. Constr. Build. Mater. **215**, 754–764 (2019)
8. Ali, R., Cha, Y.-J.: Subsurface damage detection of a steel bridge using deep learning and uncooled micro-bolometer. Constr. Build. Mater. **226**, 376–387 (2019)
9. Guo, J., Chen, J., Li, J., Fang, Q., Liu, Y.: Study on subsurface-inclined crack propagation during machining of brittle crystal materials. Appl. Phys. A **122**(5), 493 (2016)
10. Li, K., Liao, T.W.: Surface/subsurface damage and the fracture strength of ground ceramics. J. Mater. Process. Technol. **57**(3–4), 207–220 (1996)
11. Yin, J., Bai, Q., Zhang, B.: Subsurface damage detection on ground silicon wafers using polarized laser scattering. J. Manufact. Sci. Eng. **141**(10), 101012 (2019)
12. Liu, J., Zollinger, D.G., Lytton, R.L.: Detection of delamination in concrete pavements using ground-coupled ground-penetrating radar technique. Transp. Res. Rec. **2087**(1), 68–77 (2008)
13. Feng, M.Q., De Flaviis, F., Kim, Y.J.: Use of microwaves for damage detection of fiber reinforced polymer-wrapped concrete structures. J. Eng. Mech. **128**(2), 172–183 (2002)
14. Takamatsu, R., Fujisawa, K., Nakahata, K., Murakami, A.: Shape detection of multiple subsurface cavities by particle filtering with elastic wave propagation. International J. Numer. Analy. Methods Geomech. **44**(15), 2025–2041 (2020)
15. Mian, A., Newaz, G., Han, X., Mahmood, T., Saha, C.: Response of sub-surface fatigue damage under sonic load-a computational study. Composit. Sci. Technol. **64**(9), 1115–1122 (2004)
16. Cha, Y.-J., Choi, W., Büyüköztürk, O.: Deep learning-based crack damage detection using convolutional neural networks. Comput.-Aided Civil Infrastruct. Eng. **32**(5), 361–378 (2017)
17. Kim, B., Cho, S.: Automated vision-based detection of cracks on concrete surfaces using a deep learning technique. Sensors **18**(10), 3452 (2018)
18. Bhowmick, S., Nagarajaiah, S., Veeraraghavan, A.: Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from uav videos. Sensors **20**(21), 6299 (2020)

19. Dung, C.V. et al.: Autonomous concrete crack detection using deep fully convolutional neural network. Autom. Constr. **99**, 52–58 (2019)
20. Yang, Y., Nagarajaiah, S.: Dynamic imaging: real-time detection of local structural damage with blind separation of low-rank background and sparse innovation. J. Struct. Eng. **142**(2), 04015144 (2016)
21. Yang, Y., Sun, P., Nagarajaiah, S., Bachilo, S.M., Weisman, R.B.: Full-field, high-spatial-resolution detection of local structural damage from low-resolution random strain field measurements. J. Sound Vib. **399**, 75–85 (2017)
22. Xu, Y., Bao, Y., Chen, J., Zuo, W., Li, H.: Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images. Struct. Health Monitor. **18**(3), 653–674 (2019)
23. Chen, F.-C., Jahanshahi, M.R.: NB-FCN: Real-time accurate crack detection in inspection videos using deep fully convolutional network and parametric data fusion. IEEE Trans. Instrument. Measure. **69**(8), 5325–5334 (2019)
24. Gulgec, N.S., Takáč, M., Pakzad, S.N.: Convolutional neural network approach for robust structural damage detection and localization. J. Comput. Civil Eng. **33**(3), 04019005 (2019)
25. Omar, T., Nehdi, M.L.: Remote sensing of concrete bridge decks using unmanned aerial vehicle infrared thermography. Autom. Constr. **83**, 360–371 (2017)
26. Omar, T., Nehdi, M.L., Zayed, T.: Infrared thermography model for automated detection of delamination in RC bridge decks. Constr. Build. Mater. **168**, 313–327 (2018)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer, Berlin (2015)
28. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2014). Preprint arXiv:1412.6980
29. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256. JMLR Workshop and Conference Proceedings (2010)

Chapter 21
A Spatio-Temporal Model for Response and Distributed Wave Load Estimation on Offshore Wind Turbines

Karin L. Yu, Konstantinos E. Tatsis, Vasilis K. Dertimanis, Eleni N. Chatzi, and Andrew W. Smyth

Abstract Sequential Bayesian inference schemes show tremendous potential for online information extraction from sparsely instrumented, uncertain dynamical systems. Within this context, notable paradigms are the tasks of state, input-state, and joint input-state-parameter estimation. A problem that has been scarcely studied in this context is the concurrent estimation of dynamic states and distributed loads on the basis of output-only (response) measurements. Examples of particular practical interest include the estimation of wind pressure on wind turbine blades, high-rise buildings and bridges, as well as wave loading in offshore structures. In such cases, the sensing of distributed inputs is heavily constrained by the instrumentation cost and the oftentimes limited access for sensor deployment. To tackle this challenge, this contribution investigates the fusion of Gaussian process regression (GPR) models with physics-based system representations for the recursive state and distributed wave load estimation on monopile offshore wind turbines. In particular, the distributed excitation is modeled with a GPR, which enables the implementation of a spatio-temporal filtering for the input process, while the system dynamics are represented by a physics-based model, which is in turn tailored to a recursive Bayesian scheme for the solution of the state estimation problem. The proposed approach is assessed in terms of a simulated case study on the finite element model of an offshore wind turbine.

Keywords Sequential Bayesian inference (SBI) · Input-state estimation · Dual Kalman filter · Gaussian process regression (GPR) · Distributed input estimation

21.1 Introduction

In the context of virtual sensing using output-only vibration measurements, several Bayesian approaches have been proposed for concurrently estimating the unmeasured input and partially observed dynamic states in both linear [1] and nonlinear systems [2]. The problem of joint input-state estimation has been initially investigated for state-space models without a direct transmission matrix [3, 4] and recently extended in order to address the lack of optimality in terms of the mean squared error [5]. The numerical instabilities associated with the latter have been more recently addressed [6], while the state-augmentation techniques have been employed as a remedy in this regard [7]. The numerical issues related to such augmented representations were recently resolved by a dual formulation, namely the dual Kalman filter (DKF) [1, 8], while the problem of input-state estimation has been also solved using a smoothing approach [9]. These techniques require an evolution model of the input to be put in place, which is often assumed to follow a random walk assumption. Recent works [10–12] explore the adoption of Gaussian process regression (GPR) models [13] instead of the modeling of input signals.

Whilst typical input-state estimation settings assume inputs to be mostly represented by a finite set of point loads, which are uncorrelated, the more realistic setting of tackling systems whose dynamics are driven by distributed loads is not yet tackled. A first contribution to that end appears in previous work of the authoring team [17], where a Gaussian process (GP) model with the DKF was proposed for recursively estimating the state and distributed input in dynamical systems using a limited number of input measurements. In this contribution, we extend that work by fusing GPR models with physics-based

K. L. Yu · K. E. Tatsis (✉) · V. K. Dertimanis · E. N. Chatzi
Institute of Structural Engineering, ETH Zürich, Zürich, Switzerland
e-mail: karin.yu@ibk.baug.ethz.ch; tatsis@ibk.baug.ethz.ch; v.derti@ibk.baug.ethz.ch; chatzi@ibk.baug.ethz.ch

A. W. Smyth
Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY, USA
e-mail: aws16@columbia.edu

realistic system representations. The problem of recursive state and distributed input estimation is examined on the wave-induced vibrations of monopile offshore wind turbines. To this end, the distributed excitation is modeled with a GPR, which enables the implementation of a spatio-temporal filtering for the input process, while the system dynamics are represented by a physics-based model, which is in turn tailored to a recursive Bayesian scheme for the solution of the state estimation problem. The proposed approach is assessed in terms of a simulated case study on the finite element model of an offshore wind turbine.

The chapter is organized as follows: the mathematical formulation of the state and distributed input estimation problem for linear systems is presented in the first section. The data-driven representation of distributed loads by means of GPR models is presented in the second section, along with the posterior estimates obtained by means of sequential Bayesian inference. Lastly, the third and fourth sections are focused on the estimation of distributed wave loads on monopile offshore wind turbines.

21.2 Problem Formulation

In the context of virtual sensing, which constitutes the scope of this work, the dynamics of structural or mechanical systems are typically represented by first principles models, such as finite element (FE) representations. For linear time-invariant systems, these models assume the form of the following second-order differential equation:

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{S}_\mathrm{p}\mathbf{p}(t) \tag{21.1}$$

where $\mathbf{u}(t) \in \mathbb{R}^n$ is the displacement response vector, with n denoting the number of degrees of freedom (DOFs), $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K} \in \mathbb{R}^{n \times n}$ are the global mass, damping, and stiffness matrices. $\mathbf{S}_\mathrm{p} \in \mathbb{R}^{n \times n_\mathrm{p}}$ represents a selection matrix acting on the force vector $\mathbf{p}(t) \in \mathbb{R}^{n_\mathrm{p}}$.

The equation of motion, as formulated in Eq. (21.1), can be written in a state-space representation, which can be further transformed into a digital one by, for example, applying zero-order hold sampling. When additional noise is considered, the following linear stochastic difference equations for the discrete-time linear state-space model can be obtained:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{p}_k + \mathbf{w}_k \tag{21.2a}$$

$$\mathbf{y}_k = \mathbf{G}\mathbf{x}_k + \mathbf{J}\mathbf{p}_k + \mathbf{v}_k \tag{21.2b}$$

where $\mathbf{x}_k = \mathbf{x}(k\Delta t)$ is the state vector containing the displacement and velocity vectors at time instant $t = k\Delta t$, where Δt is the time step of the discretization scheme, $\mathbf{y}_k \in \mathbb{R}^{n_y}$ the output vector, and $\mathbf{p}_k \in \mathbb{R}^{n_p}$ the discretized input vector. The noise terms $\mathbf{w}_k \in \mathbb{R}^{2n}$ and $\mathbf{v}_k \in \mathbb{R}^{n_y}$ are assumed white, zero-mean, and uncorrelated Gaussian noise processes with covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ respectively. The transformation of Eq. (21.1) to a state-space form as well as the expressions of matrices $\mathbf{A} \in \mathbb{R}^{2n \times 2n}$, $\mathbf{B} \in \mathbb{R}^{2n \times n_p}$, $\mathbf{G} \in \mathbb{R}^{n_y \times 2n}$, and $\mathbf{J} \in \mathbb{R}^{n_y \times n_p}$ is widely documented in literature, and the reader is indicatively referred to [14].

In the context of recursive input and state estimation problems, the size of the input vector, which depends on the number n_p of forces driving the system dynamics, is a decisive parameter for the conditioning and performance of Bayesian estimators. This implies that the estimation of unknown inputs can become quickly ill-conditioned as the number of forces to be estimated increases, which is an intrinsic characteristic of systems subjected to multiple excitation forces, such as distributed loads. A workaround to this issue consists of identifying the projection of the inputs to some reduced-space [15, 16], instead of the physical inputs themselves, which further lifts the requirement of knowing the structure of $\mathbf{S}_\mathrm{p}$ in advance. An alternative methodology, which is further investigated below, for systems that are excited by distributed loads was initially proposed in [17] and consists of representing the input space with a spatio-temporal process; namely, a GPR model.

21.3 Gaussian Process Model

GPR is a data-driven supervised learning model that aims to identify the underlying function between an input and output space based on a known input–output data set. Such a representation offers a computationally cheap surrogate model and has the advantage of being a nonparametric Bayesian approach [13]. In the context of structural dynamics and specifically for the input-state estimation problem, GPR models are used for the representation of distributed loads, which are derived from a spatio-temporal process $p(\mathbf{s}, t)$, with $\mathbf{s} \in \mathbb{R}^{n_{\mathrm{d}}}$ denoting the spatial coordinate. This implies that the force vector $\mathbf{p}(t)$ consists of evaluations of the process $p(\mathbf{s}, t)$ in space:

$$\mathbf{p}(t) = \left[p(\mathbf{s}_1, t) \; p(\mathbf{s}_2, t) \; \ldots \; p(\mathbf{s}_{n_{\mathrm{p}}}, t) \right]^{\mathrm{T}} \tag{21.3}$$

where $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{n_{\mathrm{p}}}$ denote the corresponding locations of the applied inputs, which themselves depend on the spatial discretization of the model corresponding to Eq. (21.1) and the structure of the selection matrix $\mathbf{S}_{\mathrm{p}}$, and $\square^{\mathrm{T}}$ denotes the transpose of a matrix.

Due to the Gaussian assumption, the input process $p(\mathbf{s}, t)$ is defined by the mean value $\boldsymbol{\mu}(\mathbf{s}, t)$ and the covariance function $C^{\mathrm{p}}(\mathbf{s}, \mathbf{s}'; t, t')$, which captures small-scale spatial variations [18]:

$$p(\mathbf{s}, t) \sim \mathrm{GP}\left(\boldsymbol{\mu}(\mathbf{s}, t), \; C^{\mathrm{p}}(\mathbf{s}, \mathbf{s}'; t, t') \right) \tag{21.4}$$

It is herein assumed that a number of input observations are indirectly available at the locations where acceleration responses are measured instead, which are denoted by $\mathbf{s}_i^{\mathrm{m}} \in \mathbb{R}^{n_{\mathrm{d}}}$, for $i \in 1, 2, \ldots, n_{\mathrm{s},\mathrm{p}}$, and that the mean function can be described by the following regression model:

$$\boldsymbol{\mu}(\mathbf{s}, t) = \mathbf{f}^{\mathrm{T}}(\mathbf{s})\, \mathbf{b}(t) \tag{21.5}$$

where $\mathbf{f}(\mathbf{s}) = \left[f_1(\mathbf{s}) \; f_2(\mathbf{s}) \; \ldots \; f_{n_{\mathrm{r}}}(\mathbf{s}) \right]^{\mathrm{T}} \in \mathbb{R}^{n_{\mathrm{r}} \times 1}$ are orthonormal deterministic basis functions, whose number determines the regression order n_{r}, and $\mathbf{b}(t) = \left[b_1(t) \; b_2(t) \; \ldots \; b_{n_{\mathrm{r}}}(t) \right]^{\mathrm{T}} \in \mathbb{R}^{n_{\mathrm{r}} \times 1}$ contains the corresponding time-dependent regression coefficients. For identifiability reasons, the regression order, which essentially represents the number of input parameters to be estimated, should be less than or equal to the number $n_{\mathrm{s},\mathrm{p}}$ of acceleration measurements, which are used to provide information related to the input.

The input model introduced in Eq. (21.4) can be written in a state-space representation, which enables the recursive estimation, by means of sequential Bayesian inference, of the distribution of the input process $p(\mathbf{s}, t)$. This is herein performed by means of a DKF approach, as proposed in [17], which results in the following posterior estimates $\hat{p}_k(\mathbf{s})$ of the input process at time instant k, given measurements up to and including step k [19]:

$$\hat{p}_k(\mathbf{s}) = \mathbf{f}^{\mathrm{T}}(\mathbf{s})\mathbf{b}_k + \mathbf{c}_k^{\mathrm{m},\mathrm{p}\,\mathrm{T}}(\mathbf{s}) \left(\mathbf{C}_{k,0}^{\mathrm{p}^{\mathrm{m}}} \right)^{-1} \left(\mathbf{p}_k^{\mathrm{m}} - \mathbf{F}^{\mathrm{m},\mathrm{p}}\mathbf{b}_k \right) \tag{21.6}$$

where

$$\mathbf{c}_k^{\mathrm{m},\mathrm{p}}(\mathbf{s}) = \left[C^{\mathrm{p}}(\mathbf{s}, \mathbf{s}_j^{\mathrm{m}}; t_k) \right]_{j=1}^{n_{\mathrm{s},\mathrm{p}}\;\mathrm{T}} \in \mathbb{R}^{n_{\mathrm{s},\mathrm{p}} \times 1} \tag{21.7a}$$

$$\mathbf{C}_k^{\mathrm{m},\mathrm{p}} = \left[C^{\mathrm{p}}(\mathbf{s}_i^{\mathrm{m}}, \mathbf{s}_j^{\mathrm{m}}; t_k) \right]_{i,j=1}^{n_{\mathrm{s},\mathrm{p}}} \in \mathbb{R}^{n_{\mathrm{s},\mathrm{p}} \times n_{\mathrm{s},\mathrm{p}}} \tag{21.7b}$$

$$\mathbf{C}_{k,0}^{\mathrm{p}^{\mathrm{m}}} = \mathbf{C}_k^{\mathrm{m},\mathrm{p}} + \mathbf{R}_k \tag{21.7c}$$

In the above equations, $\mathbf{p}_k^m \in \mathbb{R}^{n_{s,p} \times 1}$ denotes the vector of input signals that can be estimated at the locations of acceleration observations, covariance matrix $\mathbf{C}_{k,0}^{p^m} \in \mathbb{R}^{n_{s,p} \times n_{s,p}}$, $\mathbf{R}_k$ is the covariance matrix of the measurement noise term, and $\mathbf{F}^{m,p} = \left[\mathbf{f}(\mathbf{s}_1^m) \; \mathbf{f}(\mathbf{s}_2^m) \; \dots \; \mathbf{f}(\mathbf{s}_{n_{s,p}}^m) \right]^T \in \mathbb{R}^{n_{s,p} \times n_r}$ is a matrix that contains the evaluations of the basis functions at the locations of acceleration measurements.

21.4 Case Study

The proposed methodology is demonstrated on a simulated offshore wind turbine, which is subjected to spatially distributed wave loads. The wind turbine support structure is modeled based on the NREL 5-MW reference offshore wind turbine defined by the National Renewable Energy Laboratory [20]. Since the exact wind turbine substructure heavily depends on site-specific conditions such as mean sea level, soil properties, and wave intensity, in this case study we focus on the analysis of a monopile with rigid foundation in shallow water. The support structure is modeled as a cantilever beam using two-noded Euler–Bernoulli beam elements. It is assumed that only the substructure is affected by wave loading. The 87.6 m high wind turbine tower with varying cross-section is composed of 10 elements, whereas the 30 m high substructure with constant cross-sectional area has 30 elements, as the goal is to investigate distributed loads acting on the underwater part of the monopile. In order to account for additional weights from connections, stiffeners, and paint, the mass density of steel has been increased to 8,500 kg/m^3. The material properties are $E = 210$ GPa and $\nu = 0.3$. The mean water depth is assumed to be 20 m, and the effect of marine growth on the substructure is omitted. A Rayleigh damping is assumed, with a ratio of 2% for the first two vibration modes.

For the hydrodynamic loads exerted on the substructure, irregular waves are considered. These are deterministic linear design waves obtained from the linear wave theory, which are derived by the JONSWAP (Joint North Sea Wave Project) wave spectrum [21]. It should be noted that breaking waves are neglected in this study as they exhibit a nonlinear and highly variable behavior. The linear wave theory only describes the horizontal water particle movement until mean sea level respective for still water level; however, above mean sea level, it leads to overestimations and therefore an inaccurate behavior. In order to account for this free-surface effect, different modifications of the linear wave theory have been proposed, such as first- and second-order stretching, and linear extrapolation. Here, we adopt Wheeler's stretching, which comprises a typically adopted first-order stretching method [22]. For the generation of a representative sea state, a mean wind speed of $U_w = 10$ m/s is considered, and the wave conditions are translated into hydrodynamic loads on the monopile using the semi-empirical Morison's equation for slender and small cylindrical structures. These loads are calculated from the sum of drag and inertia forces, considering both water particle velocity and acceleration [23]. The artificial measurement data are generated from a 600 s simulation period, which is performed with a sampling frequency of 100 Hz. The data is corrupted with 5% white Gaussian noise and subsequently used for the solution of the inverse problem, whose steps are schematically depicted in Fig. 21.1.
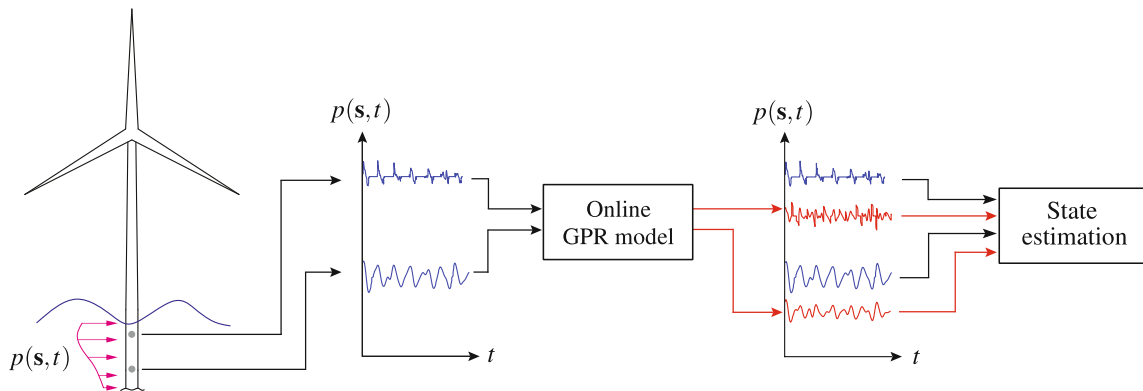

Fig. 21.1 Schematic representation of the spatio-temporal model for wave load and state estimation on offshore wind turbines

21.5   Results

The results presented in this section are based on a data set obtained from a simulation period of 600 s, which is used for the training and testing of the dual estimation scheme. Namely, the data points of the first 100 s of the simulation period are used for the training of the GPR model, which essentially comprises the extraction of the basis functions. The remaining 500 s are used for the testing of the algorithm in terms of the response and distributed input estimation accuracy. The results shown in the figures below are based on the sensor layout presented in Fig. 21.2, which consists of five accelerometers at DOFs 17, 44, 50, 59, and 68 delivering measurements in x direction (parallel to the mean sea level), two inclinometers at DOFs 42 and 93 measuring the rotation angle in y direction, and one strain measurement at DOF 56 measuring the strain component in z direction (axial). In this set-up, the location of the acceleration sensors has been optimized for input estimation using a grid search approach on the training data set, minimizing the standardized mean squared error.

The universal Kriging predictor described in Eq. (21.6), which is herein used for the estimation of the distributed hydrodynamic loads, assumes directly measured input observations. However, measuring wave forces directly is less practical and realistic. Hence, the inputs at measured acceleration locations are estimated from the equation of motion at the locations where acceleration measurements are available. The basis functions of the GPR model are determined via proper orthogonal decomposition, which is carried out during the training phase of the GPR model. In this phase, a number of different sea states is generated by means of the JONSWAP spectrum and the corresponding hydrodynamic load profiles at different time instances are collected in a so-called snapshot matrix, which is thereafter factorized by means of a singular value decomposition. The truncated eigenvectors of this decomposition are used as basis functions, thus resulting in a regression order of $n_r = 4$, which is one order less than the maximum allowable number, as dictated by the number of acceleration sensors at the region of the distributed hydrodynamic loads. Typically, it is more efficient to select a regression order equal to the number of available acceleration measurements since this would enable a more accurate representation of the spatial load profile. However, it is seen in this case study that the maximization of the regression order, which would imply the utilization of acceleration sensors in the vicinity of the mean sea level, might lead to a strong bias in the distributed load estimation, which is owed to the discontinuity of the load field at the water surface. For the covariance function, a separable stationary kernel of the Matérn class with $\nu = 3/2$ is selected.

The filtering step is initialized with a zero state vector $\mathbf{x}_0$, while the corresponding error covariance matrix $\mathbf{P}_0$ is assumed diagonal. The values assigned to this covariance matrix are split into two block matrices; one related to the displacements, whose initial values are $10^{-9} \times \mathbf{I}$, and a second one related to velocities, which is initialized with $10^{-6} \times \mathbf{I}$. Unit values are assigned to the initial vector of the input process regression coefficients $\mathbf{b}_0$, while the corresponding initial covariance matrix is equal to $10^4 \times \mathbf{I}$. The covariance matrix of the process noise $\mathbf{Q}_k$ is considered diagonal, where the values of the first n diagonal entries that correspond to displacement quantities are equal to 10^{-11} and each remaining entry corresponding to velocity terms is equal to 10^{-7}. Lastly, the covariance matrix of the measurement noise term is $\mathbf{R}_k = 2.5 \cdot 10^8 \times \mathbf{I}$, and the process noise of the colored noise process $\mathbf{Q}^\xi$ is equal to $5 \cdot 10^8 \times \mathbf{I}$.
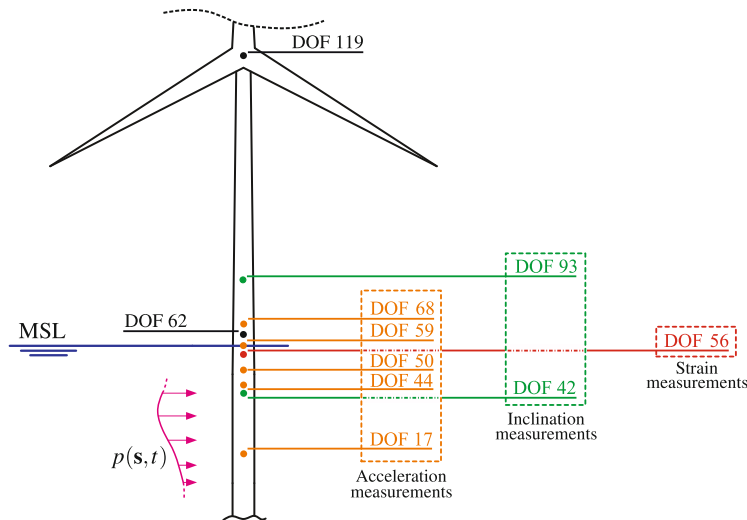

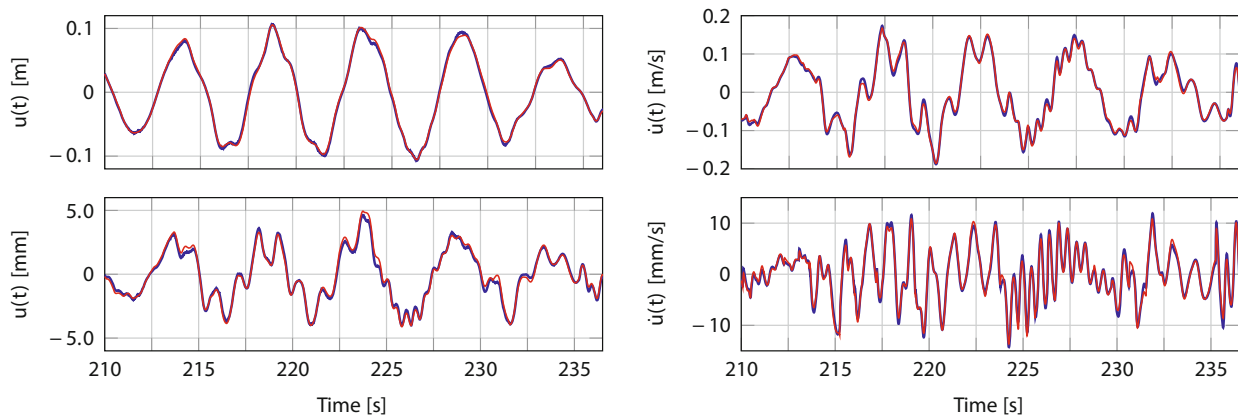**Fig. 21.2** Locations of the sensors for the input-state estimation problem

Fig. 21.3 Comparison of the vibration response estimates (red) with the actual signals (blue) for DOF 59 (bottom), at the mean sea level, and DOF 119 (top), at the tower top
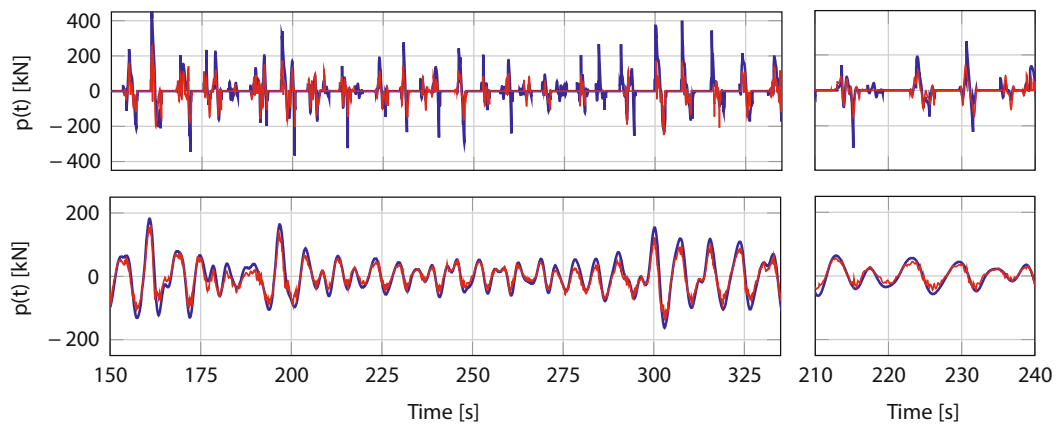

Fig. 21.4 Comparison of input estimates (red) with the actual input (blue) at DOF 44 (bottom) and DOF 62 (top)

The time history of the vibration response at DOFs 59 and 119, whose locations are displayed in Fig. 21.2, is shown in Fig. 21.3 in terms of displacements and velocities. The estimated signals are displayed in red color, while the actual system response is plotted in blue. It is seen that the state is accurately estimated in both measured and unmeasured locations. The corresponding time history estimates of the distributed loads are illustrated in Fig. 21.4, which show the estimated hydrodynamic loads at DOFs 44 and 62. With the former location being constantly underwater during the simulation period, it is observed that the hydrodynamic loads are characterized by a smooth variation, which in combination with the availability of acceleration measurements results in very accurate estimates of the forcing term. On the contrary, DOF 62 is above the mean sea level, which implies that the hydrodynamic loads are discontinuous due to the fluctuation of the water surface. This inevitably results in a less accurate estimation of the load at this position, as shown in Fig. 21.4 (top); however, the accuracy can be further improved by accordingly modifying the basis functions of the GPR model.

The accuracy of the estimated hydrodynamic loads is further illustrated in Fig. 21.5, which presents four different snapshots of the estimated load profile along the height of the substructure. The estimated profile is plotted in red, while the blue curve represents the actual loads. The red marks correspond to the locations where acceleration measurements are available, thus allowing the direct estimation of the hydrodynamic forces. The first snapshot, shown in Fig. 21.5a, presents the estimated profile at the time instant at which the largest squared error of the estimated load at DOF 62 is observed. It should be noted that there are no acceleration measurements at this DOF, which lies above the mean sea level, as shown in Fig. 21.2 and as such, the quality of estimates is limited due to the discontinuity induced by the water surface fluctuations. The snapshots shown in Fig. 21.5b and c represent the general performance of the input estimator, while Fig. 21.5d presents the profile estimate at the time instant that corresponds to the largest squared error at DOF 44. The latter emphasizes the significance of (i) the regression order selection and (ii) the basis functions, which should be chosen such that the behavior at the vicinity of the mean sea level is well modeled. In this regard, it should be noted that there are no limitations related to the
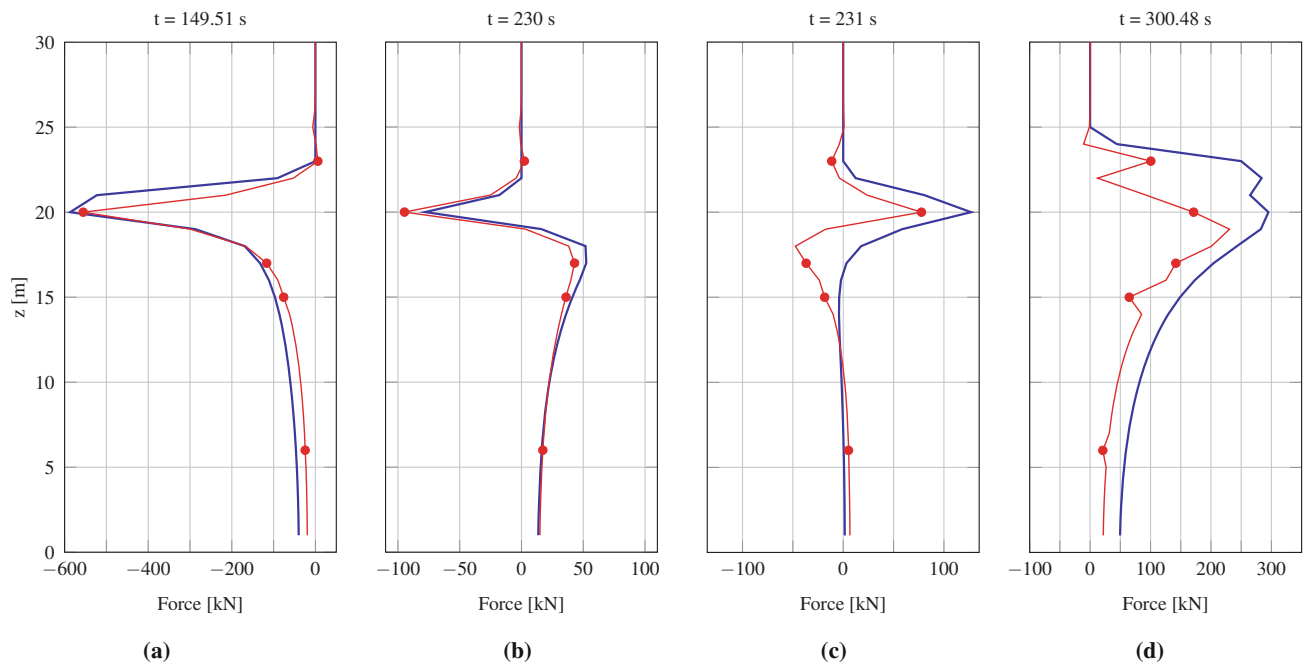
Fig. 21.5 Four snapshots of the estimated force profile (red) along the height of the substructure in comparison with the actual load profile (blue). Red dots indicate the locations of acceleration sensors

family or type of the basis functions, whereas the regression order is always limited by the number of available acceleration measurements.

21.6 Conclusions

This contribution presents the application of a sequential Bayesian inference algorithm for state and distributed input estimation of an offshore wind turbine subjected to wave loads. The algorithm is based on output-only measurements, without the need to directly observe the input process. The distributed input, which in this case consists of the hydrodynamic loads acting on the substructure of the wind turbine, is represented by a Gaussian process regression (GPR) model, which enables the sequential assimilation of measurement data by means of a dual filtering approach. It is observed that the quality of the estimated distributed loads is strongly dependent on the quality of the indirect input observations, which in turn depends on the quality of the underlying numerical model of the wind turbine. Lastly, it is further highlighted that the problem of wave load estimation on offshore wind turbines by means of GPR models calls for more informative basis functions, which should be able to account for the discontinuities of the wave pressure occurring at the mean sea level.

References

1. Azam, S.E., Chatzi, E., Papadimitriou, C., Smyth, A.: Experimental validation of the Kalman-type filters for online and real-time state and input estimation. J. Vib. Control **23**(15), 1–26 (2015)
2. Lei, Y., Xia, D., Erazo, K., Nagarajaiah, S.: A novel unscented Kalman filter for recursive state-input-system identification of nonlinear systems. Mech. Syst. Signal Process. **127**, 120–135 (2019)
3. Kitanidis, P.K.: Unbiased minimum-variance linear state estimation. Automatica **23**(6), 775–778 (1987)
4. Hsieh, C.-S.: Extension of unbiased minimum-variance input and state estimation for systems with unknown inputs. Automatica **45**(9), 2149–2153 (2009)
5. Gillijns, S., De Moor, B.: Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. Automatica **43**(5), 934–937 (2007)
6. Lourens, E., Papadimitriou, C., Gillijns, S., Reynders, E., De Roeck, G., Lombaert, G.: Joint input-response estimation for structural systems based on reduced-order models and vibration data from a limited number of sensors. Mech. Syst. Signal Process. **29**, 310–327 (2012)

7. Lourens, E., Reynders, E., De Roeck, G., Degrande, G., Lombaert, G.: An augmented Kalman filter for force identification in structural dynamics. Mech. Syst. Signal Process. **27**, 446–460 (2012)
8. Azam, S.E., Chatzi, E., Papadimitriou, C.: A dual Kalman filter approach for state estimation via output-only acceleration measurements. Mech. Syst. Signal Process. **60**, 866–886 (2015)
9. Maes, K., Gillijns, S., Lombaert, G.: A smoothing algorithm for joint input-state estimation in structural dynamics. Mech. Syst. Signal Process. **98**, 292–309 (2018)
10. Nayek, R., Chakraborty, S., Narasimhan, S.: A Gaussian process latent force model for joint input-state estimation in linear structural systems. Mech. Syst. Signal Process. **128**, 497–530 (2019)
11. Zou, J., Cicirello, A., Iliopoulos, A., Lourens, E.-M.: Gaussian process latent force models for virtual sensing in a monopile-based offshore wind turbine. In: European Workshop on Structural Health Monitoring. EWSHM 2022 (2022)
12. Rogers, T.J., Worden, K., Cross, E.J.: Bayesian joint input-state estimation for nonlinear systems. Vibration **3**(3), 281–303 (2020)
13. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press, Cambridge (2005)
14. Tatsis, K., Lourens, E.: A comparison of two Kalman-type filters for robust extrapolation of offshore wind turbine support structure response. In: Fifth International Symposium on Life-Cycle Civil Engineering (IALCCE 2016), pp. 209–216, Delft, The Netherlands (2016)
15. Tatsis, K., Dertimanis, V., Abdallah, I., Chatzi, E.: A substructure approach for fatigue assessment on wind turbine support structures using output-only measurements. Procedia Eng. **199**, 1044–1049 (2017)
16. Tatsis, K.E., Dertimanis, V.K., Papadimitriou, C., Lourens, E., Chatzi, E.N.: A general substructure-based framework for input-state estimation using limited output measurements. Mech. Syst. Signal Process. **150**, 107223 (2021)
17. Tatsis, K., Dertimanis, V., Rogers, T., Cross, E., Worden, K., Chatzi, E.: A spatiotemporal dual Kalman filter for the estimation of states and distributed inputs in dynamical systems. In: International Conference on Noise and Vibration Engineering (ISMA 2020) and International Conference on Uncertainty in Structural Dynamics (USD 2020), pp. 3591–3597 (2021)
18. Wikle, C.K., Cressie, N.: A dimension-reduced approach to space-time kalman filtering. Biometrika **86**(4), 815–829 (1999)
19. Wikle, C.K., Zammit-Mangion, A., Cressie, N.: Spatio-Temporal Statistics with R. Taylor & Francis, Milton Park (2019)
20. Jonkman, J., Musial, W.: Offshore Code Comparison Collaboration (OC3) for IEA Task 23 Offshore Wind Technology and Deployment. Technical Report. Technical Report, Colorado, USA (2010)
21. Burton, T., Jenkins, N., Sharpe, D., Bossanyi, E.: Wind Energy Handbook. Wiley, West Sussex (2011)
22. Wheeler, J.: Method for calculating forces produced by irregular waves. J. Petroleum Technol. **22**(03), 359–367 (1970)
23. Morison, J., Johnson, J., Schaaf, S.: The force exerted by surface waves on piles. J. Petroleum Technol. **2**(05), 149–154 (1950)

Chapter 22
Identification of Axial Forces in Structural Rod Members Under Compression by a Modal Approach

Volkmar Zabel and Mena Abdelnour

Abstract The identification of axial forces in structural members by means of identified modal parameters has been the subject of numerous studies in the last decade. In most cases, the estimation of tensile forces is considered. However, only a very few publications also address the identification of compressive axial forces. One problem in modal testing of compression members is the control of the boundary conditions that have also an important influence on the dynamic behavior of the system. In this contribution, an experimental study is presented that shows how the modal behavior of a steel rod in a compression test changes with an increasing axial compression loading.

22.1 Introduction

Relations between axial forces in members of truss structures or tendons and stress-dependent modal parameters have been investigated by several researchers in the past. In most cases, tension rods have been in the focus of the analyses. Solely a few studies include also members under axial compression loading. Furthermore, they usually take only the fundamental bending modes in one direction into account in their experimental studies.

In the investigation presented here, an experimental study on a steel member with an I-section that was loaded in longitudinal direction by an increasing static load is described. The investigations were performed to consider the relations between axial compressive force and natural frequencies of different modes and boundary conditions.

22.2 Description of the Tests

The structure under consideration was a 3-m-long steel beam of type IPE 120. While one end of the beam was fixed, the second support was designed as a sliding bearing to allow for the application of an axial compression load. The experimental setup is shown in Fig. 22.1.

During the static load tests, the axial force was increased incrementally until the I-beam buckled about its weak axis. At five stages, the static load was kept at a constant level for a short time such that respective dynamic tests could be performed to identify the modal parameters. During the dynamic tests, the system was excited by means of an impulse hammer at different locations and in different directions such that the measured signals included well-identifiable contributions from as many modes as possible within the frequency range of interest. The structural response was measured by means of accelerometers at 12 locations in two horizontal directions such that both bending and torsional modes could be identified appropriately.

V. Zabel (✉) · M. Abdelnour

Institute of Structural Mechanics, Faculty of Civil Engineering, Bauhaus-University Weimar, Weimar, Germany

e-mail: volkmar.zabel@uni-weimar.de; mena.abdelnour@uni-weimar.de

Fig. 22.1 Experimental setup in the laboratory
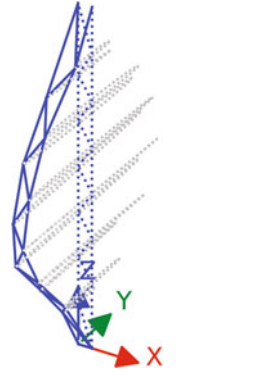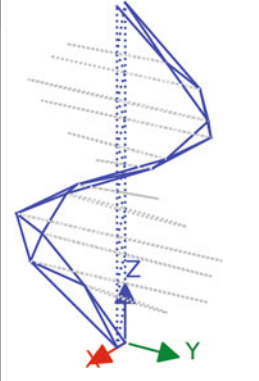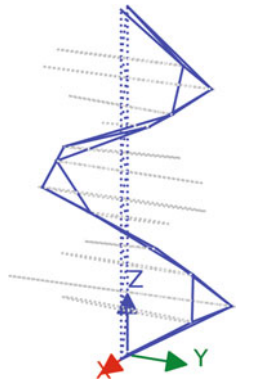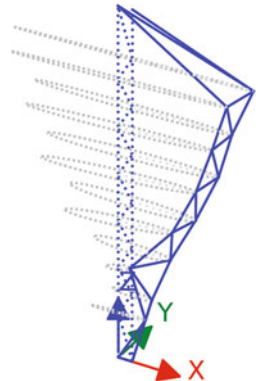
22.3 Modal Identification
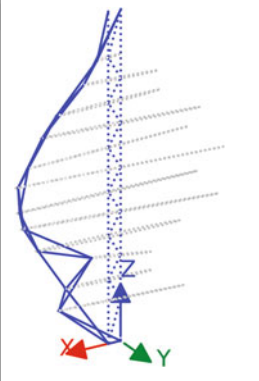
All acquired time series were analyzed by means of the covariance-driven stochastic subspace identification (SSI-COV). It was possible to clearly identify nine modes at all load levels. Table 22.1 summarizes the identified mode shapes and natural frequencies for the different load levels.

From Table 22.1, several conclusions can be drawn. (1) After application of an initial static load, the identified natural frequencies of almost all modes increased. This phenomenon is most probably caused by a cha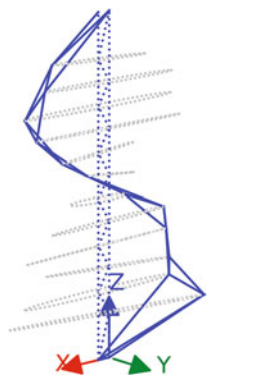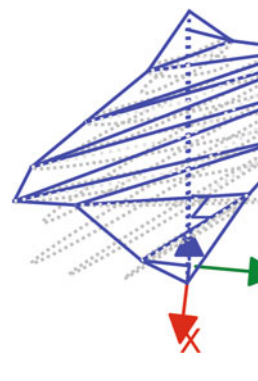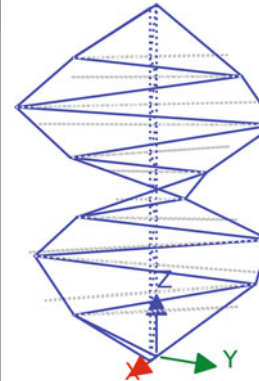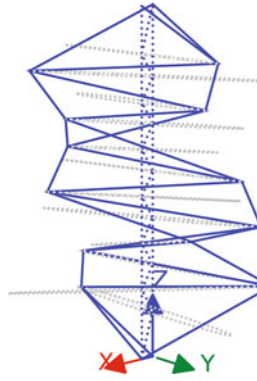nge in the boundary conditions especially at the end of the beam at which the load was applied. (2) Starting from a moderate static load level, there is a general trend of decrease of the identified natural frequencies with increasing static load. This effect can be explained by the geometrically nonlinear behavior of the system. (3) The change of the natural frequencies is most significant for the bending modes about the weak axis. The bending modes about the strong axis are affected the least. Further, the frequency shifts become also less significant with increasing order of the types of modes summarized in one block line in Table 22.1.

22.4 Numerical Analyses

The investigated steel beam was also modeled numerically. In the numerical model, the system has been considered as simply supported so far. That this assumption does not completely agree with the system that has been tested in the laboratory becomes most obvious after a look at the first mode that is characterized by a deformation about the strong axis. This mode shape clearly shows that the horizontal support at the upper end is flexible in this direction. As this flexibility has not been introduced into the numerical model so far, this mode is also not included in the results of the numerical modal analysis.

Table 22.1 Results of the modal identifications

Static load			
Bending about the weak axis			
0 kN	27.65 Hz	74.43 Hz	144.71 Hz
11 kN	28.77 Hz	77.05 Hz	158.53 Hz
50 kN	27.87 Hz	76.73 Hz	145.62 Hz
100 kN	23.31 Hz	74.21 Hz	145.62 Hz
150 kN	21.03 Hz	70.55 Hz	144.66 Hz
200 kN	16.69 Hz	66.96 Hz	142.91 Hz
Bending about the strong axis			
0 kN	31.62 Hz	70.08 Hz	210.33 Hz
11 kN	31.91 Hz	65.43 Hz	220.75 Hz
50 kN	33.68 Hz	65.81 Hz	221.75 Hz
100 kN	33.95 Hz	66.43 Hz	221.75 Hz
150 kN	33.85 Hz	66.15 Hz	222.02 Hz
200 kN	33.72 Hz	65.22 Hz	221.65 Hz
Torsion			
0 kN	53.12 Hz	122.64 Hz	212.85 Hz
11 kN	53.69 Hz	124.25 Hz	214.64 Hz
50 kN	52.71 Hz	122.12 Hz	213.95 Hz
100 kN	50.33 Hz	121.28 Hz	213.95 Hz
150 kN	49.62 Hz	120.95 Hz	211.99 Hz
200 kN	48.32 Hz	115.65 Hz	211.79 Hz

However, despite this discrepancy similar observations as obtained from the experimental study can also be made considering the numerical results. The geometrically nonlinear behavior affects as well primarily the bending modes about the weak axis. With increasing order, the respective modes are also less sensitive to changes in the axial force.

22.5 Conclusions

The observations made in this study might suggest that the identification of the fundamental mode would be absolutely sufficient for a model-based estimate of the axial force as it is most sensitive to the stress state. However, it is rather concluded that further modes should also be included in a model updating procedure that is intended to result in a realistic estimate of an axial force of a considered structural member. This is especially important for the correct identification of the boundary conditions.

Chapter 23
Digital Twin Output Functions and Statistical Performance Metrics for Engineering Dynamic Applications

Matthew S. Bonney and David Wagg

Abstract One of the most common uses of digital twins is to provide information to a user to aid in the decision-making process. The process a digital twin undertakes to generate information can be considered a *digital twin output function*. These can involve predictive simulations, historical trends, and other types of analysis using the data gathered directly from the physical twin and the models contained within the digital twin. Because of this dependency on model simulations and gathered data, the concept of trust is highly relevant to the development of digital twins. To evaluate the trust of a digital twin, specifically the quantitative aspects, the calculation of uncertainty and performance metrics is vital. This chapter considers how performance metrics can be used to compare the output functions of a digital twin to the measured quantities of interest in the physical twin, and thus provide additional information to establish trust in the digital twin that aids in decision-making. This approach will be demonstrated using an engineering dynamics example related to vibration testing.

Keywords Digital twin · Uncertainty quantification · Model updating

23.1 Introduction

Digital twins have gained a recent surge in popularity due to some of the common claims, such as mirroring between physical and digital assets, reduced operational costs, and optimal design [1–4]. While some of these claims have yet to be proven on industrial systems, recent advances in research give credence to some of these claims through proof of concepts [5–9]. In addition to proof of concepts, much work has been done in the implementation of digital twins in various industries such as aerospace [10–12], civil structures [5], and manufacturing [13, 14].

While the definition of digital twin is highly debated, especially when compared to similar technologies such as digital shadows, digital mirrors, and virtualizations [2, 15], the benefits and usability of the twins are highly documented with case studies and commercial products. However, one commonly agreed-upon characteristic of digital twins is the relationship between the physical and digital assets being acceptably identical. The concept of "acceptably identical" is the focus of this research.

The conversation around digital twins being "identical" to the physical twin is highly related to the verification and validation (VV) of the models used within a digital twin. As a digital twin-enabling technology, development and implementation of verification, validation, and uncertainty quantification (VVUQ) are paramount to the development of digital twins. While the field of VVUQ is well studied by a wide variety of researchers, the context of digital twins provides interesting specifications that are not commonly seen in other applications of VVUQ. In particular, the use of data with lifelong interactions will provide a steady stream of data for an evolving system. This chapter introduces the nomenclature of a *digital twin output function*, discusses the various types of data associated with digital twins, and compares data types via statistical performance metrics.

M. S. Bonney (✉) · D. Wagg
Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Sheffield, UK
e-mail: david.wagg@sheffield.ac.uk

23.2 Types of Data

In general, there are two major types of data, experimental and numerical data. Experimental data is taken directly from sensors measuring the physical asset during a period of time. In structural dynamics, this is typically point acceleration, multipoint velocity, or full-field displacement. Generally, there are two types of these sensors, onboard and test sensors. Onboard sensors are attached permanently to the system and are constantly collecting data. For many systems, this data is used for operation or is processed and stored at a reduced frequency. As an example, the sensors on an aircraft wing detect the angle of the flaps to inform the pilots of the current state of the wing. This data is practically never recorded and is only used for dashboard purposes. Test sensors are attached prior to regular testing that occurs and commonly have a large amount of data associated with each test. A ground vibration test for an aircraft is a common example of this, where sensors (either physical or noncontact) are used for the duration of the tests and then removed before regular operation resumes. For digital twins, it is possible to use an assortment of both types of sensors for experimental data.

Numerical data is generated from a singular or collection of models. The term "model," similar to digital twins, has a definition that is under debate, primarily depending on the context and discipline that it is being used for. In mathematics and computer science, any input/output pair is considered a model, while engineering commonly refers to a single calculation/simulation as a model and the combination of models as a process (both of which would be considered models in mathematics). This separation for engineers is useful in the understanding and fault analysis of simulation results.

There is a special subset of numerical data that utilize "black-box" or "gray-box" models. These are in contrast to "white-box" models where the simulation uses a known set of physics-based equations (commonly differential equations) to express the relationship between the parameter input and the model output. Black-box models use numerical processing of data (ideally experimental data) to create nonphysics-based equations for the model. Gray-box models use a combination of physics- and nonphysics-based equations to bridge the knowledge gap between the black- and white-box models. In practicality, white-box models have predictive capabilities but contain epistemic uncertainty due to the lack of knowledge about the physics associated with the system (made by engineering assumptions and unknown physics), while black-box models have low epistemic uncertainty since the experimental data contains the unknown physics but are typically incapable to predict outside the parameter range of the experimental data (commonly an issue in predictive failure since failure data is rare for a system). Gray-box models can bridge this gap by taking the reproducibility of the black-box model in combination with the predictive nature of white-box models with more sophisticated versions able to adjust based on the model needs [16].

23.2.1 Digital Twin Output Functions

In a digital twin for a complex engineering system, it is expected to have every form of data associated with the system and multiple models, including testing/prototyping experimental data, operational sensor data, physics-based models, and data-based/black-box models. To accommodate these multiple sources for generating information, the terminology digital twin output functions (DTOFs) is introduced in this chapter. The idea of a DTOF is similar to a model, where there is an input and output, but instead focuses on a broader class of information rather than just quantitative data. For example, in a deterministic, physics-based model, parameters are inputted and data is produced, whereas a DTOF can be constructed to answer a broader question or information input (such as "what is the current state of the system" or "what is the remaining useful life") and provides contextualized information as the output. This is particularly important for multidisciplinary digital twins where the users range from technicians to engineers to managers/CEOs. There is a discrepancy between the granularity of information required for each user, where engineers require highly granular data and managers usually only require the information for decision-making assuming that there is a high degree of trust in the digital twin [17]. This concept can be better explained using the knowledge hierarchy, seen in Fig. 23.1. In this WIKD hierarchy, information is abstracted from data through processing and refinement. Models take in parameters and data to produce desired data. In contrast, DTOF can use data from models and/or other information to give both quantitative and/or qualitative outputs that can be considered to be at the information and/or knowledge levels of the WIKD model, depending on the exact context of the problem being considered.

For each DTOF, multiple sources of data are used along with models and system-specific contextualization. This chapter focuses on the information input of "How accurate is my model?" in order to make the decision to recalibrate the model. Commonly this would be performed after a major incident (where the regularity of this is well known) or after long-term degradation occurs (where the time is not known). The planned output is a statistical measure of the model accuracy compared
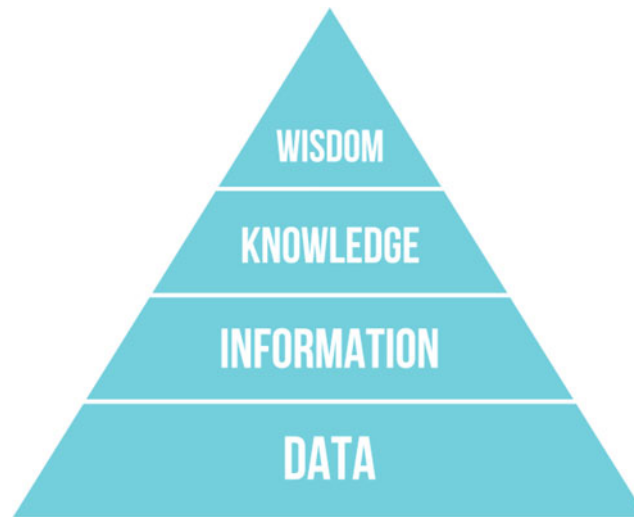
Fig. 23.1 Knowledge hierarchy/DIKW pyramid

to recently collected data. One of the main aspects of this is the comparison between numerical and experimental data that will be discussed in the next section.

23.3 Comparing Experimental and Numerical Data/Metrics

For the comparison of numerical and experimental data, one main distinction is the nature of the data. This mainly comes down to classifying the data as either deterministic or stochastic. Deterministic data is where there is no variation, and a repeat of the model/test will produce identical results. In a general sense, deterministic data can be considered to be an engineering assumption where the variability is small enough to be neglected. The other class of data, stochastic, accounts for the variability in the data. This can be done by a classification of specified distributions, such as Gaussian, and/or by the classification of statistical moments (mean, variance, skewness, etc.). Since the goal is the comparison of two pieces of data, there are four possibilities based on the classification of each.

For the basic comparison, both pieces of data are considered deterministic. This classification results in a comparison plot as seen in Fig. 23.2a and the rest of Fig. 23.2 shows the other possibilities. The comparison of two deterministic values is a simple distance measurement in the response, denoted as ϵ in the figure. Commonly, these are also called ϵ-mirrors due to the main comparison being the Euclidean distance ϵ with other possibilities discussed in [2]. As previously discussed, this scenario is typically caused by engineering assumptions on the data. There are two main assumptions that can create this scenario with the first being negligible variability. This assumption is common for traditional white-box models with deterministic equations of motion. Any variability seen in the repetitions is commonly due to phenomena such as computer bit-resolution and variations in nonexplicit differential equation solvers. These variations are orders of magnitude smaller than the output data and are therefore neglected for a large number of analyses. The other common assumption is a "mean" comparison. This looks at a specific statistical moment, such as mean, and does a comparison between those for each piece of data. Typically, this is used in preliminary analyses to ensure that the model is well validated. One interpretation of this is hypothesis testing of the statistical moments of a model (mainly asking if the experimental and numerical data are from the same distribution).

The second (and third) comparison type is between deterministic and stochastic data, as seen in Fig. 23.2b–c. This is a very common comparison in traditional structural dynamics through the use of uncertainty quantification. The quantifiable measure/metric for this comparison is typically the probability of occurrence. Ideally, the distribution would be a joint distribution for all the output. However, there are many occurrences where multiple distributions are used (as will be in this work) and a normalized probability is used. In this work, the probability is normalized by the mean probability resulting in a metric that tends to unity. While the quantification is the same for both scenarios, the interpretation/question is different. For a stochastic experimental data (either from uncertainty quantification or repeated testing) and deterministic model, the interpretation is, "Is the model within the experimental tolerance of the physical asset?" With a stochastic model (such as
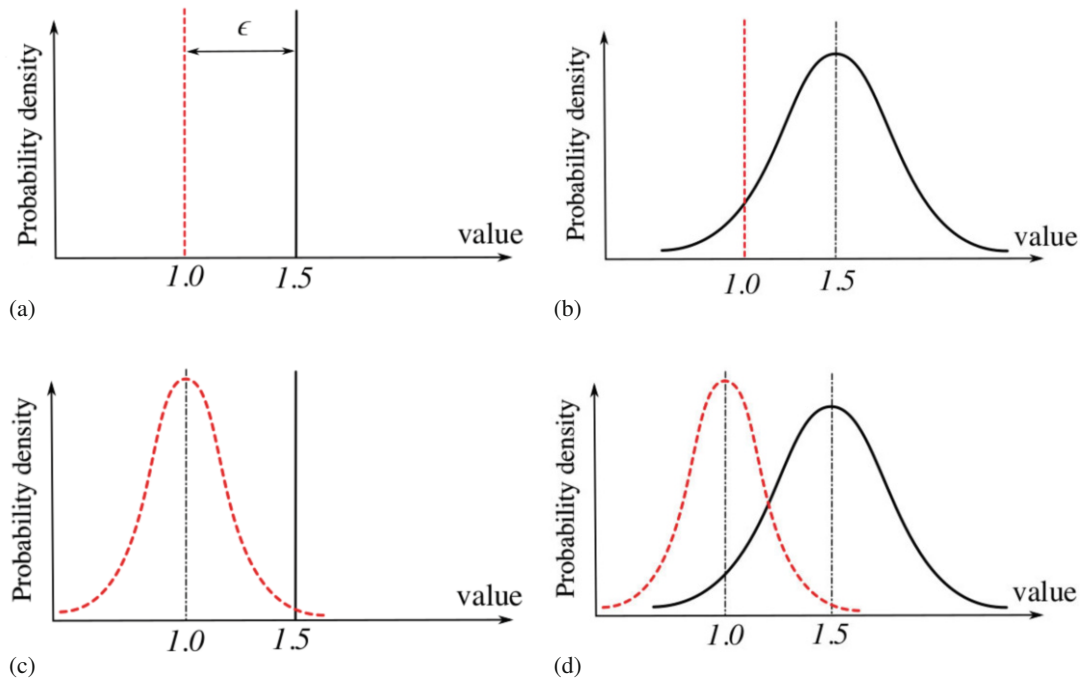
Fig. 23.2 Probability density functions for different types of classifications for comparison. Red dotted line represents experimental data (biased on the left), and black solid line represents numerical data (biased on the right). (**a**) Deterministic–Deterministic. (**b**) Deterministic–Stochastic. (**c**) Stochastic–Deterministic. (**d**) Stochastic–Stochastic

a Gaussian process model or stochastic finite element [18]) and a deterministic experimental data (such as a single-test result), the interpretation becomes, "Is the experimental data a sample from the model distribution?" While the analyses are the same, this difference in interpretation can provide some difficulties in abstraction of information from data. With stochastic experimental data (which is by far the most common and preferred scenario in practice), it is subtly assumed that the experimental data is the truth and model is being tested (e.g., validated), while with assumed deterministic experimental data, it could be interpreted that the model is correct and the sample is being tested. This subtle but significant difference can provide some difficulties in the automation associated with processing the data. While this issue is not discussed in this work, this is a current area of research for the understanding and minimizing of these difficulties.

The final comparison type is between stochastic data, as seen in Fig. 23.2d. This is the most difficult (but potentially informative) comparison due to the complexity of stochastic comparisons. Using a metric for this comparison is an active area of research, and the reader is directed to [19] for a more detailed framework for these metrics. For a brief description, there are two main types of metrics used to make this comparison with the first being *f-divergence* (also called Csiszár's ϕ-divergence) metrics. In general, these metrics take a ratio between set point of the probability using the probability density function (PDF). As the PDFs start to align/overlap, these metrics tend to unity showing a good agreement between distributions. This is similar to the metric used for the stochastic–deterministic comparison of the normalized probability of occurrence. Commonly, a metric will consist of a weighted sum of ratios to ensure that the full distribution is aligned instead of a single point (look at 1.25 in Fig. 23.2d where the ratio at that point is 1). The other main metrics are called interval metrics. These take the differences between the cumulative distribution function (CDF) of the data. Some metrics use singular points while others take an area difference between the CDFs to determine the difference in the distributions.

23.4   Demonstration

To demonstrate an example DTOF, the BAE Hawk T1a starboard wing is taken as the system of interest. Figure 23.3 shows the physical system in Fig. 23.3a and the physics-based model in Fig. 23.3b. The physical system contains 55 accelerometers attached to various points on both the top and bottom surfaces of the wing. The wing is excited using a modal shaker with a stinger attached to the bottom of the wing approximately 70% down the length of the wing away from the fuselage. While
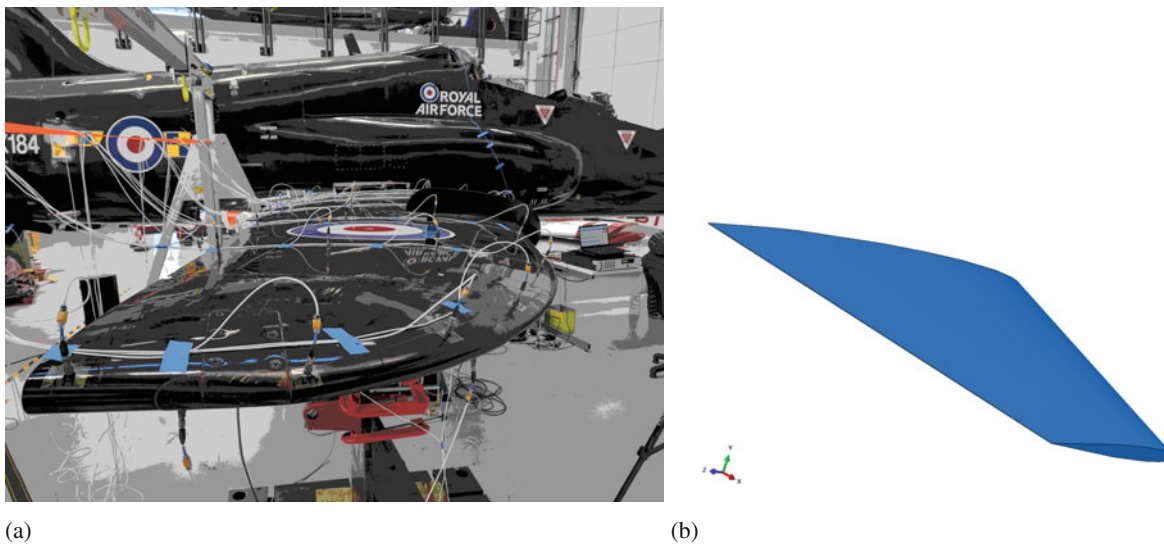
Fig. 23.3 Demonstrator system. (**a**) Physical asset. (**b**) Finite element model

many tests were performed, this section will investigate the forward sine logarithmic swept input. For the excitation level discussed, there were 10 repetitions performed to investigate variability and ensure repeatability. It is planned to make this data publicly available in the future as a possible benchmark for structural health monitoring, nonlinearity detection, and other related techniques.

The finite element model uses the "as-designed" surface geometry and applies shell elements to this surface with material properties of nominal 2024 aluminum. There are some known discrepancies between the model and the physical asset. Firstly, the tip geometry is not identical (as noted by the rounded tip closest to the camera) with the model containing the same cross-sectional shape that tapers down the length of the wing. Secondly, the internal parts of the wing are not modeled, so the model will compensate through the adjustment of shell thickness. Lastly, the boundary conditions are not identical. The model assumes a pinned joint around the connection to the fuselage, while the physical system is still attached and has motion at the connection. However, this motion, while noticeable, is small compared to the accelerometer response and does not greatly affect the quantity of interest. For simplicity, the quantities of interest are the first four bending frequencies of the system. These are matched visually through the mode shapes to ensure that there is no mode switching and to ensure proper treatment of symmetric modes.

23.4.1 Experimental Data

The system excitation investigated in this demonstration is a sine swept in frequency with logarithmic increment. While the excitation is known, an output-only methodology is chosen to better develop the tools for analyzing the aircraft in operation. There are many modes experienced by this system, but the first four resonances are taken as the quantities of interest. These lie under 30 Hz and can be seen in Fig. 23.4. There are four distinct peaks in the range and some minor possible modes due to symmetry. These minor modes are ignored and only the major peaks are under consideration.

To gather quantitative values for the natural frequencies, a circle-fitting routine is applied since each of the major resonances is well separated [17, 20, 21]. The author is aware of some of the difficulties in the circle-fitting routine for symmetric modes, but it is still used for a preliminary analysis to be incorporated into a DTOF. Since the main quantity of interest is the natural frequencies, each test contains 55 accelerometers that allow for multiple estimates per test. In total, there are 550 estimates for each of the natural frequencies due to the 10 repetitions of the test. The statistical moments from this analysis are shown in Table 23.1, and the equivalent histograms are shown in Fig. 23.5. While most of the frequencies match a relatively Gaussian distribution, the third mode is more uncertain. By investigating Fig. 23.5c, it seems to be more similar to a uniform distribution, but it is difficult to identify. Comparing the peaks in Fig. 23.4, this mode has the least defined and most noisy peak. This is also reflected in the standard deviation being much larger than the other modes of interest.
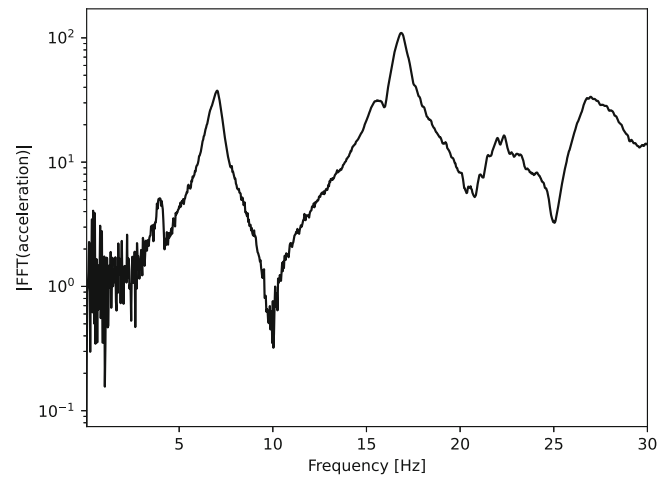
Fig. 23.4 FFT of accelerometer response near the wing tip

Table 23.1 The first two statistical moments of experimental data

Resonance	1st	2nd	3rd	4th
Mean [Hz]	7.0345	16.852	21.832	26.907
Standard deviation [Hz]	0.0333	0.2028	0.5454	0.3095

Table 23.2 Numerical results for natural frequencies of the first four resonances

Resonance	1st	2nd	3rd	4th
Natural frequency [Hz]	9.1285	14.381	18.381	19.783

23.4.2 Numerical Data

Since the wing material is aluminum, 2024-T3 aircraft grade is used for the material. It is uncertain what the exact type of aluminum/heat treatment is, so the nominal values of 2024-T3 are used in this analysis. This material has nominal properties of density of 2.78 g/cc, Young's modulus of 73.1 GPa, and a Poison's ratio of 0.3. The other main parameter of the finite element model is the shell thickness. As an engineering assumption, the thickness is treated as constant across the entirety of the wing and is modeled as 2 mm thick. While the exact thickness is not known, this value is chosen as a common thickness of the sheet metal used in fuselage constructions. It is assumed that the same material is used for the fuselage and wing since the T1a Hawk is a fixed-wing aircraft.

Running the ABAQUS's Lanczos eigenvalue solver gives many modes in this same range of 30 Hz. To uniquely identify the corresponding experimental mode, a comparison of the mode shapes is done visually to ensure mode-to-mode comparison. Future implementation will integrate criteria such as MAC or cross-orthogonality to ensure this accurate comparison. The results from this simulation can be found in Table 23.2.

23.4.3 Comparison

With the current implementation of the DTOF for inspecting the validity of the finite element model, the model is assumed to be deterministic and the experimental data can be either stochastic or deterministic. The first comparison to make is based on a deterministic—deterministic comparison of the experimental mean and numerical result. To provide an equal comparison of modes, the percentage difference of the numerical to experimental mean is averaged together to get the validation metric. This metric is −0.2718, which denotes that the model is roughly 27% out of calibration. This would suggest that there needs to be a model recalibration based on the current state of the system. Typically, a 5% error is the threshold for recalibration, so based on this metric, the digital twin would suggest a model calibration.

A second comparison treats the experimental data as a normal distribution and determines the probability of occurrence. However, as expected from the first metric, there is a large discrepancy between the model and physical system. It can also
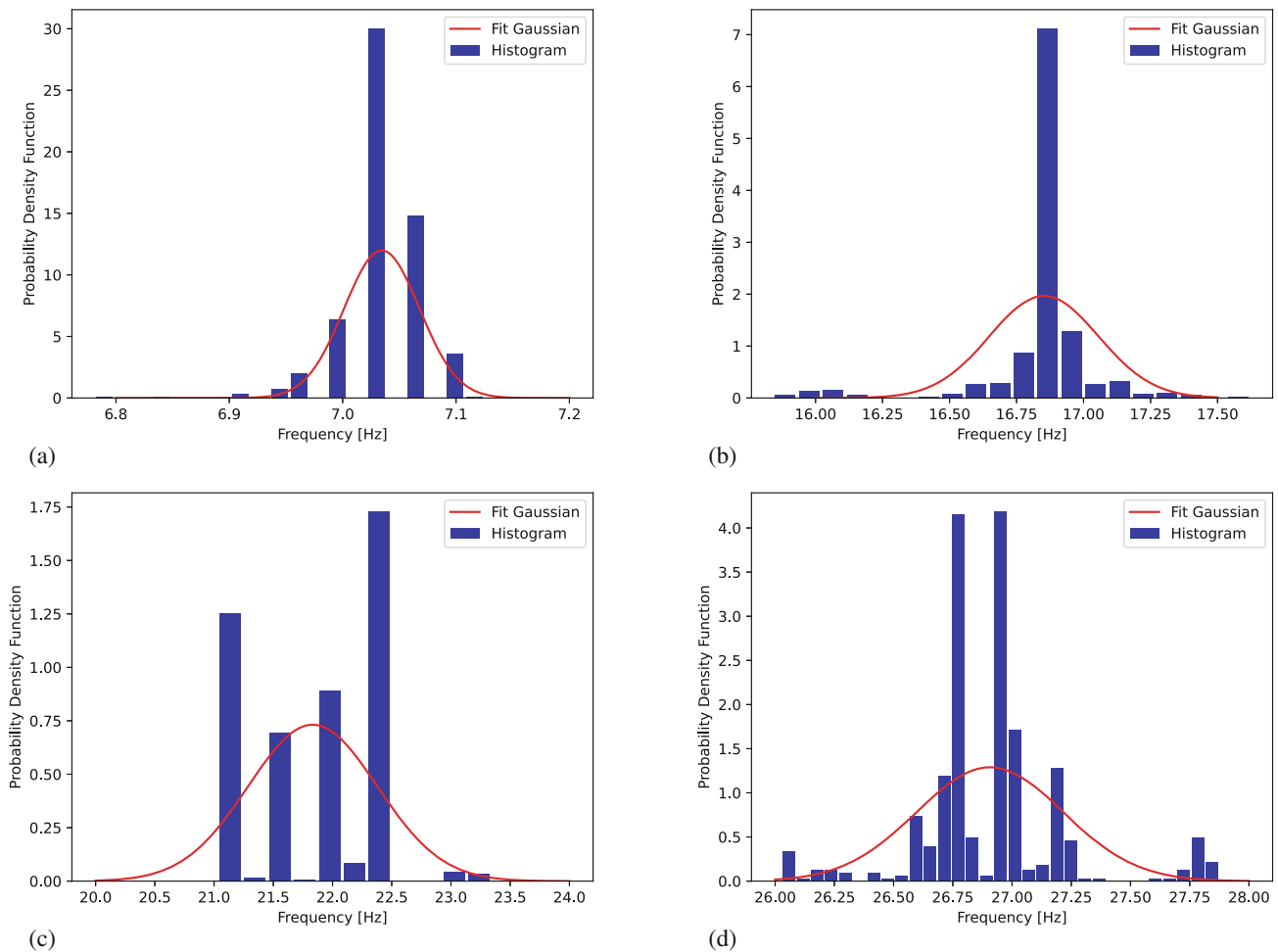
Fig. 23.5 Histogram and fitted Gaussian distribution of natural frequencies. (**a**) 1st mode. (**b**) 2nd mode. (**c**) 3rd mode. (**d**) 4th mode

be inferred from the distributions in Fig. 23.5 that the numerical data will have a very low probability of occurrence. Through the calculations, the normalized probability (taken as the ratio for the probability of occurrence and the mean probability) is averaged to 5×10^{-10}. For this metric, the ideal value would be near unity, so this being orders of magnitude smaller suggests that the likelihood is very small that the numerical simulation estimates a sample from the experimental probability density function. Similar to the deterministic metric, the DTOF would make the suggestion of recalibration.

23.5 Remarks

The digital twin output function framework gives the ability to process data from models and storage to produce information useful for decision-making. Using a question/information input is able to address some of the multidisciplinary aspects of digital twins that are not commonly addressed in research. This ability for multidisciplinary interactions and abstraction of information and knowledge is one of the key aspects of digital twins (at least in dynamical applications) that differentiate it from more traditional modeling.

The T1a Hawk is a highly complex engineering system with many issues in the modeling of the system. Particularly, this model uses the blueprints and is only able to represent the surface of the system. This inability to model has caused a mismatch between the model and physical system. The DTOF is able to quantify this mismatch with the nominal solution and would be able to quantify this for the calibrated system. While this demonstration uses the wing-only tests performed, the framework/technique is flexible enough to allow for full-system tests, stochastic finite elements, and machine learning models such as Gaussian processes.

From the viewpoint of modeling, the major future work is the wing interior modeling and further calibration. While this system is older, the material properties are most likely not the same as originally manufactured or nominal, thus requiring calibration. Additionally, the implementation of a MAC parameter for identifying comparable modes between experiments and model would aid in the automation process. While looking at the digital twin aspect, the DTOF implementation requires automation within the operational platform (such as the platforms in [17, 22]). For this work, all the data (including accessing the database for experimental data and performing finite element simulations) was gathered from the operational platform and computed manually. For the final implementation, it is desired that the user would select the test from a menu and the digital twin will output the statistical metric of comparison.

References

1. Wagg, D.J., Worden, K., Barthorpe, R.J., Gardner, P.: Digital twins: state-of-the-art and future directions for modeling and simulation in engineering dynamics applications. ASCE-ASME J. Risk Uncert. Engrg. Sys. Part B Mech. Engrg. **6**(3), 030901 (2020)
2. Worden, K., Cross, E.J., Barthorpe, R.J., Wagg, D.J., Gardner, P.: On digital twins, mirrors, and virtualizations: frameworks for model verification and validation. ASCE-ASME J. Risk Uncert. Engrg. Sys. Part B Mech. Engrg. **6**(3), 030902 (2020)
3. Grieves, M., Vickers, J.: Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems, pp. 85–113. Springer International Publishing, Cham (2017)
4. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F.: Digital twin-driven product design, manufacturing and service with big data. Int. J. Adv. Manufact. Technol. **94**(9–12), 3563–3576 (2018)
5. Bonney, M.S., Gardner, P., Wagg, D., Mills, R.: Case study of connectivity of digital twins and experimental systems. In: Proceedings of the 8th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, pp. 1416–1425, Streamed from Athens, Greece (2021)
6. Boschert, S., Rosen, R.: Digital Twin—The Simulation Aspect, pp. 59–74. Springer International Publishing, Cham (2016)
7. Kapteyn, M.G., Knezevic, D.J., Huynh, D.B.P., Tran, M., Willcox, K.E.: Data-driven physics-based digital twins via a library of component-based reduced-order models. Int. J. Numer. Methods Eng. **123**(13), 2986–3003 (2022)
8. Lee, J., Lapira, E., Bagheri, B., Kao, H.: Recent advances and trends in predictive manufacturing systems in big data environment. Manufact. Lett. **1**(1), 38–41 (2013)
9. Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., Guo, Z., Lu, S.C.-Y., Nee, A.Y.C.: Digital twin-driven product design framework. Int. J. Prod. Res. **57**(12), 3935–3953 (2019)
10. Li, C., Mahadevan, S., Ling, Y., Choze, S., Wang, L.: Dynamic bayesian network for aircraft wing health monitoring digital twin. AIAA J. **55**(3), 930–941 (2017)
11. Seshadri, B.R., Krishnamurthy, T.: Structural health management of damaged aircraft structures using digital twin concept. In: 25th AIAA/AHS Adaptive Structures Conference, p. 1675 (2017)
12. Tuegel, E.J., Ingraffea, A.R., Eason, T.G., Spottswood, S.M.: Reengineering aircraft structural life prediction using a digital twin. Int. J. Aerosp. Eng. **2011**, Article ID 154798, 14pp. (2011). https://doi.org/10.1155/2011/154798
13. Cerrone, A., Hochhalter, J., Heber, G., Ingraffea, A.: On the effects of modeling as-manufactured geometry: toward digital twin. Int. J. Aerosp. Eng. **2014**, Article ID 439278, 10pp. (2014). https://doi.org/10.1155/2014/439278
14. Rosen, R., von Wichert, G., Lo, G., Bettenhausen, K.D.: About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine, **48**(3), 567–572 (2015). 15th IFAC Symposium onInformation Control Problems in Manufacturing
15. Schwer, L.E.: An overview of the asme v&v-10 guide for verification and validation in computational solid mechanics. In: 20th International Conference on Structural Mechanics in Reactor Technology, pp. 1–10, (2009)
16. Edington, L., Dervilis, N., Abdessalem, A.B., Wagg, D.: A time-evolving digital twin tool for engineering dynamics applications. Mech. Syst. Signal Processing, **188**, 109971 (2023). https://doi.org/10.1016/j.ymssp.2022.109971
17. Bonney, M.S., de Angelis, M., Dal Borgo, M., Wagg, D.J.: Contextualisation of information in digital twin processes. Mech. Syst. Signal Process. **184**, 109657 (2023)
18. Tsialiamanis, G., Wagg, D.J., Dervilis, N., Worden, K.: On generative models as the basis for digital twins. Data-Centric Eng. **2**, e11 (2021)
19. Gardner, P., Lord, C., Barthorpe, R.J.: A unifying framework for probabilistic validation metrics. J. Verif. Validat. Uncert. Quantif. **4**(3), 031005 (2019)
20. Coope, I.D.: Circle fitting by linear and nonlinear least squares. J. Optim. Theory Appl. **76**(2), 381–388 (1993)
21. Kåsa, I.: A circle fitting procedure and its error analysis. IEEE Trans. Instrument. Measur. **IM-25**(1), 8–14 (1976)
22. Bonney, M.S., de Angelis, M., Dal Borgo, M., Andrade, L., Beregi, S., Jamia, N., Wagg, D.J.: Development of a digital twin operational platform using python flask. Data-Centric Eng. **3**, e1 (2022)

Chapter 24
Next-Generation Non-contact Strain-Sensing Method Using Strain-Sensing Smart Skin (S^4) for Static and Dynamic Measurement

Wei Meng, Ashish Pal, Sergei M. Bachilo, R. Bruce Weisman, and Satish Nagarajaiah

Abstract In this study, the next-generation non-contact strain-sensing method with the Strain-Sensing Smart Skin (S^4), based on measuring strain-induced shifts in the emission near-infrared fluorescence spectra of single-walled carbon nanotubes (SWCNTs) embedded in a thin film (S^4) on the surface of the specimen is presented. S^4 is a direct sensing method. In the existing material and structural testing, strain mapping is performed using Digital Image Correlation (DIC), which is an indirect method that compares images of undeformed and deformed configurations to calculate strain maps. The results of the finite element method (FEM) analysis, measured S^4, and DIC strain maps are presented. For static loading, the raster-scanned strain maps generated from S^4 demonstrate more accurate strain maps (especially hotspots and stress/strain concentration areas with steep strain gradients), than DIC, when tested on acrylic specimens. Further, strain maps obtained from finite element analysis matched more closely with S^4 as compared to DIC, demonstrating the superior quality of static strain mapping using S^4. For dynamic loading, the single point S^4 measurement also agreed well with the strain gauge at 3 Hz. The reference-free direct strain-sensing capability of S^4 during static and dynamic loading presented in this study demonstrates the significant potential of the novel method as a promising next-generation strain measurement technology for field applications of structural non-destructive evaluation and structural health monitoring.

Keywords Structural health monitoring · Non-destructive evaluation · Non-contact strain sensing · Single-walled carbon nanotubes · Digital image correlation · Dynamic measurements

24.1 Introduction

Strain measurements relate to the safety of structures and are used in a remarkably wide variety of heavy industries, including aviation, aerospace, and oil and gas. They are needed for testing critical structures of new designs and also for routine inspections of structures exposed to large stress/strains. The most commonly used strain-sensing technology is the resistive foil strain gauge, which can measure at only one location per sensor and needs electrical connections for data acquisition. Other contacted measurement methods, such as fiber Bragg grating (FBG) sensors, have the same limitations [1–3]. As a non-contact strain-sensing method, digital image correlation (DIC) provides full-field displacement and strain maps by comparing the digital images of the specimen surface in the un-deformed and deformed states, respectively [4]. It requires optical features on the object, and the result is sensitive to image quality, ambient light, algorithms, and operators [5–9].

W. Meng · A. Pal (✉) ·
Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA
e-mail: Ashish.Pal@rice.edu

S. M. Bachilo
Department of Chemistry, Rice University, Houston, TX, USA

R. B. Weisman
Department of Chemistry, Rice University, Houston, TX, USA

Department of Material Science and NanoEngineering, Rice University, Houston, TX, USA

S. Nagarajaiah
Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA

Department of Material Science and NanoEngineering, Rice University, Houston, TX, USA

Department of Mechanical Engineering, Rice University, Houston, TX, USA

It has been well studied that the mechanical deformation of the single-walled carbon nanotubes (SWCNTs) can be reflected on the spectral band shift, which is a good indicator for detecting strain [10–13]. Based on this idea, a photoluminescence spectroscopy-based strain-sensing method, "Strain-Sensing Smart Skin (S^4)," is proposed [14]. In this method, the strain-induced change of nanotube electronic structure can be measured as spectral shifts in the SWCNTs near-infrared (NIR) fluorescence emission. Because the nanotube sensors are uniformly distributed across the entire coated surface, strain values at arbitrary locations can be measured and combined to generate the full-field strain [15–19]. We report here the latest progress on the S^4 method for static and dynamic strain measurement.

24.2 Strain-Sensing Smart Skin (S^4)

The improved smart skin has a multi-layer structure to adapt to various materials. An opaque layer is designed at the bottom to block the possible emission from materials such as cement that emits its own fluorescence and interferes with the measurement [20]. Above the blocking layer, an isolation layer made of glossy polyurethane is applied to protect substrate material and provide a microscopically smooth base for the sensing layer, which is a submicron-thick film of SWCNTs in organic polymer poly (9,9-di-n-octylfuorenyl-2,7-diyl) (PFO). The strain of substrate structure induces the electronic structure changes of the embedded nanotubes, which is reflected on the peak shifts of the SWCNTs' NIR fluorescence emission. It is found that those spectral peaks shift consistently with the applied strain. With the pre-calibrated slope of (-1.4 nm/mε), the spectral difference between (7.6) and (7.5) peaks can be used as a reliable monitor for strain measurement shown in Fig. 24.1a. This three-layer structured base coating provides a consistent environment for the SWCNTs. It also lets us paint a speckle pattern onto the blocking layer to allow parallel measurements of DIC and S^4 strain maps on the same specimen. Figure 24.1b illustrates the three-layer coating structure.

24.3 2D Static Strain Mapping

To illustrate the full-field strain mapping by S^4 raster scanning, we used a 1.4 mm thick "I"-shaped acrylic bar with a drilled hole at the center. The central section of the specimen's top surface was coated by S^4 film with DIC speckles on the blocking layer. We attached a conventional foil strain gauge on the back of the specimen close to the hole for reference and DIC sensitivity study. The details of the specimens are shown in Fig. 24.2. The S^4 scanning step size is 0.1 mm in each direction. The strain maps were obtained by one-time scanning in a reference-free manner.

Digital image correlation measurement is also conducted on the specimen as a direct comparison with S^4. Because no out-of-plane deformation is expected in the test, a 2D-DIC system with a single camera (FLIR-Blackfly S Machine Vision Camera) is sufficient for the test purpose. The region of interest (ROI) contained approximately 850 × 850 pixels. The DIC
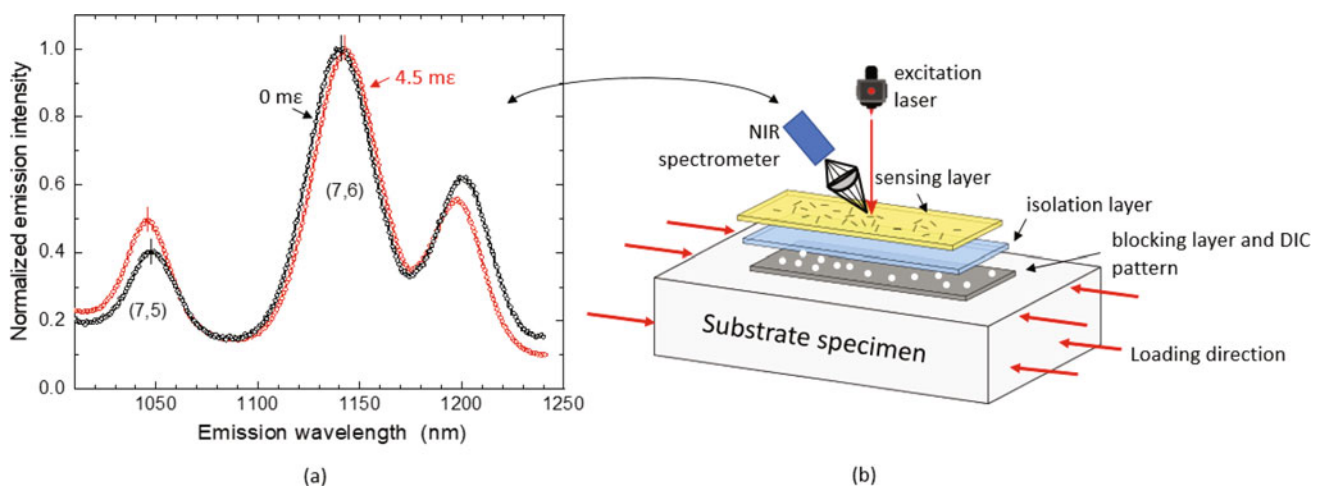

Fig. 24.1 (**a**) Emission spectrum form an S^4 film on a specimen with substrate strain of 0 (*black* points) and 4.5 milli-strain (*red* points); (**b**) layer structured S^4 film. (Adapted from Ref. [19])
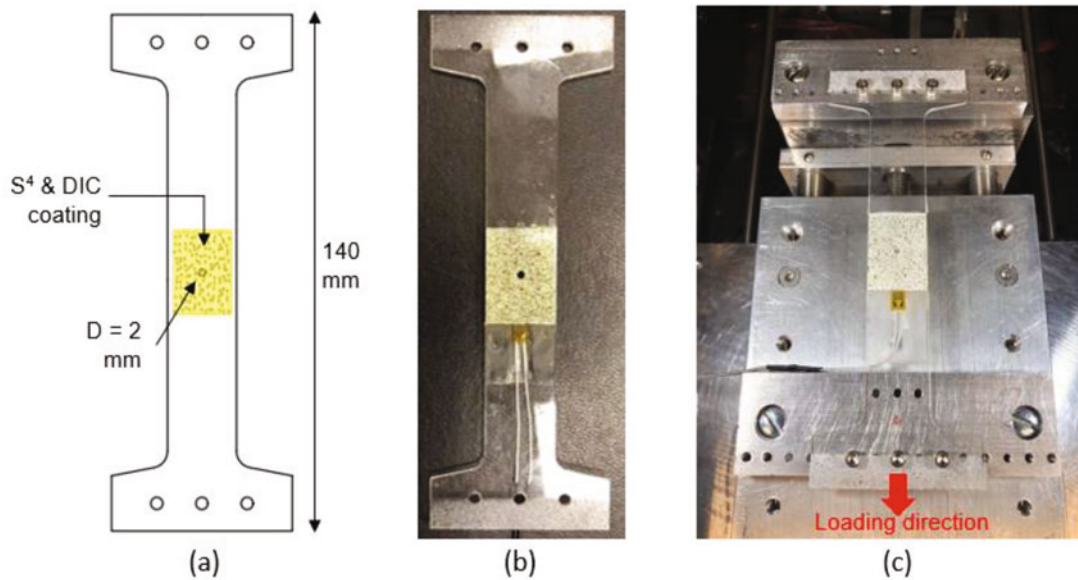
Fig. 24.2 (**a**) Sketch of the acrylic test specimen with a hole at the center; (**b**) photo of the specimen with S^4 coating and DIC speckles; (**c**) photo of the specimen on the loading device
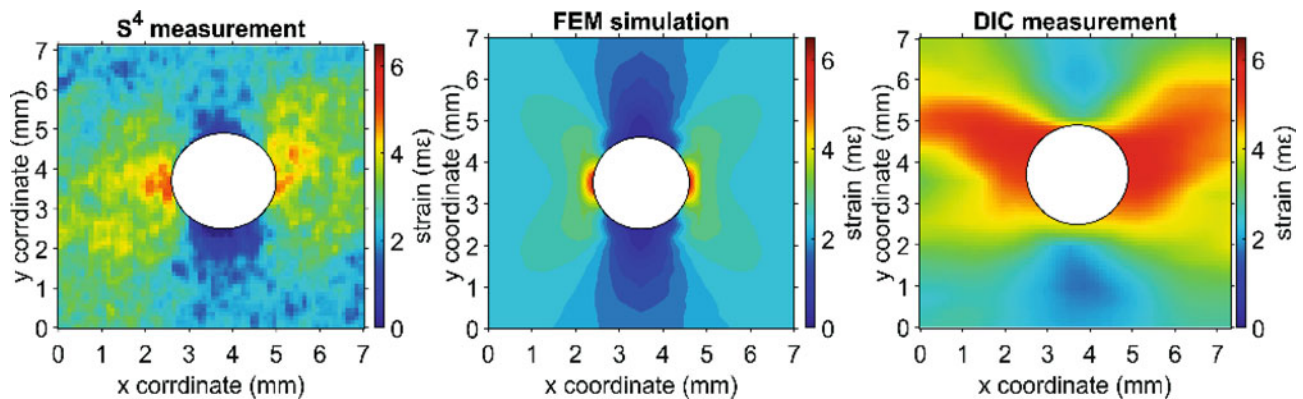

Fig. 24.3 Strain maps of the acrylic specimen with a hole measured by S^4, FEM simulation, and DIC

analysis was done by Digital Image Correlation Engine (DICe) using a subset size of 25 pixels (0.2 mm), and a step size of 12 pixels (0.1 mm).

A finite element model was developed using ANSYS 2021 R1 to compare the experimental strain maps with computational simulation. The material model is built as isotropic elasticity with a Young's modulus of 3.0 GPa and a Poisson ratio of 0.37.

Figure 24.3 displays the strain maps near the hole as obtained by S^4, FEM, and DIC. The strain map measured from S^4 reveals more spatial details and agrees more closely with the FEM simulation. The two strain concentration regions around the hole and strain transitions are clearly revealed in S^4 measurement, which has a high agreement with FEM simulation. By contrast, the strain was heavily averaged resulting in an over-smoothed strain map in DIC, where the two strain concentrations are diffused and the precise locations of them are hard to be identified.

Moreover, an additional reference is not required for post-processing in the S^4 method. However, the DIC results can be significantly influenced by key parameters such as visual strain gauge (VSG) size. A small VSG size has both higher resolution and noise level by using a small averaging area. Conversely, a large VSG size will make the strain map smoother but at the expense of poorer spatial resolution. A decision has to be made by the operators based on the sensitivity study, for which a strain measurement, such as an attached strain gauge, is required as a reference for parameter tuning. However, for an on-site inspection where a contacted reference strain reading is not available, acquiring accurate DIC strain data could be challenging.

As a direct method, S^4 measures strain from the response of many independent sub-micrometer carbon nanotube sensors. The strain measurement does not influence other regions, and there is no trade-off between measurement resolution and spatial resolution. However, DIC is an indirect method, whose accuracy heavily depends on numerical algorithms, image quality, and expert knowledge. DIC requires precisely fixed positioning of the specimen relative to the observation camera. It is therefore impractical to detect strains accumulated in service by moving/flying objects. If the cameras moved, the calibration needs to be conducted again and induced error will be hard to estimate. That means that DIC is not suitable for ground inspections to find damage to aircraft stressed in flight.

24.4 Single Point Dynamic Strain Measurement

To illustrate the dynamic strain sensing of S^4, we used a thin "I"-shaped aluminum bar with the same dimension as the above section except for thickness. The central section of the specimen's top surface was coated by S^4 film. We attached a conventional foil strain gauge at the same position on the back as a reference. The details of the specimens are shown in Fig. 24.4. At this point, we only measured the strain at a single point. The measurement results are compared with strain gauge. The time interval between each S^4 measurement is 0.05 seconds.

The strain measurements with 1, 2, and 3 Hz dynamic loads are shown in Figs. 24.5, 24.6, and 24.7. It can be seen that the dynamic behavior is captured accurately by S^4 measurement both in terms of time history and frequency. In the frequency domain, S^4 captured the same vibration frequency as strain gauge. In the time domain, the strain time history and amplitude measured by S^4 also matched well with strain gauge. The average errors between S^4 and strain gauge are 35, 45, and 66 $\mu\varepsilon$, which indicate a highly accurate strain measurement. These preliminary results indicate the potential of S^4 for accurate dynamic non-contact measurement.

24.5 Conclusion

We reported here a next-generation non-contact strain-sensing technology. The S^4 method has a multilayer structured film design. To check its performance for strain sensing, we performed a static strain mapping and a dynamic single-point strain measurement. The strain mapping is performed by raster scan on a stressed acrylic specimen with a hole at the center to concentrate distinct strain pattern. The resulting strain map was compared with results from DIC, the method currently used
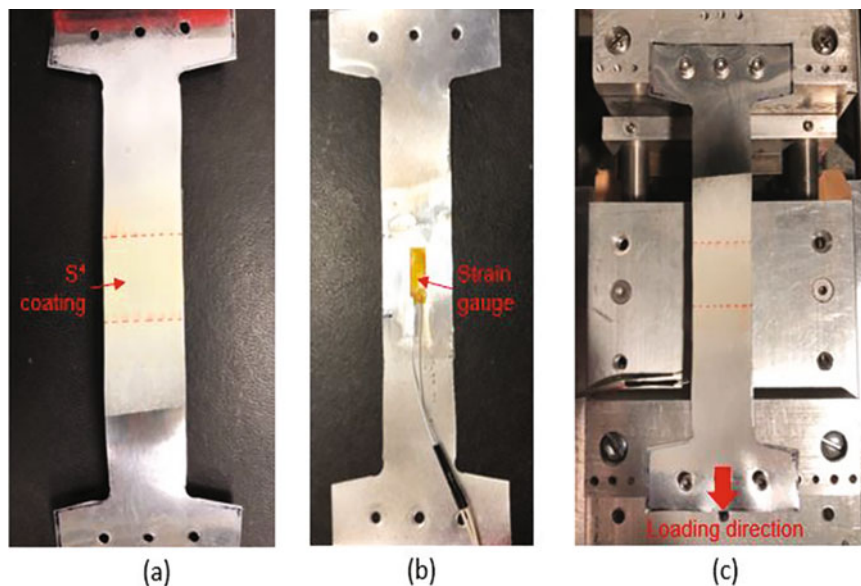

Fig. 24.4 (**a**) Front of the aluminum specimen with S^4 coating; (**b**) back of the aluminum specimen with attached strain gauge; (**c**) photo of the aluminum specimen on the loading device

Fig. 24.5 1 Hz strain measurement by S^4 and strain gauge; upper panel: strain time history; lower panel: fast Fourier transform (FFT) of the measured strain

Fig. 24.6 2 Hz strain measurement by S^4 and strain gauge; upper panel: strain time history; lower panel: FFT of the measured strain

for non-contact full-field strain measurements, and also with an FEM simulation. The S^4 map is more accurate in terms of strain concentration near the hole than the DIC map and is found to be in better agreement with the FEM simulation. In contrast to DIC, S^4 measurement is a reference-free method for which constant observation or precise registration of pre- and post-stress is not required. This feature makes it well suited to many field SHM and non-destructive evaluation applications.
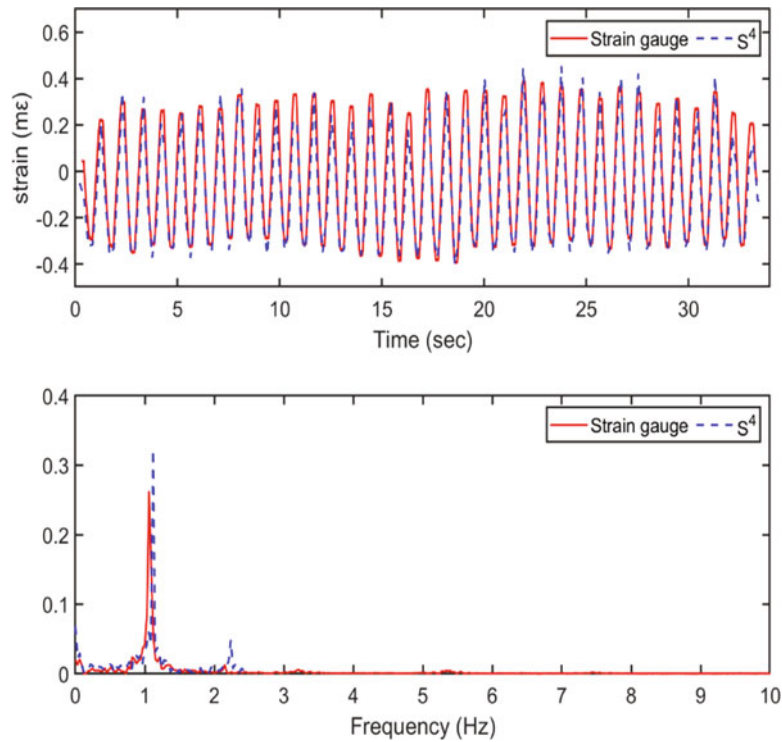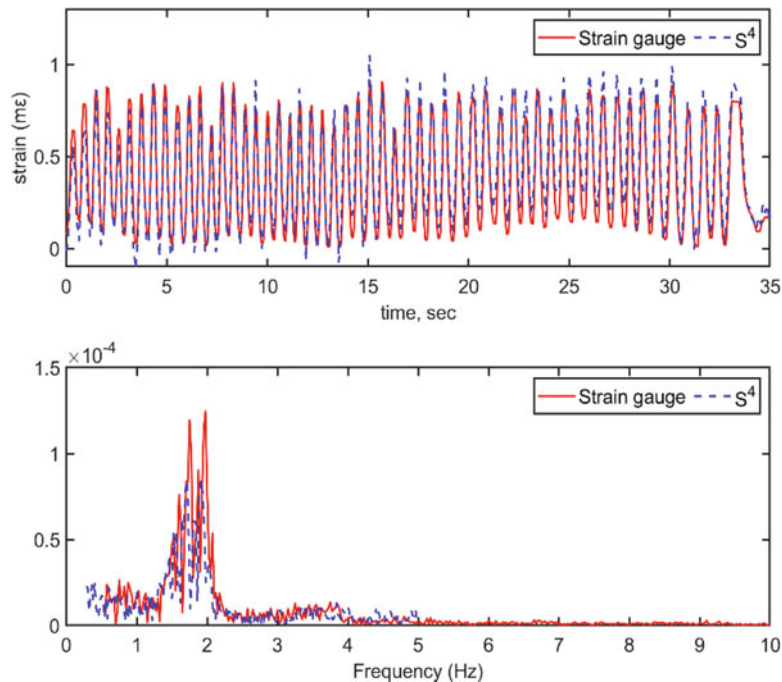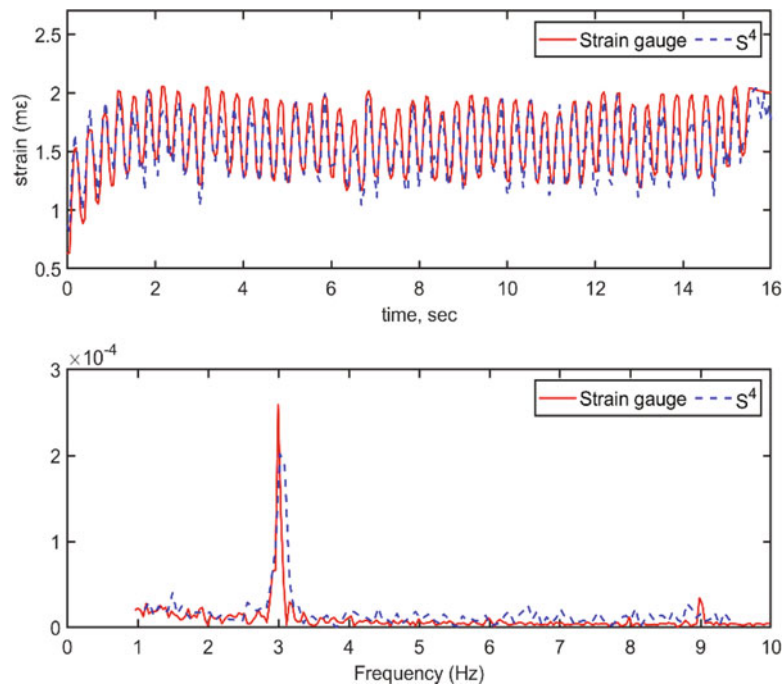
Fig. 24.7 3 Hz strain measurement by S^4 and strain gauge; upper panel: strain time history; lower panel: FFT of the measured strain

Moreover, S^4 is a direct method. There is no trade-off between strain resolution and spatial resolution. The measurement result will not be influenced by operators. The dynamic test is performed on an aluminum specimen with a single-point measurement. S^4 measurement is directly compared with the attached strain gauge. Both frequency and time history are accurately measured by S^4. The average error is less than 50 $\mu\varepsilon$. Taken together, our findings point to the value of S^4 strain measurements as a very promising alternative or complement to existing technologies for non-contact static and dynamic strain measurement. Ongoing research aims to achieve strain map measurements under both static and dynamic loading by using camera-based multi-spectral imaging to replace point-by-point scanning.

References

1. Morey, W.W., Meltz, G., Glenn, W.H.: Fiber optic Bragg grating sensors. In: Fiber Optic and Laser Sensors VII, vol. 1169, pp. 98–107. International Society for Optics and Photonics (1990)
2. Brady, G.P., Kalli, K., Webb, D.J., Jackson, D.A., Zhang, L., Bennion, I.: Recent developments in optical fiber sensing using fiber Bragg gratings. In: Fiber Optic and Laser Sensors XIV, vol. 2839, pp. 8–19. International Society for Optics and Photonics (1996)
3. Marques dos Santos, F.L., Peeters, B.: On the use of strain sensor technologies for strain modal analysis: case studies in aeronautical applications. Rev. Sci. Instrum. **87**(10), 102506 (2016)
4. Pan, B., Qian, K., Xie, H., Asundi, A.: Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. Meas. Sci. Technol. **20**(6), 062001 (2009)
5. Yu, L., Lubineau, G.: Modeling of systematic errors in stereo-digital image correlation due to camera self-heating. Sci. Rep. **9**(1), 1–15 (2019)
6. Lava, P., Cooreman, S., Coppieters, S., De Strycker, M., Debruyne, D.: Assessment of measuring errors in DIC using deformation fields generated by plastic FEA. Opt. Lasers Eng. **47**(7–8), 747–753 (2009)
7. Fayad, S.S., Seidl, D.T., Reu, P.L.: Spatial DIC errors due to pattern-induced bias and grey level discretization. Exp. Mech. **60**(2), 249–263 (2020)
8. Wang, Y., Lava, P., Reu, P., Debruyne, D.: Theoretical analysis on the measurement errors of local 2D DIC: part I temporal and spatial uncertainty quantification of displacement measurements. Strain. **52**(2), 110–128 (2016)
9. Wang, Y., Lava, P., Reu, P., Debruyne, D.: Theoretical analysis on the measurement errors of local 2D DIC: part II assessment of strain errors of the local smoothing method–approaching an answer to the overlap question. Strain. **52**(2), 129–147 (2016)
10. Li, Z., Dharap, P., Nagarajaiah, S., Barrera, E.V., Kim, J.D.: Carbon nanotube film sensors. Adv. Mater. **16**(7), 640–643 (2004)
11. Dharap, P., Li, Z., Nagarajaiah, S., Barrera, E.V.: Nanotube film based on single-wall carbon nanotubes for strain sensing. Nanotechnology. **15**(3), 379–382 (2004)
12. Zhao, Q., Frogley, M.D., Wagner, H.D.: Direction-sensitive strain-mapping with carbon nanotube sensors. Compos. Sci. Technol. **62**(1), 147–150 (2002)

13. De la Vega, A., Kinloch, I.A., Young, R.J., Bauhofer, W., Schulte, K.: Simultaneous global and local strain sensing in SWCNT–epoxy composites by Raman and impedance spectroscopy. Compos. Sci. Technol. **71**(2), 160–166 (2011)
14. Withey, P.A., Vemuru, V.S.M., Bachilo, S.M., Nagarajaiah, S., Weisman, R.B.: Strain paint: noncontact strain measurement using single-walled carbon nanotube composite coatings. Nano Lett. **12**(7), 3497–3500 (2012)
15. Sun, P., Bachilo, S.M., Lin, C.-W., Nagarajaiah, S., Weisman, R.B.: Dual-layer nanotube-based smart skin for enhanced noncontact strain sensing. Struct. Control. Health Monit. **26**(1), e2279 (2019)
16. Sun, P., Bachilo, S.M., Nagarajaiah, S., Weisman, R.B.: Toward practical non-contact optical strain sensing using single-walled carbon nanotubes. ECS J. Solid State Sci. Technol. **5**(8), M3012 (2016)
17. Sun, P., Bachilo, S.M., Weisman, R.B., Nagarajaiah, S.: Carbon nanotubes as non-contact optical strain sensors in smart skins. J. Strain Anal. Eng. Des. **50**(7), 505–512 (2015)
18. Sun, P., Bachilo, S.M., Lin, C.-W., Weisman, R.B., Nagarajaiah, S.: Noncontact strain mapping using laser-induced fluorescence from nanotube-based smart skin. J. Struct. Eng. **145**(1), 04018238 (2019)
19. Meng, W., Pal, A., Bachilo, S.M., Weisman, R.B., Nagarajaiah, S.: Next-generation 2D optical strain mapping with strain-sensing smart skin compared to digital image correlation. Sci. Rep. **12**(1), 1–12 (2022)
20. Meng, W., Bachilo, S.M., Parol, J., Nagarajaiah, S., Weisman, R.B.: Near-infrared photoluminescence of Portland cement. Sci. Rep. **12**(1), 1–6 (2022)

Chapter 25
Online Structural Model Updating for Ship Structures Considering Impact and Fatigue Damage

Jason Smith, Austin R. J. Downey, Ben Grisso, Alysson Mondoro, and Sourav Banerjee

Abstract Naval ship structures (i.e., supports, hull, driving machinery, etc.) have various damage states that develop on short-term (i.e., impact) and long-term (i.e., fatigue) time scales. An up-to-date digital twin of ship structures that can deliver condition assessment in real time would empower a real-time decision-making framework to undertake informed response management. Together, the digital twin and decision-maker will increase ship engagement survivability during combat events and reduce the severity of long-term fatigue effects. A core challenge in digital twin development is the advancement of reliable methodologies that distinguish the short-term and long-term damage states. Furthermore, these methodologies must effectively assimilate large amounts of data into physics-based or data-driven prognostics models while operating on the naval structure's resource-constrained computing environments and considering stringent real-time latency constraints. This work details the experimental validation of a specially designed multievent model updating framework that meets strict real-time latency constraints while operating on a system with limited computational resources. The proposed methodology tracks both impact and fatigue structural damage using a particle swarm that represents numerical models with varying input parameters, given set constraints for latency and computational resources. Experimental validation of the proposed methodology is undertaken using data collected from a structural testbed designed to provide responses representative of a ship subjected to fatigue and impact, considering a predetermined wave loading. Results demonstrate that a physics-based model of the structure can be updated in real time while distinguishing between plastic deformation caused by impact and continuous fatigue crack growth. Latency effects, resource-constrained accuracy, and parameter optimization of the proposed system are quantified and further discussed in this work.

Keywords Model updating · Resource-constrained · Digital twins · Real-time · Parameter optimization

25.1 Introduction

The development of digital twins and the subsequent management of next-generation structures, such as naval structures, will play a critical role in their utilization over a complete life cycle [1]. Additionally, as for the development of digital twins, structural health monitoring (SHM) [2] and real-time model updating compose the majority of the digital twin development. Digital twins without SHM or real-time model updating have hindered capabilities, accuracy, and usefulness. Moreover, if a digital twin lacks adequate real-time model updating capabilities they are unable to respond, assess, or quantify the damage caused by high-rate dynamic events. This prevents the calculations of remaining useful structural health and prognostics that are key components needed for decision-making. Real-time model updating can contain a mixture of physics-based and data-driven models that together allow for the proactive identification of the likelihood of failure, allowing for a better management of the associated logistics tail [3]. For ship structures, there are two main methods to estimate structure life span

J. Smith (✉) · S. Banerjee
Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA

A. R. J. Downey
Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA

Department of Civil Engineering, University of South Carolina, Columbia, SC, USA
e-mail: austindowney@sc.edu

B. Grisso · A. Mondoro
Naval Surface Warfare Center – Carderock Division, West Bethesda, MD, USA
e-mail: benjamin.l.grisso.civ@us.navy.mil; alysson.m.mondoro.civ@us.navy.mil

loads that use monitored data. Each life span estimation is a unique monitoring method with a specific focus. The first focuses on monitoring the environment, while the second focuses on monitoring the structural response [4]. More specifically, the ship's immediate and future surroundings are paired with sea state or wave conditions. For example, monitoring ship routing and using sensing approaches that observe and estimate wave environments in real time (including using wave height radar) can be used to update a life-cycle model of the ship structure [4]. Considering the majority of ship structure faults manifest themselves on varying time scales that consist of damage initiation on a very short time scale (i.e., impact) and damage accumulation on a long time scale (i.e., fatigue, corrosion). The real-time ship structure model updating implemented during combat and impact occurrence would lead to increases in naval system robustness.

Naval ships are equipped with an exceptionally limited amount of computational resources that are continuously allocated to an extensive range of intensive tasks that are highly dependent on the ship's current condition, posture, and mission. Moreover, the computational resources required by structural digital twins may be reallocated to more urgent needs during combat engagements such as radar signal processing, weapon system tracking, and control of power electronics [5]. Considering these limited onboard computational resources, it is necessary to allocate the needed amount of computational resources for the shortest time possible as these limited computational resources are split between multiple intensive tasks at any given time. To update a ship component with multiple models in real time while achieving this constraint, an optimal parameter model-updating algorithmic framework is needed.

Naval ships and structural ship components are active structures that are expected to experience and react to unmodeled high-rate dynamic events in real time. To model active structures in real time, any model updating methodology must use measured data to learn and adapt in real time (under 100 ms) while operating across multiple time scales (impact to life span). Moreover, the real-time structural model updating technique is unable to have a sole dependency on offline training since there are unmodeled damage event combinations that exist. Therefore, the real-time structural updating technique needs to have the ability to learn the structure's state as the unmodeled high-rate dynamic event is happening [6]. This chapter reports preliminary experimental results for the multievent real-time ship structure modeling approach proposed by the author [7]. The proposed real-time multievent framework tracks the Ship Structure and Fatigue Environment (Ship-SAFE) testbed state while it is subjected to an unmodeled dynamic event. Continuing, this proposed framework employs a swarm of particles that function together in parallel to update a linear structural model of the Ship-SAFE testbed using modal analysis. For this work, there are two direct relations to note that involve the particle swarm. The first relation is between the number of particles working together in parallel and the required computational resources, while the second relation is between the system latency and the number of iterations each of the particles is solved over. The major contribution of this work is the inclusion of data obtained from experimental modal analysis data in the proposed real-time multievent model updating framework.

25.2 Background

The proposed multievent model updating methodology is presented Fig. 25.1 while the Ship-SAFE testbed and its damage tracking parameters are depicted in Fig. 25.2. This algorithmic methodology results in an accurately updated finite element analysis (FEA) model of the Ship-SAFE beam by minimizing the error present between the current estimated system state and a sequence of altered structure FEA models. Since the fatigue crack growth occurs near the structure's left fixity and the roller support has a small position change on the right side, the structure's state changes as a result (i.e., damage). Continuing, as an unmodeled dynamic event travels through the structure, the current system state is calculated in real time by selecting the best-fit FEA model that is solved with a particle swarm approach. Moreover, the required computational resources and calculation time are reduced while increasing the robustness of the algorithmic framework as the particle swarm approach only solves a subset of a large number of potential system states rather than using all the computational resources to solve every possible system state.

The real-time algorithmic framework for multievent FEA model updating is presented in Fig. 25.1 and consists of n number of solved FEA models, each with varying independent input boundary conditions and damage state. Continuing, Eq. (25.1) shows that the extracted mode and frequency data is used in the truncated flexibility matrix. Where d_i is a mass normalization constant for the ith mode, $\bar{\phi}_i$ is the mode shape matrix, and ω_i is related to the modal frequencies matrix [8]. Next, $\Delta \mathbf{F}_{\text{trun}}$ is the difference computation between the flexibility matrices of the true (damaged) structure and the trial FEA model. $\mathbf{F}_{\text{turn}}^{\text{true}}$ is the true (damaged) matrix, and $\mathbf{F}_{\text{turn}}^{\text{trial}}$ is the trial FEA model. Lastly, by minimizing $\Delta \mathbf{F}_{\text{trun}}$, the correlating $F_{\text{turn}}^{\text{trial}}$ model is used as the updated model for the structure.
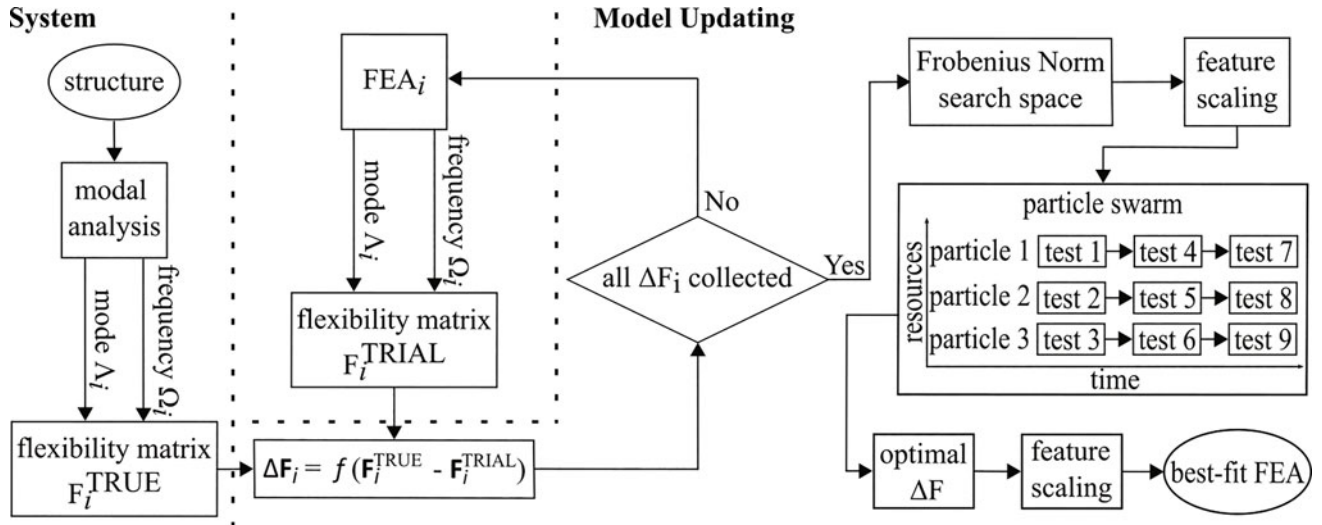
Fig. 25.1 Methodology for the real-time multievent model updating framework used in this work
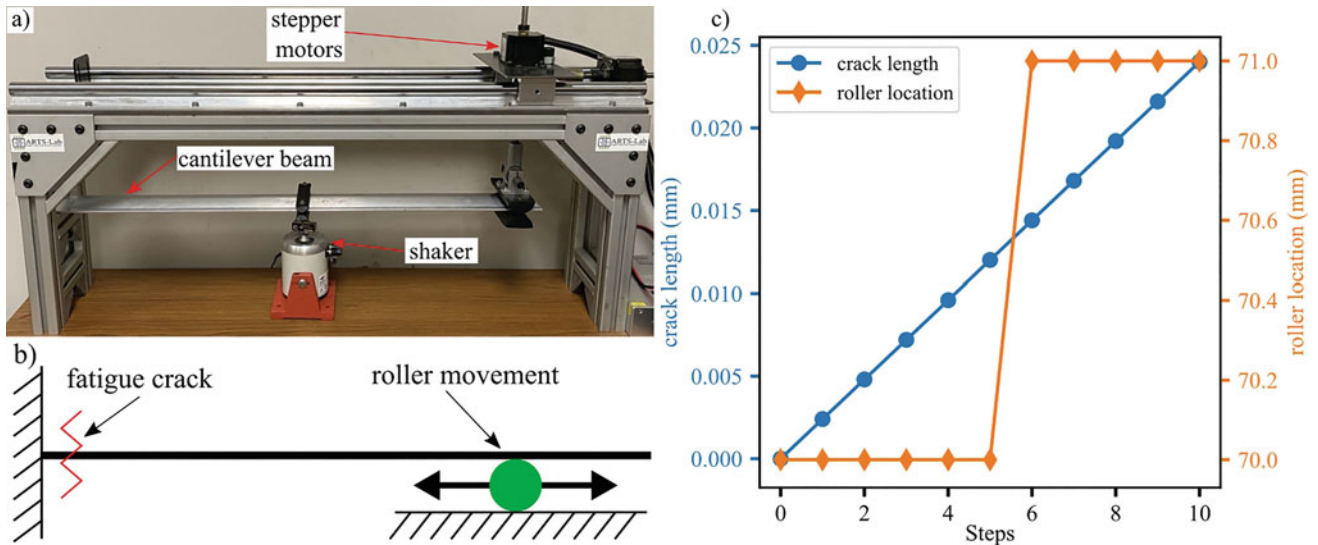

Fig. 25.2 Ship-SAFE testbed used for this work, showing (**a**) physical testbed that consists of a cantilever beam with movable support on the right-hand side and a shaker to represent the structure of interest and (**b**) a 2-D representation of the structure showing the location of the fatigue crack and roller movement

Continuing, Fig. 25.1 shows the algorithmic framework for real-time multievent FEA model updating and consists of n number of constructed FEA models with varying independent boundary conditions and damage cases. Next, Eq. (25.1) shows frequency and mode data are extracted and used in a truncated flexibility matrix.

$$\mathbf{F}_{\text{trun}} = \sum_{i=1}^{n} \left(\frac{d_i}{\omega_i} \right)^2 \bar{\phi}_i \, \bar{\phi}_i^{\text{T}} \tag{25.1}$$

where d_i is a mass normalization constant for the ith mode, $\bar{\phi}_i$ is the mode shape matrix, and ω_i is related to the modal frequencies matrix [8]. $\Delta \mathbf{F}_{\text{trun}}$ is computed as the difference between the flexibility matrices of the true (damaged) structure and the trial FEA model. $\mathbf{F}_{\text{turn}}^{\text{true}}$ is the true (damaged) matrix, and $\mathbf{F}_{\text{turn}}^{\text{trial}}$ is the trial FEA model.

$$\Delta \mathbf{F}_{\text{trun}} = \mathbf{F}_{\text{turn}}^{\text{true}} - F_{\text{turn}}^{\text{trial}} \tag{25.2}$$

Table 25.1 Damage cases for the Ship-SAFE testbed considered in this work

	Roller location (m)	Crack length (m)
Damage case 1	0.700	0.0080
Damage case 2	0.710	0.0100

For this work, $F_{\text{turn}}^{\text{trial}}$ is optimized using a particle swarm by iteratively improving $\Delta \mathbf{F}_{\text{trun}}$ for the FEA models. An important aspect of this project is determining the particle swarms' optimal parameters since naval ship environments have constrained computational resources. If nonoptimal parameters are used, many inefficiencies occur but only the two main issues are discussed. The first is finding the optimal FEA model parameters after an extended time frame (using excessive computational resources), while the second is a returned optimal location that results in higher error (using insufficient computational resources). Continuing, both inefficiencies are very problematic; in either case, a high error solution is returned, or an excessive amount of the naval ship's computational resources are allocated over an extensive time frame. To balance both time constraints and computational resource parameters, the optimal combination is determined and used for this work.

This work used the Ship-SAFE testbed cantilever beam (Fig. 25.2) to represent a ship component. The testbed consists of a large 40×40 extruded aluminum frame that secures a stepper motor that is attached to a fixed aluminum beam. The thin aluminum beam is 76.2 mm wide with a free length consisting of 914.4 mm and a thickness of 1.59 mm. Moreover, Fig. 25.2b shows a 2D representation of the physical testbed with the roller movement along the length of the beam while Fig. 25.2c shows potential steps for a typical crack length and roller location test. For the finite element model, a defined thickness 2D shell element was used for computational efficiency.

The two damage cases that can be considered by the Ship-SAFE are (1) fatigue crack growth and (2) a sudden impact represented by a sudden change in roller location. To expand, consider the temporal tracking parameters reported in Fig. 25.2c. For steps 1–5, the roller location is held at a single location while the fatigue crack length growth is constant; in step 6, the roller location is suddenly changed (representing an impact that happens on a short time scale) while the crack length growth remains at a constant rate. After the impact damage, the roller location remains at the new location while the crack length grows to its final state by step 10. However, in this preliminary work, experimental results for only two damage cases are used. The damage cases are presented in Table 25.1.

25.3 Analysis

The first five numerical mode shapes of the Ship-SAFE cantilever beam are presented in Fig. 25.3. The specific modes of interest for this work are the vertical bending (Bending-Z) modes, which are modes 2, 4, and 5 from Fig. 25.3, since the input conditions act in the vertical bending direction. Continuing, the vertical bending modes were experimentally validated by placing accelerometers on the beam in Fig. 25.2a. The vertical mode shape comparison between the numerical and experimental modes is shown in Fig. 25.4. Continuing, Fig. 25.4 is composed of the following: (1) scaled FEA modes, (2) accelerometer measuring points, (3) interpolation points, and (4) 1-D fit through measured and interpolation points. Together, these aspects provide a visual mode comparison that is later mathematically evaluated using two separate methods. To initially determine the optimal curve fitting method, many fitting methods were explored such as linear, 3–6 degree polynomial, log, 1-D fit, sine, and cosine. For this first test only, the measured points were used for each fitting method; this test resulted in the 1-D fit as the most optimal fitting method. Next, each of the methods was tested again with an interpolation point on various locations and compared to each other, with the best results coming from the 1-D fit again. Lastly, the results from each 1-D fit method were compared, with the best fit resulting from the 1-D fit with an interpolation point. It is important to note that the misplaced nodes for the FEA modes are caused by the fatigue crack in the model.

Since the optimal fit method is determined, the numerical and experimental modes can be compared. Starting with mode 1, there is a good fit between the experimentally measured and numerical mode with only a small shift present. Moving to mode 2, the fit between each is better than in mode 1 but a small shift is still present. Finally, in mode 3 the fit comparison between each method is less than that in the previous modes; this is caused by a larger shift and larger difference for the first peak. To quantify and show the comparisons between the numerical and experimental modes, two methods were chosen. The first method is a modal assurance criterion (MAC) plot, which provides a decent degree of mode shape consistency. The second method is an orthogonality check, which determines whether a mode can be constructed by a linear combination of other modes. Both methods were used to evaluate the first three vertical bending modes and are shown in Fig. 25.5. Here, the MAC plot shows a good correlation between modes 1 and 2 with a lesser correlation for mode 3, while the orthogonality plot shows a high correlation for modes 1–3 with a small correlation between experimental mode 3 and numerical mode 2. This small correlation is likely apparent since both are bending modes and are similar in shape.

Fig. 25.3 First five mode shapes of the FEA model for the Ship-SAFE testbed
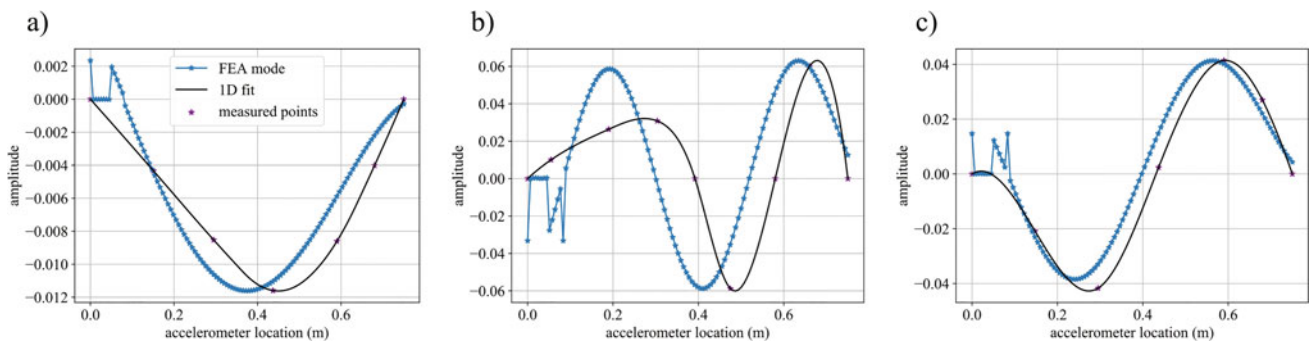

Fig. 25.4 Experimental and numerical bending mode comparisons for (**a**) mode 1, (**b**) mode 2, and (**c**) mode 3, each showing a small horizontal shift

The MAC and orthogonality plots both have a strong diagonal, which indicates a good correlation between like modes from the numerical and experimental methods while having weaker off-diagonal values indicating less correlation between unlike modes (i.e., modes 1 and 2).

The experimental Frobenius Norm search space surface plots are presented in Fig. 25.6. Where Fig. 25.6a is damage case 1 and Fig. 25.6b is damage case 2, as detailed in Table 25.1. The global minimum coordinate for each damage state is of interest as it corresponds to the true state of the structure that produces the smallest calculated error. Moreover, the global minimum of the experimental Frobenius Norm search space is found through a particle swarm implementation that uses random particle starting locations in the search space [7]. This search method results in a global minimum that is quickly, reliably, and efficiently determined by optimizing the particle swarm parameters. Of the tested particle-iteration parameter combinations, the optimal combination found in the author's prior work to 10 and 25 for a number of particles and number of particle iterations respectively was reused in this work [7]. Results obtained are reported in Table 25.2. These results show that the proposed real-time multievent model updating framework is capable of tracking multiple error types in the considered experimental structure.
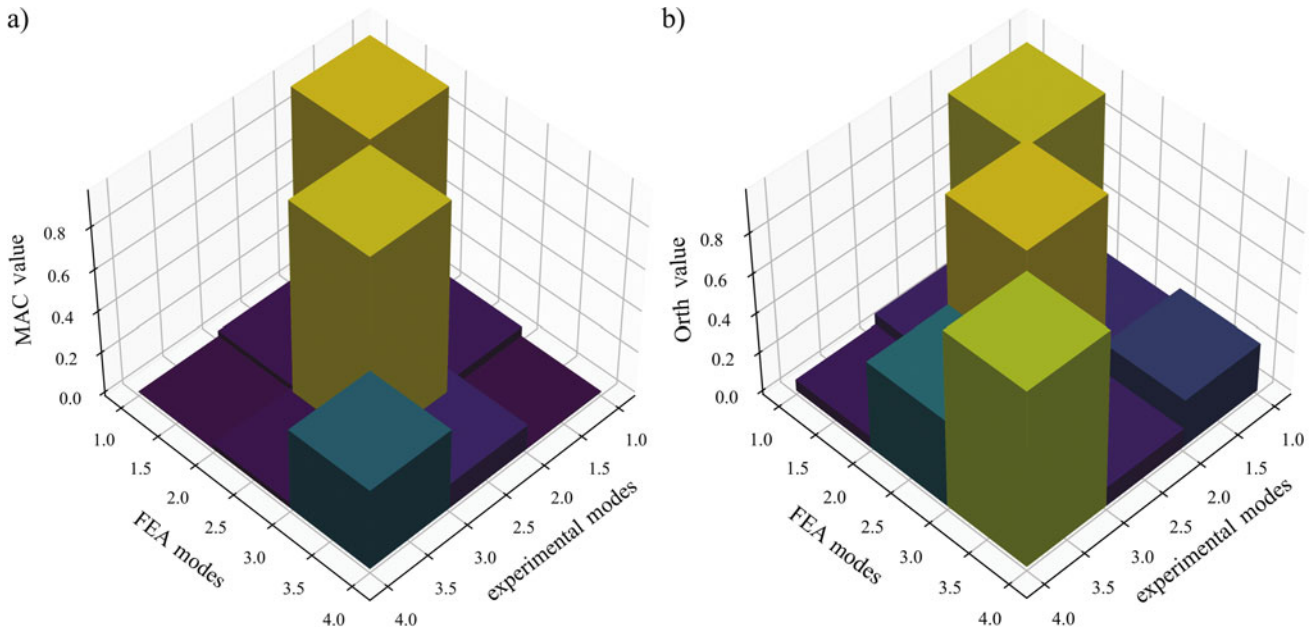
Fig. 25.5 Experimental and numerical mode shape comparison results, showing (**a**) modal assurance criterion (MAC) plot and (**b**) orthogonality plot
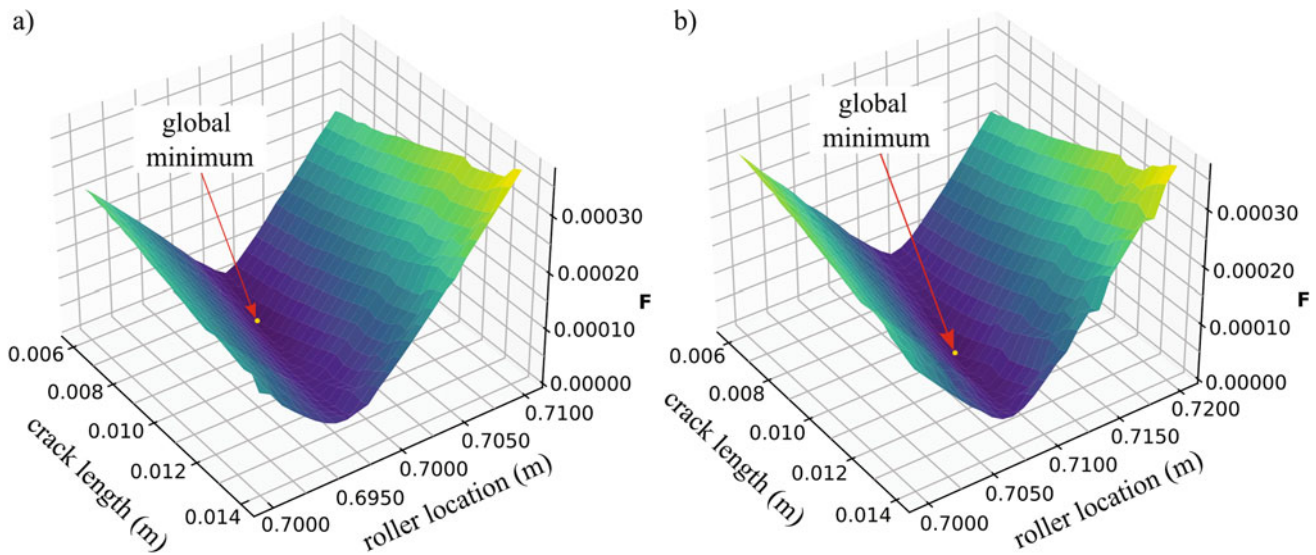
Fig. 25.6 Experimental Frobenius Norm search space surface plot showing (**a**) damage case 1 and (**b**) damage case 2

Table 25.2 Results for the considered damage cases

	Ground truth (m)		Estimated (m)		Error (%)		
	Roller location	Crack length	Roller location	Crack length	Roller location	Crack length	F
Damage case 1	0.700	0.0080	0.700	0.0076	0	5.26	8.60E–06
Damage case 2	0.710	0.0100	0.71	0.0103	0	−2.91	7.20E–06

25.4 Conclusion

There is an essential need for the development of a real-time model updating methodology for ship structures that considers impact damage and fatigue damage without neglecting the ships' computational resources. This work reports the initial development of a multievent model updating framework for ship structures that examine impact and fatigue damage. To

provide repeatable structure conditions and evaluate the real-time multievent model updating framework for naval ship structures, a cantilever beam testbed was developed. The experimental Frobenius Norm search space surface plots contained two damage inputs. The first considered damage case is a steady slow-growing fatigue crack and the second is sudden impact damage that results in a boundary condition change. This boundary condition change caused by the impact is represented as a sudden roller connection movement on the cantilever beam. The convex experimental Frobenius Norm search space developed by the two damage events was solved using a particle swarm optimizer with experimentally determined optimal parameters. Two damage steps were investigated and the particle swarm optimizer returned near-global minimum coordinates in the Frobenius Norm search space. With near-global minimum coordinates correctly returned for both steps, the optimal parameters for the structure are known. Thus, the structural model could be updated with correct values for fatigue crack length and changes in connections (i.e., roller location). The preliminary work undertaken in this work resulted in an updated structural model with a near-global minimum returned by a particle swarm optimizer.

References

1. Okasha, N.M., Frangopol, D.M.: Integration of structural health monitoring in a system performance based life-cycle bridge management framework. Struct. Infrastruct. Eng. **8**, 1–18 (2012)
2. Worden, K., Farrar, C. R., Manson, G., Park, G.: The fundamental axioms of structural health monitoring. Proc. R. Soc. A Math. Phys. Eng. Sci. **463**(2082), 1639–1664 (2007)
3. Drazen, D., Mondoro, A., Grisso, B.: Use of digital twins to enhance operational awareness and guidance. In: Proc. Conf. Comput. IT Appl. Marit. Ind, vol. 18, pp. 344–351 (2019)
4. Mondoro, A., Grisso, B.: On the integration of SHM and digital twin for the fatigue assessment of naval surface ships. In: Structural Health Monitoring 2019. DEStech Publications, Lancaster (2019)
5. Schoder, K., Stanovich, M., Vu, T., Vahedi, H., Edrington, C., Steurer, M., Ginn, H., Benigni, A., Nwankpa, C., Miu, K., Ferrese, F.: Evaluation framework for power and energy management shipboard distribution controls. In: 2017 IEEE Electric Ship Technologies Symposium (ESTS). IEEE, Piscataway (2017)
6. Downey, A., Hong, J., Dodson, J., Carroll, M., Scheppegrell, J. : Millisecond model updating for structures experiencing unmodeled high-rate dynamic events. Mech. Syst. Signal Process. **138**, 106551 (2020)
7. Smith, J., Huang, H.T., Downey Jr, A., Mondoro, A., Grisso, B., Banerjee, S.: Multi-event model updating for ship structures with resource-constrained computing. In: Zonta, D., Su, Z., Glisic, B. (eds.), Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2022. SPIE, Bellingham (2022)
8. Kurata, M., Kim, J.H., Lynch, J.P., Law, K.H., Salvino, L.W.: A probabilistic model updating algorithm for fatigue damage detection in aluminum hull structures. In: ASME 2010 Conference on Smart Materials, Adaptive Structures and Intelligent Systems, Volume 2. ASMEDC, Houston (2010)

Chapter 26
Detuning Optimization of Nonlinear Mistuned Bladed Disks Using a Probabilistic Learning Tool

Evangéline Capiez-Lernout and Christian Soize

Abstract This chapter deals with the detuning optimization of a mistuned bladed disk in the presence of geometrical nonlinearities. A full data basis is constructed by using a finite element model of a bladed disk with cyclic order 12, which allows all the possible detuning configurations to be computed. It is then proposed to reformulate the combinatorial optimization problem in a probabilistic framework using and adapting the recent probabilistic learning on manifolds (PLoM) tool to the detuning context. The available full data basis is used in order to validate the proposed method.

Keywords Detuning/mistuning · Geometrical nonlinearities · Probabilistic learning on manifold · PLoM · Uncertainty quantification · Stochastic optimization

26.1 Introduction

This research concerns the improvement of the vibratory performances of turbomachines by using a detuning optimization strategy that allows for reducing the amplifications induced by the unavoidable random blade mistuning of bladed disks. In a green aviation context for which fan blade design yields larger blades made up of lighter materials, nonlinear geometrical effects are taken into account. The detuning is described by using alternating patterns of several different sector types. A full data basis is constructed by using a finite element model of a bladed disk with cyclic order 12 [1, 2], which allows the random responses of all the possible detuning configurations to be identified [3]. Such a detuning optimization requires to solve a high-dimensional combinatorial optimization problem for which the cost function is evaluated from a nonlinear stochastic reduced computational model [high-fidelity computational model (HFCM)], which has previously been constructed [3, 4]. In practical situations, only a small data training set, issued from the HFCM and which does not a priori include any optima, is available. The main idea is then to construct a continuous approximation of this cost function based on the use of the probabilistic learning on manifolds (PLoM) [5, 6], and that is used in the learning step. Several difficulties inherent to the definition of the cost function require reformulating the definition of the optimum. The available full data basis is then used to validate the proposed methodology.

26.2 Background

The bladed disk is assumed to have n_w blades with two types of blades (labeled by integer 0 and 1). A detuning configuration ℓ is then parameterized by a vector $\mathbf{w}^{c,\ell} = (w_1^{c,\ell}, \ldots, w_{n_w}^{c,\ell})$, in which for $k \in \{1, \ldots, n_w\}$, w_k^c is equal to 0 or to 1. In the frame of the mistuning, the resonance of the most responding blade is defined by the $\mathbb{R}^+$-valued random variable A^ℓ, in which upperscript ℓ corresponds to the detuned configuration number ℓ and whose realization θ_k is denoted by $a^{\ell,k} = A^\ell(\theta_k)$. In order to get a robust scalar quantity for characterizing the random nonlinear dynamical behavior of the detuned structure, an estimate of the maximum extreme value statistics of random variable A^ℓ is constructed. The number of Monte Carlo numerical simulations is written as $n_{\text{sim}} = \nu_r \, \nu_e$ (for $n_{\text{sim}} = 500$, $\nu_e = 10$ and $\nu_r = 50$). For $\ell \in \{1, \ldots, \nu_e\}$, we then define the quantity $\underline{a}_M^\ell$, such that

E. Capiez-Lernout (✉) · C. Soize
Université Gustave Eiffel, Laboratoire MSME UMR 8208 CNRS, Marne La Vallée, France
e-mail: evangeline.capiez-lernout@univ-eiffel.fr; christian.soize@univ-eiffel.fr

$$\underline{a}_M^\ell = \frac{1}{v_r} \sum_{r=1}^{v_r} a_M^{\ell,r} \quad \text{with} \quad a_M^{\ell,r} = \max_{k \in \{v_e(r-1)+1,...,rv_e\}} a^{\ell,k} \quad r \in \{1,\ldots,v_r\}.$$

It should be noted that such quantity of interest is neither issued from a mean value or from an extreme value but is defined as an average of a set of $v_r = 50$ maxima taken in a subset of $v_e = 10$ realizations. The observation of the detuned ℓ-configuration with mistuning is then defined as the amplification factor $q^{c,\ell}$ with respect to its tuned counterpart with pure mistuning, which is written as $q^{c,\ell} = \underline{a}_M^\ell / \underline{a}_M^t$, in which superscript t is related to the tuned configuration. It is thus interpreted as a highly nonlinear function of $\mathbf{w}^{c,\ell}$ that is to say $q^{c,\ell} = f_{\text{HFCM}}(\mathbf{w}^{c,\ell})$. An available data basis [3] is constructed using the finite element model of the blisk described in [1], yielding 352 detuning configurations that are restricted to the set $\mathcal{C}_c \subset \mathcal{N}_c$ of the $n_c = 216$ detuning configurations having a majority of blades with type 0.

The detuning optimization consists of solving the combinatorial optimization problem such that optimum $\mathbf{w}_{\text{un}}^{c,\text{opt}}$ is defined by

$$\mathbf{w}_{\text{un}}^{c,\text{opt}} = \arg \min_{\mathbf{w}^c \in \mathcal{C}_c} \mathcal{J}_c(\mathbf{w}^c) \text{ with } \mathcal{J}_c(\mathbf{w}^c) = f_{\text{HFCM}}(\mathbf{w}^c) \quad , \quad \mathbf{w}^c \in \mathcal{C}_c. \tag{26.1}$$

Such an optimization problem cannot in general be performed, the number of possible detuning configurations increasing exponentially with the number of blades. Let $\mathcal{D}_d \subset \mathcal{C}_c \subset \mathcal{N}_c$ be an N_d-dimensional small training set with $N_d \ll n_c$ that does not a priori contains any improving detuning configurations and for which $\mathcal{J}_c(\mathbf{w}^c)$ is computed with HFCM for all $\mathbf{w}^c \in \mathcal{D}_d$. A new approximation of the optimization problem is formulated, yielding the optimum $\mathbf{w}^{c,\text{opt}}$ to be defined by

$$\mathbf{w}^{c,\text{opt}} = \arg \min_{\mathbf{w}^c \in \mathcal{C}_c} J_{\text{ar}}(\mathbf{w}^c) \text{ with } J_{\text{ar}}(\mathbf{w}^c) = \mathcal{E}\{Q|\mathbf{W} = \mathbf{w}^c\} \quad , \quad \mathbf{w}^c \in \mathcal{C}_c, \tag{26.2}$$

where $(Q, \mathbf{W})$ is a $\mathbb{R}^{1+n_w}$-valued random variable whose joint probability density function is constructed from the training data set using the modified multidensity Gaussian KDE [7] by adjusting a bandwidth parameter to get for each marginal density function of W_i a bimodal probability density function centered around 0 and 1. The authors refer to [5–7] for the theory and the algorithm of the probabilistic learning on manifolds (PLoM). This machine learning tool allows for generating a large number N_{ar} of learned realizations, denoted by $(q_{\text{ar}}^k, \mathbf{w}_{\text{ar}}^k)$, $k \in \{1, \cdots, N_{\text{ar}}\}$, while preserving the concentration of the learned probability measure on the manifold defined by the graph $(q(\mathbf{w}), w \in \mathbb{R}^{n_w})$.

Due to the numerous local minima and weak contrast of highly nonlinear cost function $J_{\text{ar}}(\mathbf{w})$, the formulation of this optimization problem must be improved, leading us to reformulated the detuning optimization problem as follows. By fixing a parameter n_s defining the dimension of the set $\mathcal{W}_{n_s}^{\text{opt}} = \{\mathbf{w}^{c,\ell_1}, \ldots, \mathbf{w}^{c,\ell_{n_s}}\} \subset \mathcal{C}_c$ and by sorting $J_{\text{ar}}(\mathbf{w}^{c,\ell_1})$ for $j = 1, \ldots, n_s$, according to its first n_s increasing values, the optimal detuning configuration is then defined by

$$\mathbf{w}^{c,\ell_{\text{opt}}} = \arg \min_{\mathbf{w}^{c,\ell_j}, \, j=1,\ldots,n_s} \{\mathcal{J}_c(\mathbf{w}^{c,\ell_1}), \ldots, \mathcal{J}_c(\mathbf{w}^{c,\ell_{n_s}})\}, \tag{26.3}$$

the existence of such an optimum being conditioned by the relation $\mathcal{J}_c(\mathbf{w}^{c,\ell_{\text{opt}}}) < \min_{\ell=1,\ldots,N_d} f_{\text{HFCM}}(\mathbf{w}^{c,\ell})$.

26.3 Analysis

The proposed approach is applied using the available full data basis issued from the stochastic nonlinear computational model of a detuned bladed disk structure with 12 blades and with 2 types of sectors. With such a data basis, there are $n_c = 216$ possible detuning configurations with a majority of blades with type 0, and the dimension N_d of subset $\mathcal{D}_d$ can vary from 1 to 202. Table 26.1 summarizes the main optimization results. It is seen that the improving detuning configurations represented in Fig. 26.1 and highlighted in green can be found as optimal solutions with a reasonable number N_d of training points. Note that such an algorithm converges to the optimal solution $\mathbf{w}_{\text{un}}^{c,\text{opt}}$ as defined in Eq. (26.2) and represented in orange in Fig. 26.1, when all the $n_c = 216$ possible detuning configurations are included in the training set, which makes this formulation as a coherent one.

Table 26.1 Results as a function of the number N_d of training points in $\mathcal{D}_d$—results obtained for $N_d = n_c = 216$ (reference). Bold results are related to the optimal value of the cost function and defines the optimal quantity of interest for a given number N_d of training points

N_d	$j = 1, \ldots, 6$	$\mathcal{J}_c(\mathbf{w}^{c,\ell_j})$, $\mathbf{w}^{c,\ell_j} \in \mathcal{W}_{n_s}^{\mathrm{opt}}$						ℓ_{opt}	$q^{c,\ell_{\mathrm{opt}}}$
50	q^{c,ℓ_j}	0.9995	0.9974	0.9834	1.0000	**0.9822**	1.0015	141	0.9822
75	q^{c,ℓ_j}	0.9974	0.9882	1.0000	1.0015	1.0067	**0.9529**	166	0.9529
100	q^{c,ℓ_j}	0.9974	1.0000	0.9995	**0.9538**	0.9822	0.9882	123	0.9538
202 (whole $\mathcal{D}_d$)	q^{c,ℓ_j}	1.0000	1.0015	0.9882	**0.9538**	1.0067	0.9822	123	0.9538
Reference	q^{c,ℓ_j}	**0.9476**	0.9529	0.9538	0.9822	0.9834	0.9824	104	0.9476
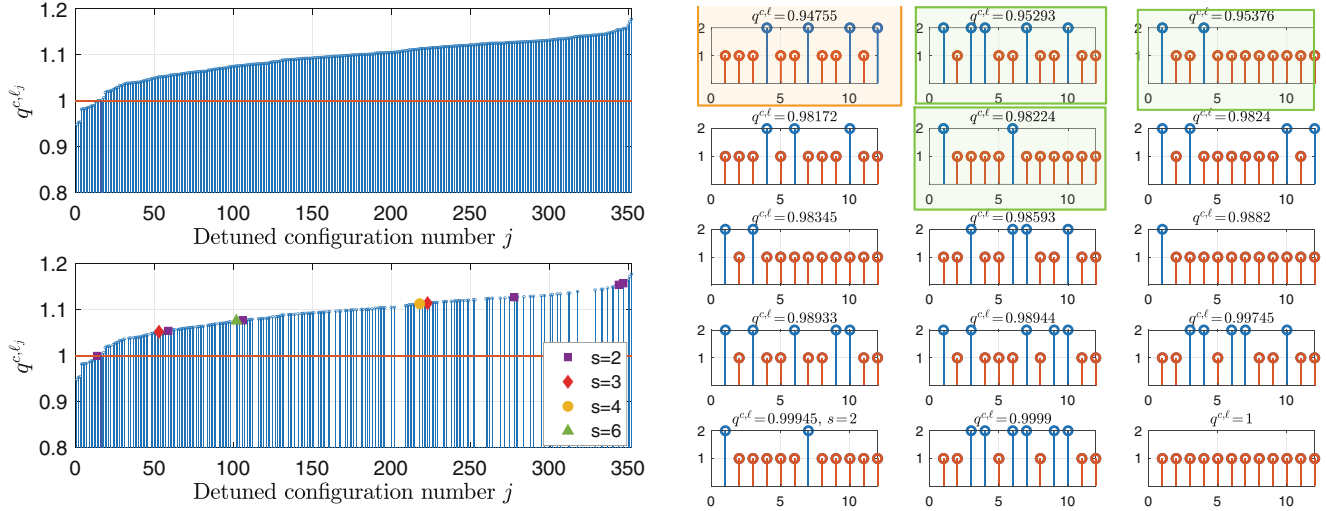
Fig. 26.1 Quantity of interest according to the detuned configuration. Graph of function $j \mapsto q^{c,\ell_j}$ for the $n_c = 352$ possible detuned configurations (left upper figure) and for the $n_c = 216$ detuned configurations with a number of blades with type 2 less than or equal to 6 (left lower figure). Sub-cyclicity order s is also given. Characteristics of the improving detuned configurations: graph of $i \mapsto w_i^{c,\ell_j}$, $j \in \{1, \ldots, 15\}$ (right figures)

26.4 Conclusion

In a detuning optimization context involving a large number of possible detuned configurations, we have proposed a reformulation in a probabilistic framework of the combinatorial optimization problem, which is adapted to a probabilistic machine learning tool in order to limit the number of evaluations of the cost function with the high-fidelity computational model. The methodology proposed has been validated for a 12-bladed disk structure for which the exact optimal detuning configuration in the presence of random mistuning has been previously identified. A good prediction of the optimum has been obtained with this method, which demonstrates the efficiency and the capability of the proposed methodology.

References

1. Bladh, R., Castanier, M.P., Pierre, C.: Component-mode-based reduced order modeling techniques for mistuned bladed disks. Part 2: application. ASME J. Eng. Gas Turbines Power **123**(1), 100–108 (2001)
2. Han, Y., Murthy, R., Mignolet, M.P., Lentz, J.: Optimization of intentional mistuning patterns for the mitigation of the effects of random mistuning. ASME J. Eng. Gas Turbines Power **136**(6), 062505 (2014)
3. Capiez-Lernout, E., Soize, C.: Nonlinear stochastic dynamics of detuned bladed-disks with uncertain mistuning and detuning optimization using a probabilistic machine learning tool. Int. J. Non Linear Mech. **143**, 104023 (2022)
4. Picou, A., Capiez-Lernout, E., Soize, C., Mbaye, M.: Robust dynamic analysis of detuned-mistuned rotating bladed disks with geometric nonlinearities. Comput. Mech. **65**(3), 711–730 (2020)
5. Soize, C., Ghanem, R.: Data-driven probability concentration and sampling on manifold. J. Comput. Phys. **321**, 242–258 (2016)
6. Soize, C., Ghanem, R.: Probabilistic learning on manifolds (PLoM) with partition. Int. J. Numer. Methods Eng. **123**(1), 268–290 (2022)
7. Soize, C., Ghanem, R.: Physics-constrained non-Gaussian probabilistic learning on manifolds. Int. J. Numer. Methods Eng. **121**(1), 110–145 (2020)

Chapter 27
Model-Based Inspection Planning for Large-Scale Structures Using Unmanned Aerial Vehicles

Zihan Wu, Jice Zeng, Zhen Hu, and Michael D. Todd

Abstract This chapter proposes a model-based inspection planning framework for damage diagnostics and maintenance optimization of large-scale deteriorating structures. An objective function is constructed using the definition of cost per unit time (CPUT) based on the three key parameters of unmanned aerial vehicle (UAV) inspection: inspection distance, inspection interval, and crack limit for repair action. To accelerate the process of optimizing the expensive-to-evaluate function, a Bayesian optimization is employed which improves the time and computational efficiency. The results demonstrate that the proposed method is able to efficiently determine the optimal parameters of UAV inspection parameters and continuously update the information model.

Keywords Unmanned aerial vehicles · Miter gates · Inspection planning · Bayesian optimization

27.1 Introduction

Unmanned aerial vehicle (UAV) systems are gaining extensive attraction as an unmanned inspection technique. The UAV system allows for the implementation of high-resolution cameras and computer vision-based sensors, providing an efficient way of acquiring data for structural health monitoring (SHM). Current UAV-based SHM strategies focus more on improving data precision and consistency, which overlooks the crucial need to integrate UAV mission planning into the broader context of structure life-cycle management. In most applications, UAVs are given a flying path to cover the full field of the objective structure, where the economic cost of each UAV mission is not considered. When applied to large-scale structures, such as miter gates and dams, the inspection cost for each UAV mission becomes a significant factor, which should be carefully treated. This chapter proposes a model-based inspection planning framework that aims to provide a more informative mission planning for the UAV system by considering the physics of the inspection objectives and its damage risk–based profile. To simulate the process of obtaining damage condition assessment from structural analysis, a high-fidelity finite element (FE) model is first developed to represent large-scale structures that are experiencing multiple fatigue cracking. With an iterative global-local modeling method, fatigue cracks at multi-locations are realized and studied through FE analysis [1]. An objective function is constructed as cost per unit time (CPUT) where several factors that relate to UAV inspection are considered: the inspection distance, inspection interval, and critical crack limit for repair action. Based on the cost analysis, a Bayesian optimization is then performed to accelerate the process of finding the optimal parameters for UAV inspection.

27.2 Framework for Inspection and Maintenance Planning

As shown in Fig. 27.1, each UAV mission is defined by the inspection distance and crack limit of repair action. The inspection distance determines the flying path of the UAV, which further contributes to the inspection cost. Different inspection distances also impact the image quality of the obtained observations, which directly affects the confidence level of the damage

Z. Wu (✉) · M. D. Todd
Department of Structural Engineering, University of California San Diego, La Jolla, CA, USA
e-mail: z5wu@eng.ucsd.edu

J. Zeng · Z. Hu
Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, MI, USA
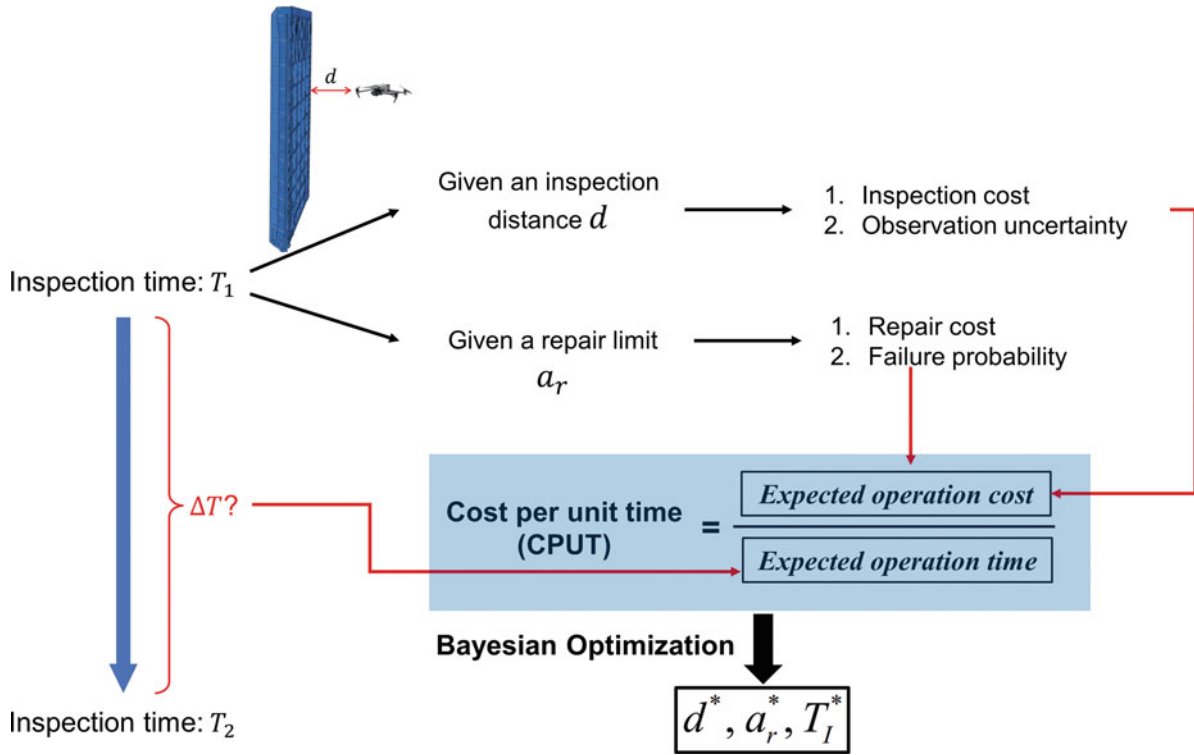
Fig. 27.1 Framework for inspection and maintenance planning

diagnostics and failure prognostics [2]. Similarly, the crack limit describes the protocol such that if a crack of a certain length is detected, a decision must be made as to whether the structure requires immediate repair. The decision of how to perform each UAV mission will result in different consequences and potential cost probability. Another significant factor, inspection interval, defines the operation time of the structure before the next inspection. Based on the remaining useful life of the structure predicted by the model, different operation times will undergo different risks and costs from potential structural failure (structures may fail between two inspection intervals). To quantitatively analyze the objective function for a particular UAV inspection mission, the CPUT is calculated given the inspection distance, a crack limit, and an inspection interval. The whole framework aims to find the optimal parameters for the inspection action and following inspection interval, which minimizes the objective function.

27.3 Bayesian Optimization

Considering the expensive-to-evaluate objective functions in CPUT, a Bayesian optimization is performed to efficiently find the parameters that lead to a minimum cost [3]. A kriging model is first built and trained based on a few training samples. To find the next potentially optimized point in the whole input space, the model is then updated by adding training points gradually, maximizing the expected improvement function at each iteration. In this chapter, the expected improvement is defined as

$$\text{EI}(\mathbf{x}) = \left(\mu(\mathbf{x}) - y^*\right) \Phi\left(\frac{\mu(\mathbf{x}) - y^*}{\sigma(\mathbf{x})}\right) + \sigma(\mathbf{x}) \phi\left(\frac{\mu(\mathbf{x}) - y^*}{\sigma(\mathbf{x})}\right)$$

where $\Phi(\bullet)$ and $\phi(\bullet)$ are the cumulative distribution function and probability density function of a standard Gaussian variable and y^* is the maximum prediction of the model based on current training samples. By maximizing EI, we can sequentially find new training points that gradually approach the optimal values.

27.4 Results

To illustrate the proposed framework, the crack growth patterns at three different locations of the miter gate are synthetically generated through an FE model as shown in Fig. 27.2. The three cracks with different initial crack lengths indicate that without proper inspection and corresponding repair action, the failure crack limit will soon be reached, leading to structural failure.

Table 27.1 shows some typical results of different combinations of UAV inspection parameters. It can be concluded that several strategies may lead to a minimum CPUT: setting a short inspection interval or setting a small crack limit (once find a crack, repair it). Otherwise, long inspection intervals without proper repair action may lead to catastrophic loss.

Figure 27.3 shows that the expected improvement is reduced periodically as the Bayesian optimization keeps finding the "weakest point" of the kriging model. Meanwhile, the CPUT is optimized and gradually converged to the true minimum value.
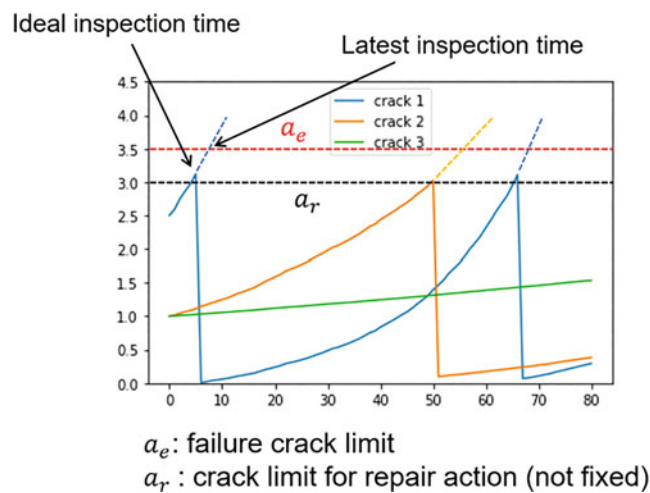
Fig. 27.2 Synthetic crack growth patterns at three different locations of the miter gate

Table 27.1 CPUT for different combinations of UAV inspection parameters

Inspection distance d (in.)	Crack limit for repair action (in.)	Inspection interval (months)	CPUT
0.72	3.25	2	5.05
0.89	3.5	10	1181.9
0.44	1.5	8	14.75
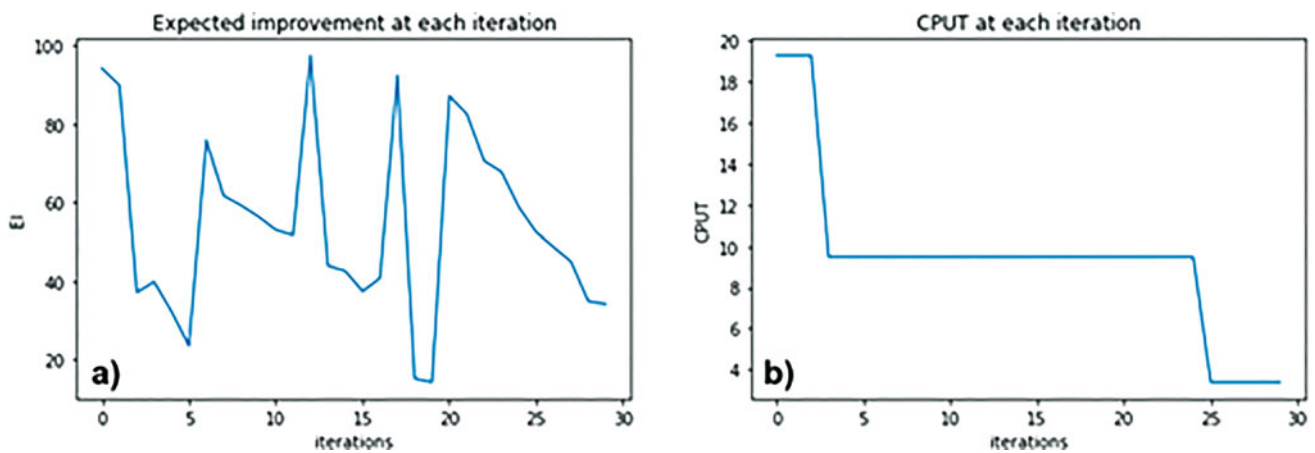
Fig. 27.3 (**a**) Expected improvement at each iteration and (**b**) CPUT at each iteration

27.5 Conclusion

The numerical results show that the proposed mission planning strategy can significantly reduce the economic cost without affecting the accuracy of potential crack estimation. It has the potential to provide accurate and high-quality SHM information for online diagnosis and prognosis. In addition, the incorporated Bayesian optimization successfully improves the computational efficiency of finding optimal solutions for UAV missions.

References

1. Fillmore, T.B., et al.: A surrogate model to accelerate non-intrusive global–local simulations of cracked steel structures. Struct. Multidiscip. Optim. **65**(7), 1–20 (2022)
2. Wu, Z., et al.: Diagnostics and prognostics of multi-mode failure scenarios in miter gates using multiple data sources and a dynamic Bayesian network. Struct. Multidiscip. Optim. **65**(9), 1–20 (2022)
3. Hu, Z., Xiaoping, D.: Mixed efficient global optimization for time-dependent reliability analysis. J. Mech. Des. **137**(5), 051401 (2015)

Chapter 28
The Effect of Temporal Correlations on State Estimation Through Variational Bayesian Inference

Motahareh Mirfarah, Alana Lund, and Shirley J. Dyke

Abstract Effective health monitoring in dynamic systems hinges on the proper estimation of the system's state. As one of the most powerful methods of state estimation, variational Bayesian inference provides a flexible framework for making probabilistic inference about the state of a system. This method approximates the latent variables' posteriors by minimizing the evidence lower bound (ELBO) loss function, which is an estimate of the information lost by approximating the true posterior with the selected variational family. This chapter focuses on the flexibility of the variational family and investigates the effect of considering dependency between the latent variables on the quality of posteriors, with respect to applications in structural health monitoring. The most common assumption made for the variational family is mean-field Gaussian, restricting the posterior's space by independent Gaussian distributions. As this space does not realistically represent the relation between the dynamic system's parameters, in this study, temporal correlations are added to the family's covariance matrix structure. To evaluate the effect of the temporal off-diagonal covariance matrix on the performance of the inference, a numerical simulation of a linear single-degree-of-freedom system is utilized. The results of employing this covariance structure in the variational family are discussed from the accuracy, robustness, and computational cost standpoints. The findings show that for this linear system, the estimates of the parameters would not be meaningfully enhanced by adding the temporal dependency to the mean-field structure of the variational family.

Keywords State estimation · Variational inference · Mean-field Gaussian · Covariance structure · Latent parameter dependency

28.1 Introduction

Variational Bayesian inference is a powerful framework for approximating the posterior distribution. The key feature of this approach is the optimization of a hypothesis class, called the "family of variational distributions," to find the nearest distribution to the true posterior. In the literature, the widely held assumption for the variational family is the "Gaussian mean-field," which considers the unknown variables to be independent and follow Gaussian distribution [1, 2]. However, the variational family of distributions needs to have adequate flexibility to be a good representative of the posterior. The mean-field assumption does not necessarily lead to realistic outcomes when it comes to the models with highly correlated variables, as the structure of the variational posterior does not allow such dependencies [3–5].

This research focuses on relaxing the mean-field assumption by considering dependent variables inside the structure of the variational family, providing a closer resemblance to the model. We employ this approach for the identification of a linear single-degree-of-freedom system, whose local parameters (displacement and velocity) at each time step are dependent with respect to the Markov-based dynamic model.

M. Mirfarah (✉)
School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA
e-mail: mmirfara@purdue.edu

A. Lund
Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada

S. J. Dyke
School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA

Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA
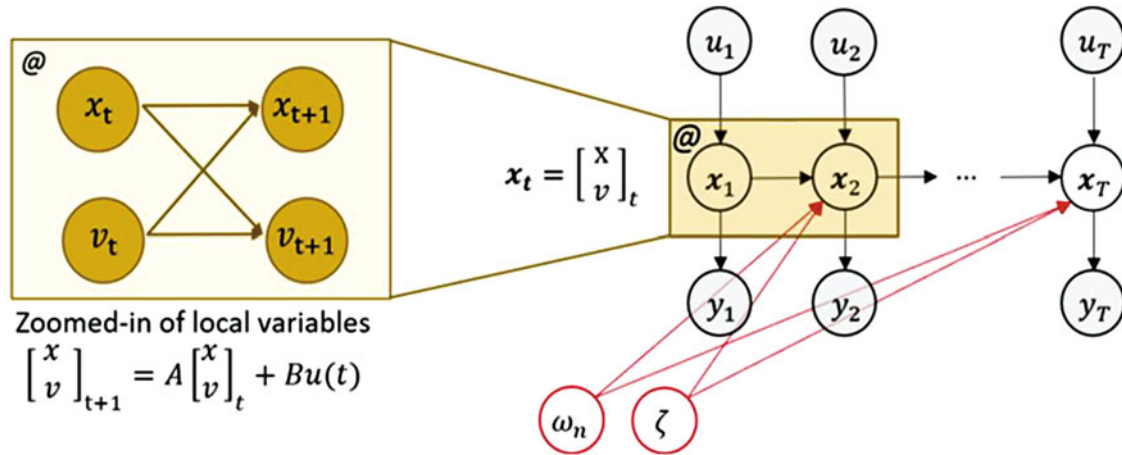
Fig. 28.1 Graphical model of the linear oscillator system

28.2 Methodology

The graphical model of the linear oscillator system investigated in this study is shown in Fig. 28.1. Variational inference is implemented for the identification of the latent variables in this system. This includes the estimation of local variables, displacements, and velocities ($x_1, \ldots, x_T, v_1, \ldots, v_T$), as well as the global parameters, natural frequency ($\omega_n = 20.0$ rad/s), and damping ratio ($\zeta = 0.1$), while assuming the mass to be known ($m = 25$ kg). The observations from the system are the noisy acceleration response at each time step (denoted by y_i). The input base excitation (represented by u_i) is a band-limited white noise signal with a maximum amplitude of 15 m/s^2, a cutoff frequency of 16 Hz, and a duration of 2 s. A 2% root-mean-square measurement and process noise is included in the numerical simulation of the system response.

In this study, two different arrangements are considered for the variational family. The first structure is the mean-field, which assumes independency between all the inferred variables. The other structure incorporates the temporal dependency between the local variables (displacements and velocities), while keeping the remaining global variables in the same mean-field configuration. This relationship can be equivalently represented by additional off-diagonal terms in the structure of the covariance matrix, as can be seen in Fig. 28.2. The variational inference approach implemented here is described in detail in [6]. The parameters of ω_n and ζ are constrained to the positive values, and a log-transformation is used to make the optimization unconstrained.

28.3 Analysis

The results for the implementation of the mean-field approximation and the temporally dependent variational family, with respect to the posterior of the global parameters, are shown in Fig. 28.3a, b, respectively. The results include the mode, 2.5 percentile, and 97.5 percentile of the posterior distribution for 50 different prior sets. As can be seen, the estimations of both structures are concentrated around the actual values with a similar spread, though the temporal dependence case (Fig. 28.3b) does seem to remove the bias evident in the mean-field case (Fig. 28.3a) at the cost of a small increase in the spread of the posterior means. Regardless, the temporally dependent covariance structure does not meaningfully improve the estimation results and incurs a higher computational cost due to the more complex structure than the diagonal mean-field. Therefore, we can conclude that for a linear dynamic system, the mean-field approximation sufficiently describes the true posterior and is an appropriate representation of the system.

28.4 Conclusion

The results of this study show that considering a mean-field or a temporally dependent structure for the variational family does not have a considerable influence on the identification of a linear single-degree-of-freedom oscillator. Both structures
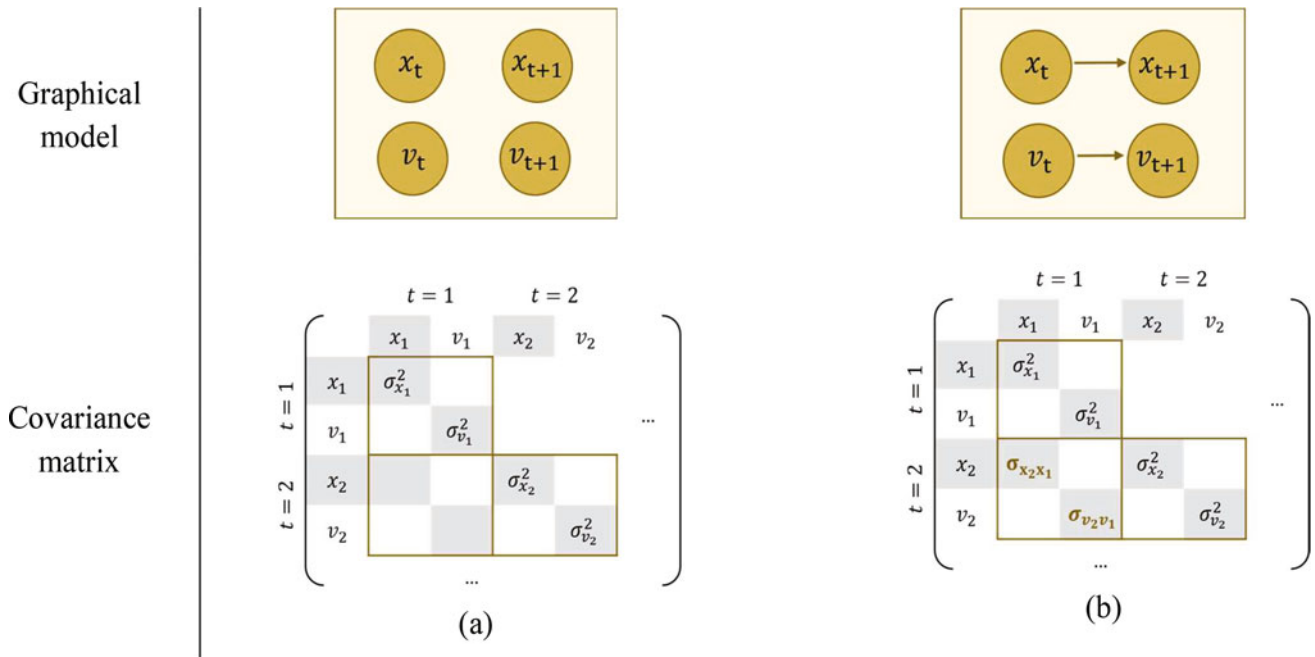
Fig. 28.2 The structure of variational family for (**a**) mean-field and (**b**) temporal dependency
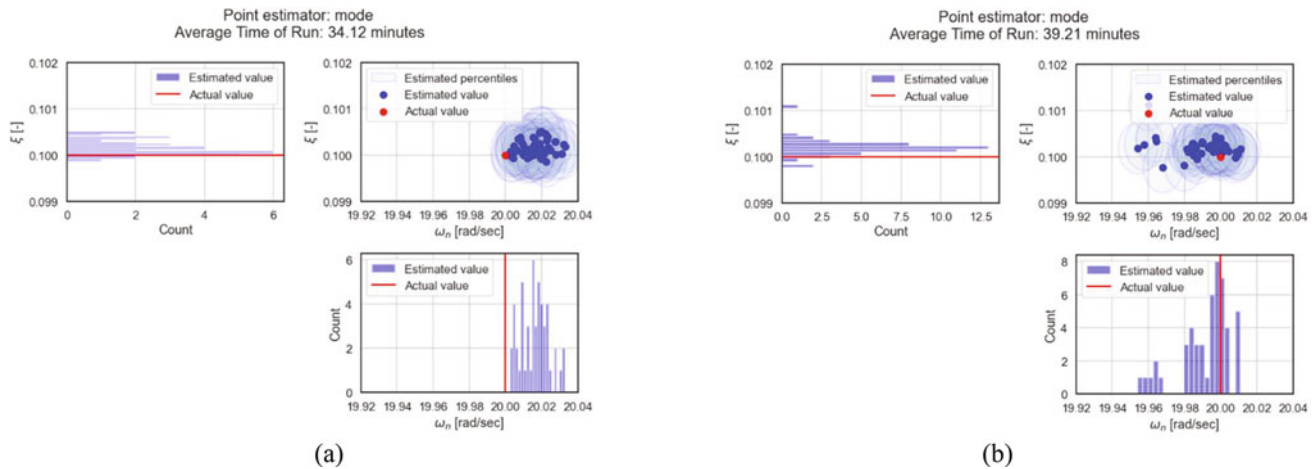
Fig. 28.3 Estimated posteriors for the natural frequency and damping ratio in 50 prior sets with (**a**) mean-field variational family and (**b**) temporally dependent variational family

provide estimations close to the actual values. Additional investigation into the posterior covariance structure for nonlinear systems could provide further insights into the application of these methods for full-scale structural health monitoring.

References

1. Bishop, C.M., Nasrabadi, N.M.: Pattern Recognition and Machine Learning, no. 4. Springer (2006)
2. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational inference: a review for statisticians. J. Am. Stat. Assoc. **112**(518), 859–877 (2017)
3. MacKay, D.J., Mac Kay, D.J.: Information Theory, Inference and Learning Algorithms. Cambridge University Press (2003)

4. Giordano, R.J., Broderick, T., Jordan, M.I.: Linear response methods for accurate covariance estimates from mean field variational Bayes. Adv. Neural Inf. Process. Syst. **28**, 1–9 (2015)
5. Tran, D., Blei, D.M., Airoldi, E.M.: Copula variational inference. In: International Conference on Neural Information Processing Systems (NIPS), pp. 1–9 (2015)
6. Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., Blei, D.M.: Automatic differentiation variational inference. J. Mach. Learn. Res. **18**, 1–45 (2017)

Chapter 29
On the Selection and Validation of Component Damage Models for Prediction of Damage-State Behavior of a Truss Bridge

James Wilson, Paul Gardner, Graeme Manson, and Robert J. Barthorpe

Abstract Structural health monitoring has seen significant progress in recent decades and offers major potential benefits in terms of life-cycle management of engineering infrastructure compared to traditional monitoring and maintenance methods. However, many challenges remain, including the lack of availability of sufficient damage-state data from structures of interest on which to validate physics-based models, which can be used to simulate the behavior of structures in their damaged conditions. This can potentially be avoided by validating the damage models at the components or subassemblies where damage would be expected to be found. It is hypothesized that the uncertainty quantified at the component level can then be propagated to the assembly level, thereby avoiding the requirement for damage-state data from the assembly.

This chapter presents an investigation of the process described above, where component-level damage-state data is used to select and validate predictive damage models of the struts of a truss bridge, which are then built into an assembly model using dynamic substructuring. The validated assembly-level damage predictions are then compared to a test dataset covering a range of damage conditions in the assembly, and the candidate component-level damage models are compared against each other in terms of accuracy of fit to the test data.

Keywords Validation · Hierarchy · Dynamic substructuring · Physics-based modeling · Structural health monitoring

29.1 Introduction

A significant outstanding issue in the implementation of structural health monitoring (SHM) strategies in practice is the lack of availability of data from the structure of interest in its damage states. This is usually due to the high cost of structures appropriate for health monitoring: deliberately introducing damage to these structures in order to acquire training datasets is not feasible. One way to provide training data for damage detectors is to simulate the data using physics-based models; however, validation data is still required across the predictive domains of these models. The problem of model validation can be reduced by validating the submodels of a substructured assembly. This means that damage-state testing could, in principle, be reduced to a few components of interest where damage would be expected to manifest within the larger assembly, thus significantly reducing the cost of validation data acquisition.

Dynamic substructuring provides the means for breaking down a complex model of an assembly into a set of submodels representing substructures and components of the larger assembly [1]. This has multiple benefits, notably enabling the construction of hierarchical models for SHM; a demonstration of the use of dynamic substructuring in an SHM context can be found in [2], where a representative numerical example showed that a set of uncertain submodels could be used to make uncertain predictions on damage-sensitive features for an assembly structure.

Models have been implemented for SHM in hierarchical systems in the literature previously [3, 4]; however, these studies have not investigated the potential benefits for model validation explicitly. Where substructuring with a view to model validation has been investigated, the findings have not been extended to real-life structures [5].

This chapter aims to present hierarchical validation as a legitimate means for tackling the lack-of-data problem in existing SHM strategies. This is done by applying an initial framework for hierarchical validation to a substructured model of a laboratory-scale truss bridge, where uncertainty quantified through validation tasks at the submodel level will be shown to allow quantification of uncertainty at the assembly level. When compared to assembly-level test data, the hierarchically

J. Wilson (✉) · P. Gardner · G. Manson · R. J. Barthorpe
Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Sheffield, UK
e-mail: jwilson11@sheffield.ac.uk; p.gardner@sheffield.ac.uk; graeme.manson@sheffield.ac.uk; r.j.barthorpe@sheffield.ac.uk
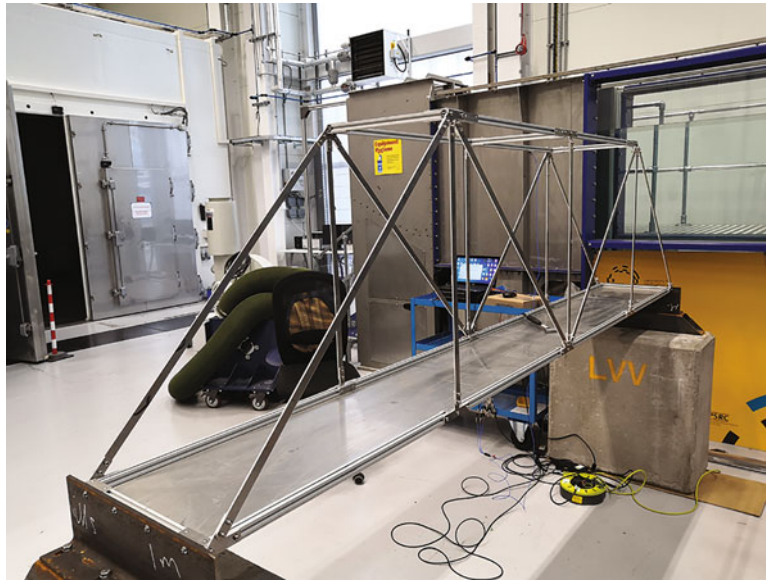
Fig. 29.1 The bridge set up for roving hammer data acquisition

validated model will be shown to offer accurate probabilistic predictions of damage-sensitive features across a range of damage conditions in the assembly.

29.2 Truss Bridge: Structure, Model, and Substructures

The structure of interest in this study is a laboratory-scale truss bridge, shown in Fig. 29.1. The bridge is representative of common designs implemented in the real world [6] and is composed of three main sets of substructures: the struts, the upper framework, and the deck. The struts of a truss bridge are the main load-bearing components of truss bridges and were therefore the key components of interest for damage detection. The struts were cut from 5083-'O'/H111 grade aluminum plate. The deck was composed of two components: a plate section (also 5083-'O'/H111 grade aluminum) and an aluminum Rexroth border. The upper frame was also constructed of aluminum Rexroth sections. Joint components (a series of brackets and connecting bolts) are also part of the assembly; however, they were neglected from the model in this research.

Submodels were constructed using ANSYS MAPDL to represent the struts, deck, and upper frame as substructures of the bridge. The beam sections of these submodels (struts, Rexroth) were meshed using BEAM188 elements, and the deck plate was meshed using SHELL181 elements. The two elements types were verified for modal analysis against analytical solutions for beams and plates [7], and a grid convergence analysis was carried out to set the optimum element size for each substructure [8].

The assembly model was constructed from these submodels using primal dynamic substructuring applied in the physical domain. The nodes of the submodels at the joint interfaces were identified for each submodel and these were constrained to each other in all degrees of freedom to create a model of the assembly.

29.3 Submodel Validation

In order to attain an assembly-level predictive damage model of the bridge, the individual strut submodels were validated against experimental data. This entailed calibration of the material parameters of the submodels, and calibration, selection, and validation of the crack models.
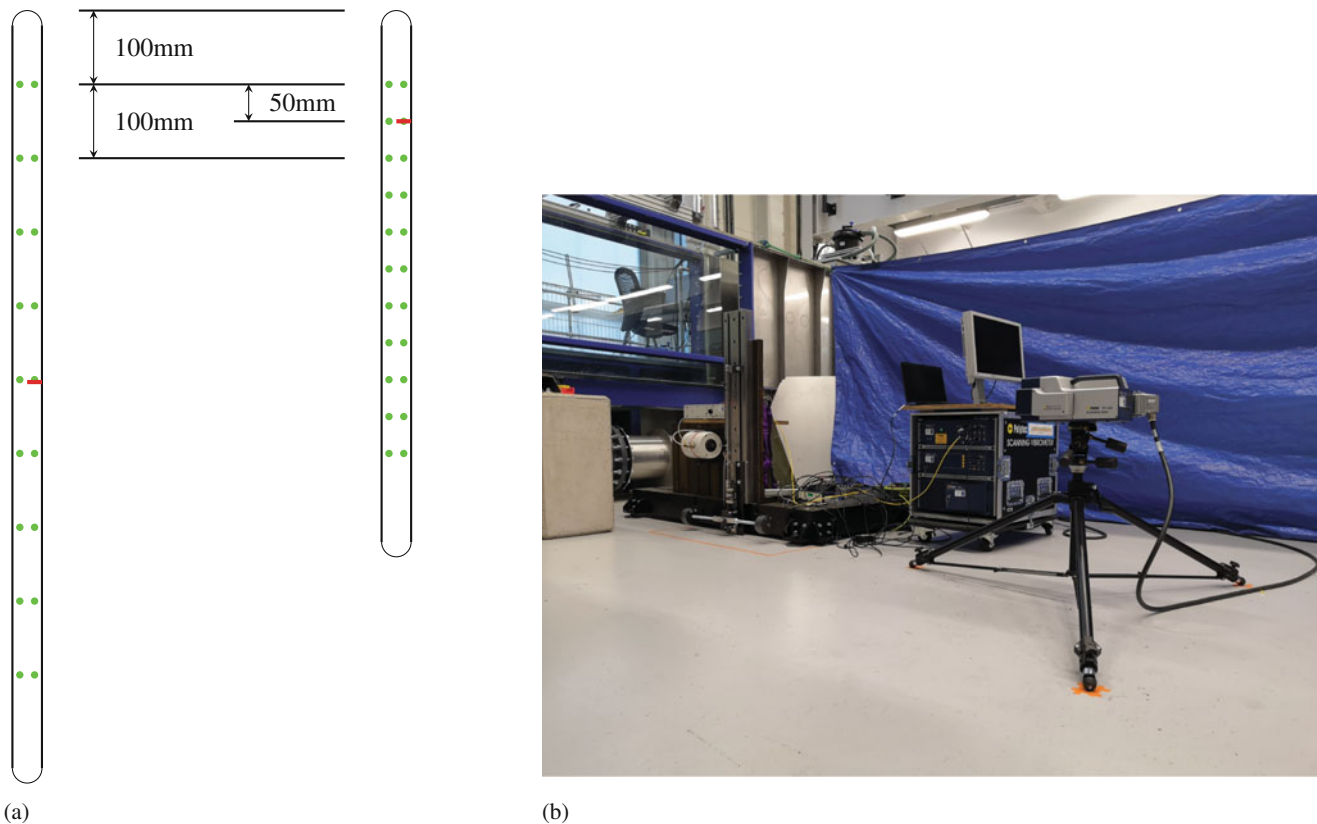
Fig. 29.2 The experimental setup for strut-level validation data. (**a**) The damage (red lines) and measurement locations (green dots) for each strut. (**b**) The strut set up with scanning laser vibrometer

29.3.1 Experimental Data

Validation data was gathered for both the vertical (shorter) and diagonal (longer) strut types in isolation across a range of damage extents and load conditions. The data was recorded to enable the confident predictions of the struts' behavior under damaged conditions as part of the bridge assembly. The struts were excited using a white noise signal provided by a piezoelectric shaker, and the dynamic response was recorded at a number of locations using a scanning laser vibrometer (see Fig. 29.2). Damage was introduced as a saw cut at 4 mm intervals at the midpoint of the diagonal strut and 150 mm from the top end of the vertical strut. The struts were clamped at each end and a weighted runner was used to apply a range of compressive and tensile load conditions along the length of each strut.

Mode matching between the experimental data and model predictions was carried out by comparing the mode shapes and natural frequencies from each; this then provided a set of natural frequencies obtained from each dataset on which calibration and validation could be carried out.

29.3.2 Parameter Calibration

The calibration of the strut models in their undamaged condition was carried out in order to determine the uncertainty in the material parameters of the models and ensure the accuracy of their predictions. The calibrated models could then be used in the validation of the damage model applied to both struts, which would allow for confidence to be established in the predictions of the models under damage conditions.

The material parameters of the strut models were Young's modulus, density, and Poisson's ratio. The nominal values of these for the struts were 72 GPa, 2650 kg/m^3, and 0.33, respectively. The Young's modulus was optimized against model data from the strut-level tests using the Nelder–Mead simplex algorithm [9]. The optimization was carried out at each load case for each strut to minimize the mean difference between the first four natural frequencies of the strut. A Gaussian distribution

was then fitted to the results of these optimizations based on the assumption that the vertical and diagonal struts had the same underlying material parameter distributions. This resulted in a calibrated strut Young's modulus distribution with a mean of 71.5 GPa and a standard deviation of 1.8 GPa. The mean of this distribution was then validated by testing it against the fifth natural frequency of each strut, which showed concurrent error levels compared to the errors from the first four natural frequencies it was trained on, and a reduction compared to the nominal settings across the full range of test points.

29.3.3 Model Selection and Validation

Predictive damage models for cracks in beams can be broadly categorized as element stiffness reduction models, distribution models, joint models, and element removal models [10]. The strut submodels used in this study used 2D beam elements and were therefore most adaptable to the element stiffness reduction models and distribution models. Three models based on these techniques were tested in this study and are described below.

Model 1 was essentially an element stiffness reduction damage model. The inputs to the model were crack depth and location on the strut. The element at the crack location was then reduced in stiffness by reducing its cross-sectional depth by the depth of the crack, with the centroid offset from the undamaged elements as shown in Fig. 29.3. The key parameter of the model was crack width, which determined the size of the element for stiffness reduction. Model 2 was an extension of model 1 with an additional parameter that controlled the Young's modulus of the element at the damage location, α; the contribution of this parameter is described in Eq. 29.1.

$$E_{cross-section} = E_{strut} \left(\frac{d_{strut} - d_{crack}}{d_{strut}} \right)^{\alpha}$$

(29.1)

Model 3 was a stiffness distribution damage model based on a Gaussian distribution to describe the Young's modulus at each node of the model, as was demonstrated in [11]. The inputs were the crack location (which was set as the mean of the Gaussian distribution) and depth (which determined the maximum likelihood of the distribution), and the tunable parameter was crack width (which was set as the standard deviation of the Gaussian distribution). The form of the function is given in Eq. 29.2.

$$E_{node} = E_{strut} \left(1 - \frac{d_{crack}}{d_{strut}} e^{-\frac{1}{2} \left(\frac{x_{node} - x_{crack}}{w_{crack}} \right)^2} \right)$$

(29.2)

Posterior distributions of the parameters of the above damage models were estimated using approximate Bayesian computation (ABC) [12]. This algorithm draws samples of the model parameters from a set of prior distributions, and then tests the model predictions based on the samples against experimental data. An error metric is defined such that any sample points that produce an error greater than the threshold on this metric are discarded, while those that produce a lower error are retained and used as an estimate of the posterior distribution. This makes ABC a likelihood-free method for estimating the posterior distributions of model parameters; this was the reason ABC was selected for this research as it removed the task of fitting a formal likelihood distribution to the experimental data while enabling the incorporation of prior knowledge on the model parameters. The prior for crack width was set to a Gaussian distribution with a mean of 5 mm and a standard
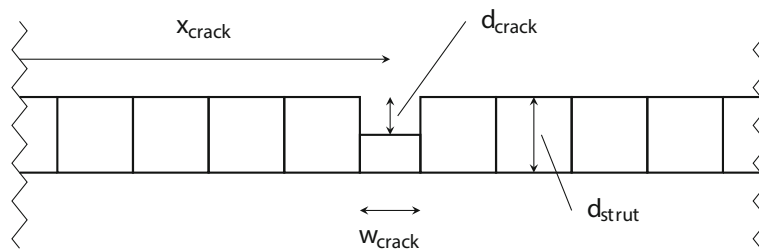

Fig. 29.3 Graphical representation of damage model 1 with inputs and parameters marked (x, d and w represent distance, depth, and width, respectively)

deviation of 1 mm; the prior for α was set to a Gaussian distribution with a mean of 0 and a standard deviation of 0.1. These were sampled 1000 times each alongside the material parameter distributions using Latin hypercube sampling to create 1000 candidates for ABC from each damage model.

The training set utilized the proportional change in natural frequency from the undamaged condition at damage extents of 8 and 16 mm on the diagonal strut, while the validation set used the same feature at 4 and 12 mm, on both struts. The error function upon which the ABC threshold was set was the mean of the difference between the predictions and experimental data at the damage extents described above, taken across all damage and load cases and across the first five modes. The minimum acceptance rates for the posterior were set to 10% for each model, which yielded a set of posterior estimations for each parameter of each damage model; these were then propagated upward for the assembly-level predictions.

Based on analysis of the acceptance rates and error against the validation data, model 3 performed the best of the crack models, followed by model 2 and finally model 1. However, it should be noted that model 1 is the most direct physical representation of the true damage in an engineering sense, while models 2 and 3 are more "functional" in their attempts to fit to the training data. The parameters of model 3 in particular (where crack width is set as the standard deviation of the Gaussian distribution) lack physical meaning. This may mean that the model would not generalize as well when exposed to novel data—model 3 did in fact perform the least well when comparing the nominal and calibrated settings against the validation data (models 1 and 2 showed significant improvements from the nominal settings after calibration). All three models were then tested at the assembly level in order to determine the success of the validation processes and to see whether model selection at this stage (which would favor model 3) is robust at the assembly level.

29.4 Assembly-Level Prediction

The aim of this chapter is to demonstrate that assembly-level damage-state predictions can be made with the associated quantified uncertainty based on validation tasks carried out at the subassembly level. This is shown in this section by comparing the predictions of the assembly models to experimental data across a range of damage states.

29.4.1 Experimental Data

A new set of data was required for this section covering the behavior of the assembly under a range of damage conditions. Two datasets were recorded in immediate succession, the first being a roving hammer test to acquire a high-resolution set of mode shapes for the assembly in its healthy state and the second being a test of the assembly response to damage. Both tests used a PCB Piezotronics model 086C03 impact hammer to excite the structure and PCB Piezotronics accelerometers. Spectral response data was recorded in the range 0–128 Hz at a resolution of 0.0625 Hz and modal features were extracted using the PolyMAX curvefitting algorithm.

The first dataset here allowed for mode matching between the nominal model (pre-validation) and the true structure based on natural frequencies and mode shapes. Following the mode matching, the nominal models were then used to determine a set of features that showed sensitivity to damage in each strut where damage was introduced (see Fig. 29.4). These features were the proportional change in two natural frequencies as damage progressed, selected as described to indicate damage in each strut. These were labeled "Feature 1" and "Feature 2" for simplicity in this chapter.

For the damage-state testing, a single impact location was used with multiple accelerometers, as shown in Fig. 29.4. Damage was introduced in 2.5 mm increments by saw cut to the midpoints of struts 1, 5, 2, and 6, up to a maximum of 17.5 mm. Following the damage test on each strut, the strut was removed and replaced with an undamaged version before commencing the damage test on the next strut. The bolts fastening the components of the assembly together were tightened to a torque of 8 Nm using a torque wrench in an effort to maintain consistent boundary conditions across the testing.

29.4.2 Model Testing

The uncertainty quantified at the strut level was propagated to the assembly level by running each posterior sample through the assembly process for a given damage input. The healthy struts in the assembly were assigned new samples from the calibrated distribution for Young's modulus at each posterior sample point. This yielded a set of uncertain predictions of the
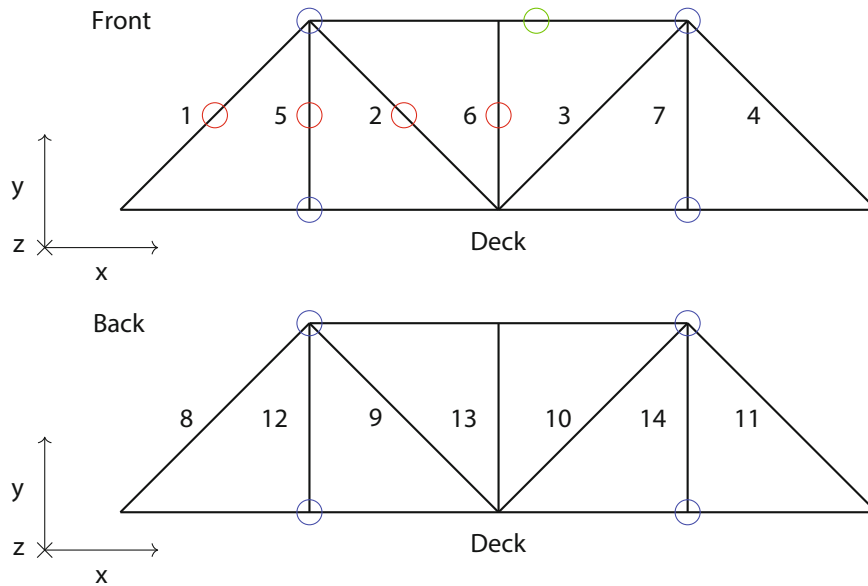
Fig. 29.4 Diagram of tap location (green), accelerometer locations (blue), and damage locations (red) for the assembly-level damage-state bridge testing
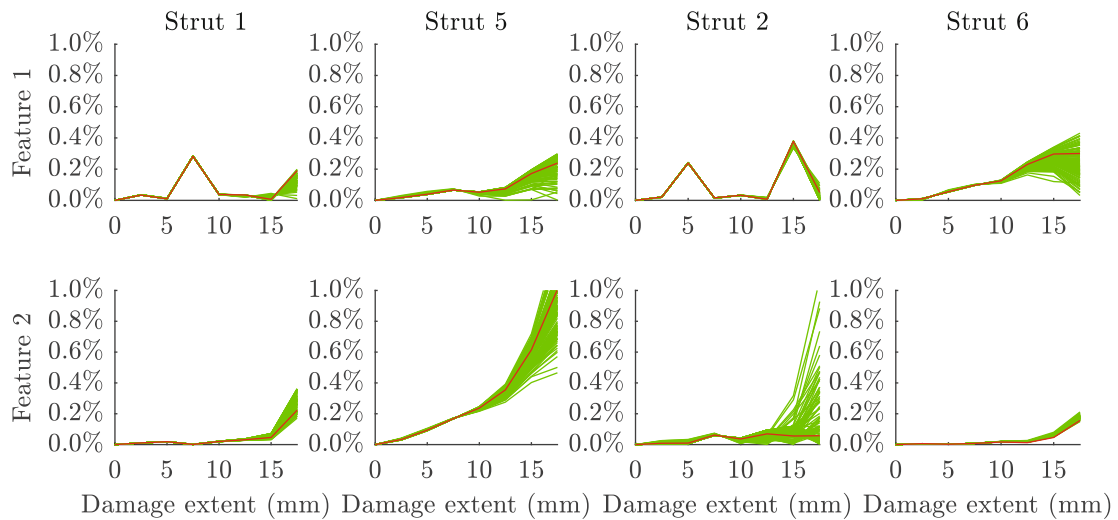

Fig. 29.5 The prediction error as damage progresses for the nominal (red) and validated (green) predictions for model 1

features of interest for each given damage input. These could, in future work, be used to train a set of damage classifiers for implementation on future data drawn from the structure during ongoing monitoring, as was demonstrated in [13]. Analysis of the classification results would then give more information on the legitimacy of the component-level validation and selection processes.

The assembly model was then run at each of the damage conditions used in the experimental data, and the accuracy of the predictions was compared against their experimental equivalents. To interrogate the success of the strut-level validation process, the model error for the validated uncertain predictions was plotted alongside the error at the nominal settings for each model (the prediction error in this context was simply the difference between the predicted and experimental features). These comparisons are illustrated in Figs. 29.5, 29.6 and 29.7.

These figures show that the validation process has allowed for estimation of significant uncertainty in the feature predictions that would have been unavailable otherwise—in the absence of validation data no inference can be made on prediction uncertainty. In addition to this, it can be seen that there are a number of cases (e.g., model 2, strut 5; model 3, strut 6) where the overall prediction accuracy is improved compared to the nominal models. However, for the majority of cases
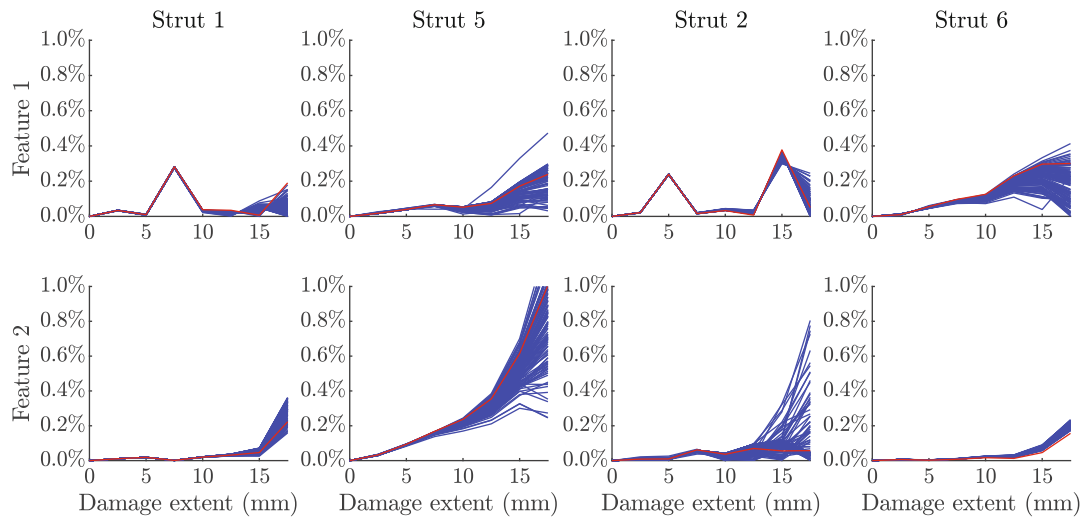
Fig. 29.6 The prediction error as damage progresses for the nominal (red) and validated (blue) predictions for model 2
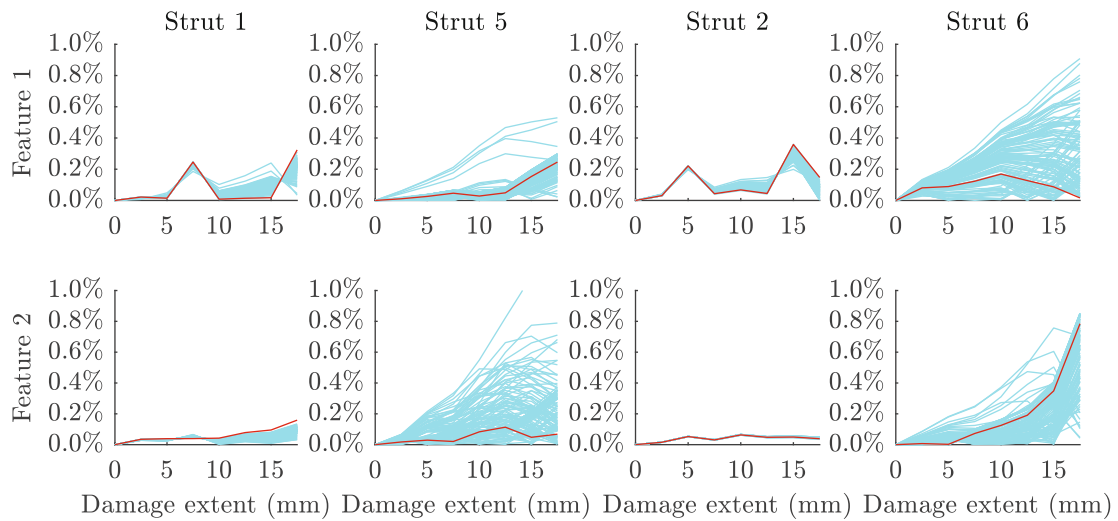
Fig. 29.7 The prediction error as damage progresses for the nominal (red) and validated (cyan) predictions for model 3

no clear improvement in predictive accuracy is discernable; this is likely due to experimental inaccuracies in the validation data combined with the relatively well-selected prior parameters of the nominal models.

Despite this, the results indicate success in the hierarchical validation process. Most importantly, it can be clearly seen that carrying out validation in this hierarchical sense is possible for predictive damage models and potentially provides the means for carrying out informed and probabilistic health-state classification in the future.

29.5 Conclusions

This chapter presented a novel method for validating physics-based models for SHM and demonstrated its potential through the use of realistic experimental datasets. Hierarchical validation in conjunction with dynamic substructuring of predictive damage models offers the potential for developing models for SHM without the need for assembly-level validation data—a key issue in the implementation of SHM strategies.

A truss bridge model was validated for damage-state prediction at the component level, and the results of the validation process allowed for the assembly-level model to be used in a probabilistic sense; the predictions for a given set of damage inputs gave a predicted feature distribution that was derived from comparison to real data. This allowed for quantification

of uncertainty at the assembly level and showed good performance compared to the unvalidated model predictions when compared against assembly-level test data with the structure of interest in a range of damage states.

Further work would entail the development of a damage detector based on the validated model predictions. This would enable analysis of the success of the model selection process by contrasting the classification accuracy of the damage detectors trained by each model.

References

1. Allen, M.S., Rixen, D., van der Seijs, M., Tiso, P., Abrahamsson, T., Mayes, R.L.: Substructuring in Engineering Dynamics: Emerging Numerical and Experimental Techniques, 1st edn. Springer, Cham, Switzerland (2020)
2. Wilson, J., Gardner, P., Manson, G., Barthorpe, R.J.: Hierarchical model verification and validation for structural health monitoring using dynamic substructuring. In: European Workshop on Structural Health Monitoring 2022, pp. 533–542. Cham (2022)
3. Su, W.C., Huang, C.S., Hung, S.L., Chen, L.J., Lin, W.J.: Locating damaged storeys in a shear building based on its sub-structural natural frequencies. Eng. Struct. **39**(1), 126–138 (2012)
4. Wang, S., Xu, Y., Gu, C., Xia, Q., Hu, K.: Two spatial association-considered mathematical models for diagnosing the long-term balanced relationship and short-term fluctuation of the deformation behaviour of high concrete arch dams. Struct. Health Monitor. **19**(5), 1421–1439 (2019)
5. Li, C., Mahadevan, S.: Role of calibration, validation, and relevance in multi-level uncertainty integration. Reliabil. Eng. Syst. Safety **148**(1), 32–43 (2016)
6. Maes, K., Van Meerbeck, L., Reyners, E.P.B., Lombaert, G.: Validation of vibration-based structural health monitoring on retrofitted railway bridge KW51. Mech. Syst. Signal Process. 165, 108380 (2022)
7. Blevins, R.D.: Formulas for Natural Frequency and Mode Shape, 1st edn. Krieger, Malabar, Florida (2001)
8. Roache, P.J.: Fundamentals of Verification and Validation, 1st edn. Hermosa Publishers, Socorro, New Mexico (2009)
9. Nelder, J.A., Mead, R.: A simplex method for function minimization. Comput. J. **7**(4), 308–313 (1965)
10. Friswell, M.I., Penny, J.E.T.: Crack modelling for structural health monitoring. Struct. Health Monitor. **1**(2), 139–148 (2002)
11. Bruns, M., Hofmeister, B., Grießmann, T., Rolfes, R.: Comparative study of parameterizations for damage localization with finite element model updating. In: 29th European Safety and Reliability Conference, pp. 1125–1132, Hannover (2019)
12. Beaumont, M.A., Zhang, W., Balding, D.J.: Approximate Bayesian computation in population genetics. Genetics Soc. Amer. **162**(4), 2025–2035 (2002)
13. Barthorpe, R.J., Hughes, A.J., Gardner, P.: A forward model driven structural health monitoring paradigm: Damage detection. In: Proceedings of the Society for Experimental Mechanics Series, pp. 119–126, Cham (2022)

Chapter 30
Surrogate Aerodynamics Modeling Applied to Surrogate Structural Dynamical Systems

Jonathan R. Smith

Abstract Surrogate modeling can be used to rapidly develop and test a complex system whose broad characteristics can be simplified to answer specific questions. In turn, the inputs to these systems can also be converted into surrogates that represent real test scenarios, all of which drive the design process. The driver is to accelerate product realization by means of digital engineering practices.

Our team at Sandia National Laboratories has developed a workflow that combines aerodynamic and structural surrogate models to analyze the effects of broad component parameters on the primary modes of a system. Given a desired preliminary design, a surrogate model is created which can be used to obtain mode shapes and frequencies that, ideally through proper modeling and understanding, are representative of the full system. Furthermore, our team has developed a range of surrogate structural dynamics models in the interest of a "ground-up" approach, wherein the simpler models are used to inform each subsequent, more complex iteration.

An aerodynamic model is also created from the expected conditions the structure will be subjected to. Power spectral density (PSD) is extracted and converted into pressure data and further into forces applied to the structure, providing the capability to fully analyze the modal behavior of the system via the interfacing of surrogate models. An additional challenge addressed is a quick-turn model credibility process to match the speed at which these surrogate models are developed.

The surrogates facilitate rapid simulation and design changes, significantly cutting the time and computing power required to obtain modal and test information compared to high-fidelity alternatives.

Keywords Surrogate model · Environmental forces · Structural dynamics · Digital engineering

30.1 Introduction

Digital engineering practices emphasize the need for early development analysis in the interest of improving down-selection by incorporating "low-fidelity," quick turnaround results. The surrogate models which form the foundation of this analysis can be linked while still retaining relative accuracy around the quantities of interest. Selectively reducing the degrees of freedom reduces computation costs and time to the point at which modeling simulations can be run on local machines via MATLAB and Simulink, rather than requiring high-powered computing resources.

The SOLSTICE[1] team at Sandia National Labs has developed a workflow that utilizes the surrogate methodology and applies it to multiple physics domains to acquire structural dynamic analysis of a system. As a broader goal of this team,

---

[1] Simulation Of Linked multi-physics Surrogate Time-domain models In Combined Environments.

J. R. Smith (✉)

Virtual Technologies and Engineering Department, Sandia National Laboratories, Albuquerque, NM, USA
e-mail: jonsmit@sandia.gov

the capability of producing targeted, rapid results is crucial toward mission progress with the intention of pushing digital engineering to the forefront of our efforts at Sandia.

30.2 Background

High-fidelity models are necessary parts of the design process to assure reliability before and during the production and testing of a product; however, these models understandably require large amounts of time and computing power, which can be prohibitive when it comes to development timelines. SOLSTICE is a means by which to incorporate the early design portion of the digital engineering architecture. By developing a surrogate capability that can showcase the potential use cases of a system that is deliberately targeted to primary quantities of interest, down-selection can be better informed sooner and high-fidelity modeling can begin with more information and a narrower range of simulation boundaries.

SOLSTICE exists solely within the commercial-off-the-shelf (COTS) environment of Mathworks' MATLAB and Simulink. This allows the team to operate with the full suite of tools provided by Mathworks while providing the freedom to develop models and interfaces for a wide variety of problems. Further, SOLSTICE can interact with other software to perform as the key analysis tool for designs along a product's life cycle. As designs change, capabilities that have been developed can accommodate those changes more rapidly than high-fidelity modeling, thus acting as continuous support and corroboration.

For this capability, the intention was to take a system within a presumed operating environment, produce a surrogate pressure model along limited points of interest, and normalize those pressures into forces that can be applied to a surrogate structural dynamics model of the system. In doing so, SOLSTICE is capable of tying together the physics domains of aerodynamic analysis and structural dynamic analysis.

30.3 Process

The general architecture for this workflow can be seen in Fig. 30.1, wherein the surrogate pressure model processes an environment scenario and subsequently produces power spectral density (PSD) pressure data in the frequency domain to be passed off to the surrogate mechanical model. This is done by way of an inverse Fourier transform process that turns the PSDs – with units of Pa^2/Hz – into normalized surface forces in the time domain to be simulated within the structural dynamics piece of the workflow. Ultimately, the product of this analysis is the *structural dynamics* PSD as a measure of the structural integrity of the system when in operation.
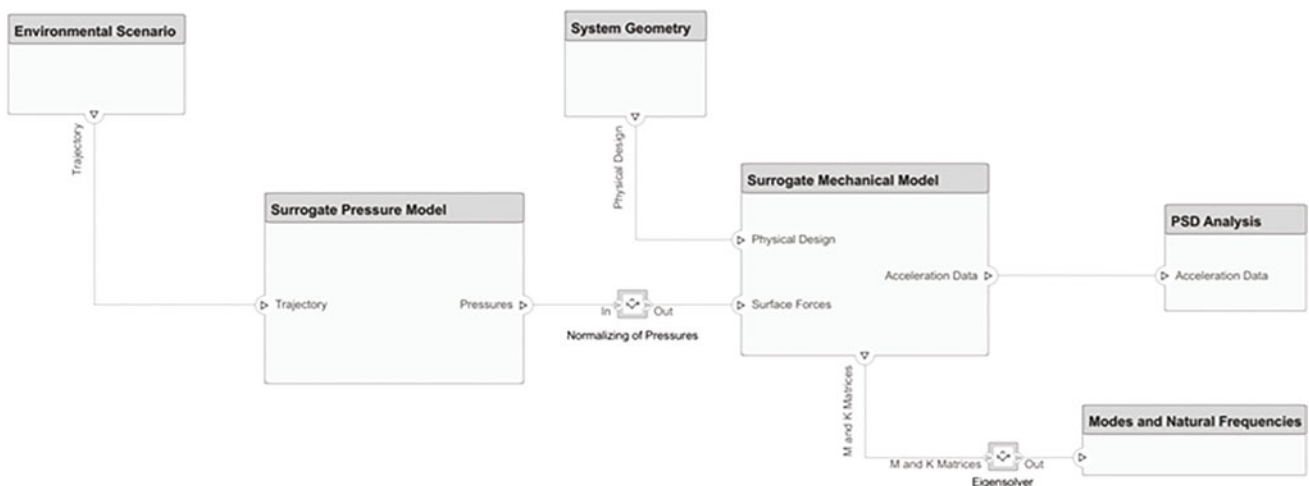

Fig. 30.1 Surrogate model architecture diagram

30.4 Surrogate Pressure Model

The aerodynamic model that serves as the foundation for the surrogate pressure results can be produced from any operational scenario given the trajectory and shape of the body. The primary contributors to consequential pressures are velocity, angle of attack (for flight vehicles), boundary layer thickness, and surface friction based on the fluid medium. While analysis within SOLSTICE primarily results in time domain data, the surrogate pressure model in this workflow produces PSDs in the frequency domain, largely due to expected test data and energy normalization.

The surrogate state comes after a reduction of the aerodynamic model when decreasing the number of analysis points along a single surface axis. A desired level of averaging can be applied to capture fewer locations on the surface on which to analyze the PSD; the overall energy theoretically remains the same while the energy of individual points increases or decreases inversely with the number of points selected.

These pressure PSDs can be normalized into surface forces which can then be applied to the structural dynamics model. The number of surface points these forces can be applied to is dependent upon desired fidelity. Typically, it is more acceptable to reduce the number of surface points when there are fewer components in the model; however, as complexity increases it is more appropriate to keep a greater number of points through which to average the pressure across the surface.

30.5 Surrogate Mechanics Model

When it comes to SOLSTICE modal analysis, simulations are conducted in the time domain with any subsequent frequency domain analyses following from the results. This is due to Simulink's inherent time-domain solving capabilities, which are ideal when solving a time-dependent ordinary differential equation (ODE) such as the mass-spring-damper equation:

$$F\,(t) = M\frac{d^2u}{dt^2} + C\frac{du}{dt} + Ku \tag{30.1}$$

This foundational equation applies to all structural dynamics analysis, from simple mass blocks and springs to full finite element modeling. The representation of this equation in Simulink can be seen in Fig. 30.2; the complexity of any model comes down to the construction of the mass and stiffness matrices, which are completely dependent upon the number of degrees of freedom, system geometry, and material properties.

While the number of elements in a mesh can simply be decreased, the model can be further reduced by way of Craig-Bampton reduction. Such a model relies upon only a limited number of primary modes of the system and removing relatively superfluous higher frequencies, thus decreasing the computational complexity when solving the resultant mass-spring ODE. The solver finds the acceleration, velocity, and position of each degree of freedom over a determined simulation time based on the timescale used to capture the pressure PSDs. Accelerations are isolated to find PSDs of each degree of freedom as desired, which can then be used to analyze the stresses at those points.
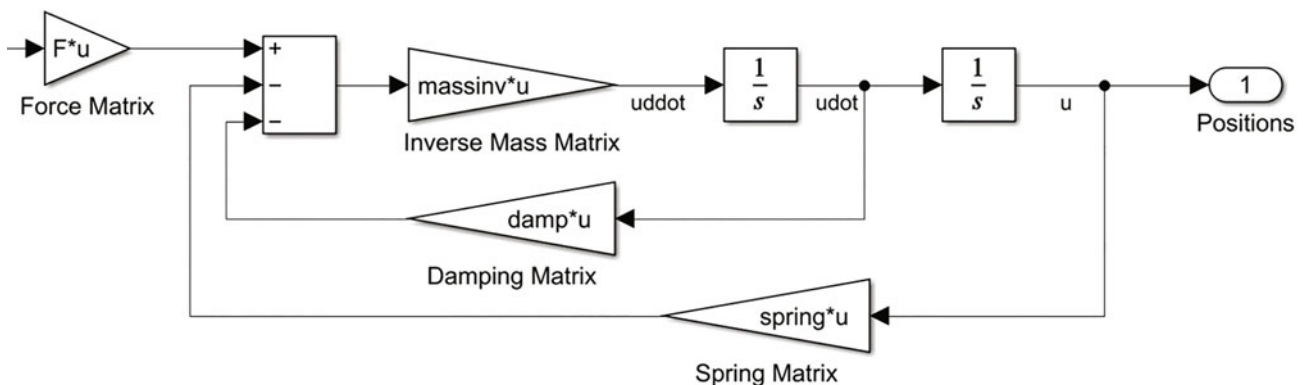
Fig. 30.2 Mass-spring-damper equation represented in Simulink

Independent of the pressure model, the natural frequencies and mode shapes of the system are found for each iteration based on mass and stiffness properties, prior to Craig-Bampton reduction. This is done by using MATLAB's built-in eigenvalue solver, *eig()*.

30.6   Conclusion

Digital engineering efforts are moving forward within Sandia National Labs and the broader engineering community; SOLSTICE is pushing ahead with rapid analysis to facilitate early development decision-making. Surrogate models are capable of targeting specific quantities of interest in order to distinguish primary system concerns and supporting higher fidelity efforts throughout the product design life cycle. The SOLSTICE team has demonstrated that surrogate models depicting various physics domains can be combined to support the evidence needed for further simulation and testing. These models present an opportunity to greatly reduce computation cost and time while also providing information to other analysis tools that can increase simulation efficiency. Each step of this workflow ultimately ties into the digital engineering life cycle, and as development continues more support can be added throughout the design process.

Chapter 31
Footbridge Vibration Predictions and Interaction with Walking Load Model Decisions

Lars Pedersen and Christian Frier

Abstract Vibrations in footbridges generated by pedestrians are a matter of concern, typically because there is the risk that vibration thresholds may exceed resulting in an unacceptable serviceability-limit-state.

There are challenges involved with predicting vibration levels at the design stage as the engineer in charge of predictions needs to make a number of choices for his calculations, for instance, regarding the load model for the pedestrians and the adjoining parameters (walking parameters). Through sensitivity studies employing artificial footbridges, the chapter will investigate the impact selected choices will have on the outcome of bridge vibration response predictions. In the chapter, a stochastic representation of the load will be considered, and hence the response calculations will end up in a stochastic representation of footbridge response.

The way to arrive at stochastic representations of bridge response will be by employing Newmark time integration and Monte Carlo simulations. The action in focus is the vertical load by pedestrians, and likewise, it will be the vertical footbridge response that is focused on.

Keywords Footbridge · Vibrations · Walking · Serviceability

Nomenclature

a	Bridge acceleration
F	Walking load
f_1	Bridge fundamental frequency
f_s	Step frequency
i	Integer
L	Bridge length
l_s	Step length
m_1	Bridge modal mass
Q	Modal load
t	Time
v	Pacing speed
W	Weight of pedestrian
Θ	Phase
Φ	Mode shape
α	Dynamic load factor
μ	Mean value
σ	Standard deviation
ζ_1	Bridge damping ratio

L. Pedersen (✉) · C. Frier
Department of the Built Environment, Aalborg University, Aalborg, Denmark
e-mail: lp@civil.aau.dk

31.1 Introduction

For footbridges, the serviceability-limit-state can be of concern. An unwanted scenario is excessive vibrations brought about by pedestrians. The Millennium Bridge vibrations [1] serve as an example, where resonant excitation caused by pedestrians resulted in excessive vibrations.

Different efforts have been devoted to describing the actions of humans whilst crossing a footbridge. Deterministic approaches were introduced in [2–4]. Later, it was recognised that it might be more appropriate to model the action of a pedestrian in a stochastic manner, modelling walking parameters as stochastic variables and to employ a stochastic framework for predicting footbridge vibrations [5–10].

Having settled on a stochastic approach for addressing the mechanisms of walking loads and recognised that a stochastic approach for studying predictions of footbridge vibrations is useful, there are yet other decisions to be made by the engineer in charge of computations.

In literature, there are different suggestions as to how to model the random nature of the different walking parameters, as will be shown in the chapter, and some of the different possible modelling approaches (in the form of the different potential settings of parameters for statistical distributions or fundamentally different relationships) will be applied in numerical simulations of pedestrian-induced vibrations in footbridges in this chapter. This is with a view to examine how different approaches affect estimates of bridge acceleration quantiles, the latter being considered a relevant parameter for assessing the serviceability-limit-state of a footbridge.

Occasionally, in a numerical simulation (modelling walking parameters as random variables and assuming mutual dependency between walking parameters), there will be scenarios where values for a walking parameter end up outside the range in which it was originally calibrated from experiments. These scenarios need to be appropriately handled in simulations. The authors of this chapter might not in all previous publications have addressed this issue thoroughly as they have been asked questions regarding this issue on different occasions. Hence, in this chapter, different assumptions as to how to possibly consider and address out-of-range properties of some walking parameters in numerical simulations are addressed and evaluated.

Section 31.2 describes the footbridges assumed for the studies of this chapter, and why a set of bridges are considered. Section 31.3 outlines the load modelling approach, different approaches for setting up a statistical framework for computing bridge acceleration response, and different approaches for handling out-of-range properties for walking parameters in numerical simulations. Section 31.4 outlines the overall methodology, and Sect. 31.5 presents the results. Conclusions are provided in Sect. 31.6.

31.2 Bridges Assumed for the Study

Assumed for this chapter are three single-span pin-supported single-degree-of-freedom (SDOF)-bridges. The general idea is to provide results for more than a single bridge so as to widen the basis for assessments and conclusions. The modal characteristics for the three bridges are provided in Table 31.1, along with adjoining assumptions regarding the bridge length, L.

The modal properties for the bridges are those valid for the bending axis in focus when considering the vertical action generated by pedestrians. The potential second, third, or higher bending modes are not considered for the studies.

Overall, the properties of the artificial footbridges, lined up in Table 31.1, are believed to be sensible for simple single-span pin-supported footbridges. For the investigations of this chapter, it is judged sensible to consider fairly simple bridges to allow the focus to be on possible implications of employing different load modelling assumptions in simulations.

Table 31.1 Modal properties of bridges (f_1, ζ_1, m_1) and bridge lengths (L)

Property	f_1	ζ_1	m_1	L
Bridge I	1.6	0.5	61.7	53.8
Bridge II	1.9	0.5	43.8	45.3
Bridge III	2.2	0.5	32.6	39.1
Unit	Hz	%	10^3 kg	m

31.3 Modelling of Walking Loads

31.3.1 Basic Load Model Assumptions

The vertical action of a pedestrian, $F(t)$, is modelled as a summation of the two contributions outlined in Eqs. (31.1) and (31.2).

$$F_i(t) = W\alpha_i \sum_{\overline{f}_j = i - 0.25}^{i + 0.25} \overline{\alpha}_i\left(\overline{f}_j\right) \cos\left(2\pi \overline{f}_j f_s t + \theta\left(\overline{f}_j\right)\right) \tag{31.1}$$

$$F_i^S(t) = W\alpha_i^S \sum_{\overline{f}_j^S = i - 0.75}^{i - 0.25} \overline{\alpha}_i^S\left(\overline{f}_j^S\right) \cos\left(2\pi \overline{f}_j^S f_s t + \theta\left(\overline{f}_j^S\right)\right) \tag{31.2}$$

In this chapter not all details about the modelling approach are explained, neither is the summation of load contributions which is somewhat more complex than explained right above. Interested readers can study [9]. The general idea is that the modelling approach accounts for the main harmonics and subharmonics of excitation to be present and that it accounts for the leakage of energy around the excitation frequencies, which some load models do not. Basically, it is considered the most refined time-domain load model of walking loads available. Hence, this model is chosen for the investigations of this chapter.

Having determined $F(t)$, the modal load $Q(t)$ can be derived using Eq. (31.3).

$$Q(t) = \Phi(t)F(t) \tag{31.3}$$

Here, $\Phi(t)$ is the mode shape function assumed for the first mode of bending action and Eqs. (31.4) and (31.5) outline the assumptions in that regard.

$$\Phi(t) = \sin\left(\pi vt/L\right) \tag{31.4}$$

$$v = f_s l_s \tag{31.5}$$

For employing Eq. (31.4), the pacing speed, v, is required and it is calculated using Eq. (31.5). In this equation, f_s is the step frequency and l_s is the step length of the pedestrian.

For the studies of this chapter, in most cases, these two parameters will be modelled as random variables, assuming Gaussian distributions. Mean values and standard deviations are shown in Table 31.2.

In Table 31.2 two different statistical distributions for f_s (both suggested in the literature) are introduced to allow examining the sensitivity of choosing one or the other distribution as the basis for computing bridge accelerations.

The static weight of the pedestrian, W, is assumed to take on a value of 750 N, hence it is considered being a deterministic property for the studies of this chapter.

Adding to the complexity of possible choices to be made by the engineer in charge of computations is that there would be another way to arrive at the modal load $Q(t)$ in that it would be possible to employ Eq. (31.6) for determining the step length of the pedestrian.

$$l_s = 0.2011 f_s^3 - 0.6021 f_s^2 + 0.6462 f_s + 0.2547 \tag{31.6}$$

This is a relationship introduced for the first time in [8] based on a graph shown in [12]. Equation (31.6) is a polynomial approximation to the graph presented in [12], which had no adjoining mathematical formula. The polynomial fit was made for f_s in the range [1.0 Hz; 2.7 Hz]. In Eq. (31.6), the value of f_s is to be inserted in the unit Hz for arriving at l_s –values in the unit meters.

Table 31.2 Mean values (μ) and standard deviations (σ)

–	μ	σ	Reference
l_s	0.71 m	0.071 m	[9]
f_s	1.87 Hz	0.186 Hz	[9]
f_s	2.20 Hz	0.300 Hz	[11]

Table 31.3 Mean values (μ) and standard deviations (σ) [7, 9]

–	α_2	α_3	α_4	α_5
μ	0.07	0.05	0.05	0.03
σ	0.030	0.020	0.020	0.015

Table 31.4 Handling of out-of-range properties for Eq. (31.7)

Approach	Description
A	Setting values of μ equal to zero when f_s is outside the range [1.0 Hz; 2.7 Hz]
B	Setting values of μ equal to the value attained at $f_s = 1.0$ Hz also for values of $f_s < 1$ Hz and setting the values of μ equal to values attained at $f_s = 2.7$ Hz also for values of $f_s > 2.7$ Hz
C	Using Eq. (31.7) also when f_s is outside the range [1.0 Hz; 2.7 Hz]

As for the first main harmonic, α_1, for use in Eq. (31.1), its mean value, μ, and standard variation, σ, may be calculated using Eq. (31.7).

$$\mu = -0.2649 f_s^3 + 1.3206 f_s^2 - 1.7597 f_s + 0.7613; \quad \sigma = 0.16\mu \qquad (31.7)$$

This is a suggestion introduced in [7] as regards μ, calibrated to measured data for values of f_s in the range [1.0 Hz; 2.7 Hz].

The load model, Eq. (31.1), also requires settling on mean values and standard deviations for other main harmonics beyond the first harmonic. Table 31.3 defines the assumptions made for the remaining main dynamic load factors assumed in the load model.

There is also a set of subharmonic dynamic load factors to be entered into Eq. (31.2). These are computed following lines of procedures outlined in [9]. Here it is just mentioned that they depend on the main harmonic load factor, α_1.

31.3.2 Out-of-Range Considerations

Having outlined the general load assumptions, the focus is on Eqs. (31.6) and (31.7). These are assumed valid for f_s in the range [1.0 Hz; 2.7 Hz]. In numerical simulations, handling f_s as a random property, there will be outcomes of f_s outside this range.

This can be handled in different ways. For the studies of this chapter, three different approaches (A, B, and C) are considered. The approaches are outlined in Table 31.4.

Table 31.4 has focused on μ for the first main harmonic load factor as function of f_s. Exact similar approaches to handling l_s as function of f_s (in Eq. (31.6)) are also considered for the studies of this chapter. The general idea is to examine how sensitive the predicted bridge acceleration response is to the choice between the three different approaches.

31.4 Methodology

Using Newmark time integration, it is possible to generate a load-time history and to compute vertical bridge acceleration response whilst a pedestrian is assumed to cross the bridge. Repeating this exercise using Monte Carlo simulations allows for obtaining a statistical representation of bridge acceleration response. From each simulated bridge crossing, the maximum acceleration at the bridge midspan, denoted a, is extracted.

This exercise is carried out for the three different bridges, for instance, for the different assumptions regarding how to handle out-of-range properties of f_s as well as for other studied scenarios. For each bridge and study condition, 100,000 bridge crossings by a single pedestrian (at each crossing assuming different values for f_s and l_s) are simulated.

From the simulation results, it is possible to extract different quantiles of a. It is believed that the higher quantiles are those most relevant for consideration as it is believed to be these that might be problematic in the matter of bridge serviceability, but the result section of this chapter (in some cases) will also present acceleration quantiles lower than a_{95}, for the sake of completeness.

31.5 Results

This section presents and evaluates results obtained for acceleration quantiles for the three different bridges. Three investigations are made (I, II, and III).

31.5.1 Investigation I

Table 31.5 shows the results obtained for a_{95} for bridge I, II, and III computed using the relationship $v = f_s l_s$ where both f_s and l_s are handled as random variables. The assumption as for the distribution of l_s is the values set out in Table 31.2 for mean value and standard deviation of this parameter. However, computations were made on the two different assumptions for the distribution of f_s also presented in Table 31.2, and Table 31.5 compares the results.

In Table 31.5, model FS1 refers to the model for f_s in which $(\mu, \sigma) = (1.87 \text{ Hz}, 0.186 \text{ Hz})$ is assumed and model FS2 to the model in which $(\mu, \sigma) = (2.20 \text{ Hz}, 0.30 \text{ Hz})$ is assumed. For all calculations, approach B for handling out-of-range conditions is employed.

It is apparent that there is a significant difference between the a_{95}-values computed for the three different bridges (for instance, compare results for FS1). This is not surprising, as the likelihood of resonant excitation is different for the three bridges. For bridge II (having a fundamental frequency of 1.9 Hz), the a_{95}-value is quite high for model FS1, which is because the mean value of f_s is relatively close to the fundamental frequency of the bridge. For bridges I and III, again looking at results for FS1, the a_{95}-values are lower, as would be expected.

A similar tendency is seen in results obtained assuming FS2, where it is apparent that the a_{95}-value peaks when the mean value of f_s equals the bridge frequency (being the case for bridge III).

Overall, the results suggest that assumptions made about the statistical representation of f_s have a relatively high impact on a_{95}-values. This is because the values of this acceleration quantile differ quite a lot assuming model FS1 or model FS2 for simulations.

31.5.2 Investigation II

Next up is addressing the influence of the choice made regarding setting up a model for arriving at values for l_s. For the investigation, there are two possible ways referred to as model LS1 and LS2. For both models, the basic assumption $v = f_s l_s$ will be used, and it will be assumed that f_s follows a Gaussian distribution which $(\mu, \sigma) = (1.87 \text{ Hz}, 0.186 \text{ Hz})$. In model LS1, it is assumed that the value of l_s is derived also assuming a Gaussian distribution with values $(\mu, \sigma) = (0.71 \text{ m}, 0.071 \text{ m})$, in accordance with Table 31.2. In model LS2, the values of l_s are computed using Eq. (31.6). Table 31.6 presents the results obtained using the two different approaches. For all calculations, approach B for handling out-of-range conditions is employed.

Table 31.5 Computed values of acceleration quantile a_{95} for two different assumptions for modelling step frequency

Bridge	I		II		III	
Bridge frequency f_1	1.60 Hz		1.90 Hz		2.20 Hz	
Model for f_s	FS1	FS2	FS1	FS2	FS1	FS2
Mean value for f_s	1.87 Hz	2.20 Hz	1.87 Hz	2.20 Hz	1.87 Hz	2.20 Hz
a_{95}	0.0932 m/s^2	0.0338 m/s^2	0.3456 m/s^2	0.2267 m/s^2	0.2728 m/s^2	0.4973 m/s^2

Table 31.6 Computed values of acceleration quantiles a_{95} for two different assumptions for modelling step length

Bridge	I		II		III	
Bridge frequency f_1	1.60 Hz		1.90 Hz		2.20 Hz	
Model for l_s	LS1	LS2	LS1	LS2	LS1	LS2
a_{95}	0.0932 m/s^2	0.0932 m/s^2	0.3456 m/s^2	0.3456 m/s^2	0.2728 m/s^2	0.2728 m/s^2

Table 31.7 Computed values of acceleration quantiles (on the basis of the FS1 assumption)

Bridge	I			II			III		
Approach	A	B	C	A	B	C	A	B	C
a_{95}	0.0909	0.0932	0.0932	0.3442	0.3456	0.3456	0.2674	0.2728	0.2728
a_{75}	0.0306	0.0308	0.0308	0.1316	0.1308	0.1308	0.0630	0.0634	0.0634
a_{50}	0.0227	0.0227	0.0227	0.0629	0.0629	0.0629	0.0364	0.0367	0.0367
Unit	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2

Table 31.8 Computed values of acceleration quantiles (on the basis of the FS2 assumption)

Bridge	I			II			III		
Approach	A	B	C	A	B	C	A	B	C
a_{95}	0.1818	0.0338	0.0338	0.4646	0.2267	0.2267	0.8082	0.4973	0.4973
a_{75}	0.0216	0.0204	0.0204	0.0705	0.0604	0.0604	0.1946	0.1562	0.1562
a_{50}	0.0171	0.0165	0.0165	0.0406	0.0379	0.0379	0.0833	0.0746	0.0746
Unit	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2	m/s^2

Studying the a_{95}-values, it is apparent that they differ for the different bridges, as would be expected. Another thing to look for is the difference between values obtained employing model LS1 and model LS2 for simulations. In fact, no difference is seen in results even when employing four digits for displaying a_{95}-values, suggesting that the choice between employing LS1 and LS2 is not important for obtaining values of a_{95}.

31.5.3 Investigation III

In this section, the focus is on the results of computed acceleration quantiles for the three different bridges (I, II, and III) assuming the three different approaches (A, B, and C) for handling out-of-range conditions for the equations for the walking parameters f_s and l_s.

First up is a presentation of results of acceleration quantiles obtained using $v = f_s l_s$ where both f_s and l_s are handled as random variables. The walking parameter f_s is modelled with a distribution in which $(\mu, \sigma) = (1.87$ Hz, 0.186 Hz$)$ is assumed (FS1) and l_s is modelled with a distribution in which $(\mu, \sigma) = (0.71$ m, 0.071 m$)$ is assumed, in accordance with Table 31.2. Table 31.7 presents the results.

As expected, the values of acceleration quantiles turn out different for the three bridges, and highest for Bridge II, explained by the fact that the fundamental frequency of this bridge is very close to the mean value assumed for f_s. The results also show that for any of the presented acceleration quantiles for a specific bridge, there appears to be a very limited difference having calculated them on the basis of approach A, B or C. It suggests that the choice of approach for handling out-of-range conditions for f_s is not of significant importance. At least in the studied case, with the mean value of f_s at 1.87 Hz and a relatively small standard deviation (0.186 Hz), the likelihood of scenarios in simulations with a pedestrian crossing using a step frequency $f_s < 1.0$ Hz or $f_s > 2.7$ Hz (the out-of-range conditions for Eq. (31.7)) is very small, at the same time explaining why almost similar values of acceleration quantiles are obtained in simulation for the three approaches (A, B, and C).

However, there are other assumptions for the statistical distribution for f_s available in the literature. There is the proposal suggested in [11], see Table 31.2. In this $(\mu, \sigma) = (2.20$ Hz, 0.30 Hz$)$ is assumed (FS2). Here the mean value of f_s is higher than assumed for the previous study and so is the standard deviation. Hence, the likelihood of reaching simulation conditions where $f_s > 2.7$ Hz is higher than in the previous study.

Table 31.8 presents the results for acceleration quantiles for the three different bridges, again using $v = f_s l_s$ and the model for l_s is yet again based on the assumption $(\mu, \sigma) = (0.71$ m, 0.071 m$)$.

It comes as no surprise that on these assumptions for computation of acceleration quantiles, it is bridge III that experiences the highest acceleration levels as bridge III has a fundamental frequency in fact equal to the mean value assumed for f_s (2.20 Hz). The interesting part is to examine the difference between outcomes of acceleration quantiles computed for the three different approaches (A, B, and C) employed for handling outcomes of f_s outside the range [1.0 Hz; 2.70 Hz]. It is computed that the probability of obtaining values of $f_s > 2.70$ Hz in simulations is at 4.78% when assuming $(\mu, \sigma) = (2.20$ Hz, 0.30 Hz$)$ for f_s as done for these simulations whereas it was only at 0.0004% when assuming $(\mu, \sigma) = (1.87$ Hz, 0.186 Hz$)$.

For all three bridges, it is seen that there is no difference between the acceleration quantiles computed employing approaches B and C. Furthermore, it is seen that the acceleration quantiles computed employing approach A differ rather significantly from those computed on the assumption of approaches B and C, at least for the a_{95}-estimates. Thinking it through, this is not all that surprising as in approach A for every outcome of $f_s > 2.70$ Hz in simulations, a figure of 0 enters for the first main harmonic load factor, α_1. This basically has the effect that the pedestrian force is set to zero for this pedestrian crossing, which cannot be meaningful, and consequently, the computed distribution function for bridge acceleration is messed up, rendering the acceleration quantiles not meaningful. However, it is comforting to find that approaches B and C result in close to identical estimates of the acceleration quantiles and that approach B has been used by the authors of this chapter in previous studies published in the literature.

Other simulation runs were made studying the implications of choosing approach A, B, or C for handling out-of-range properties for l_s when using the relationship outlined in Eq. (31.6) for l_s. Results are not shown here but they generally confirm that there are no differences in estimates of acceleration quantiles between choosing approach B or C for these simulations. A fun fact is that it turned out that using approach A resulted in infinitely long simulation runs. This was and is because setting values of l_s to zero (and this will happen during simulations using approach A), the value of v using $v = f_s l_s$ will be set to zero, resulting in an infinitely long time for the pedestrian to cross the bridge. Hence, these simulation runs are not meaningful.

31.6 Conclusion and Discussion

The chapter has addressed the implications of decision-making for the engineer in charge of predicting bridge accelerations using a probability-based approach for modelling the action. This is by doing simulation studies.

The results suggest that acceleration quantiles of bridge response are sensitive to choices made for the statistical distribution representing the step frequency of walking, f_s. Different ways to modelling the walking parameter l_s representing the step length of a pedestrian were also studied. It turned out so that the manner in which this parameter (l_s) was modelled did not have much bearing on the outcome of simulation results for bridge acceleration quantiles.

Another matter of interest for the studies of this chapter was to examine how sensitive simulation results in terms of bridge acceleration quantiles would be to different ways of handling out-of-range conditions for walking parameters in numerical simulations. As mentioned in the introduction, questions have been asked and the chapter has provided results.

The results suggest that basically either approach B or approach C studied in this chapter can be employed without inflicting differences in estimates of bridge acceleration quantiles, at least for the SDOF bridges assumed for the studies of this chapter. Approach A would be the approach not to choose, for the reasons described in Sect. 31.5.

For the study of possible out-of-range study conditions for the dynamic load factor, it should be mentioned that it encompassed an assumption where the probability of arriving at out-of-range conditions is relatively high (by employing model FS2). It is worth mentioning that this is the only model for the step frequency that the authors of this chapter have encountered that translates into out-of-range probabilities for the main dynamic load factor that are that high. The data basis and sample size for calibrating this model are not very-well described in [11]. However, as it is a model available in the literature, it was considered relevant to address it for the context of the studies of this chapter.

References

1. Dallard, P., Fitzpatrick, A.J., Flint, A., Le Bourva, S., Low, A., Ridsdill-Smith, R.M., Wilford, M.: The London Millennium Bridge. Struct. Eng. **79**, 17–33 (2001)
2. Ellis, B.R.: On the response of long-span floors to walking loads generated by individuals and crowds. Struct. Eng. **78**, 1–25 (2000)
3. Bachmann, H., Ammann, W.: Vibrations in Structures – Induced by Man and Machines IABSE Structural Engineering Documents 3e. IABSE, Zürich (1987)
4. Rainer, J.H., Pernica, G., Allen, D.E.: Dynamic loading and response of footbridges. Can. J. Civ. Eng. **15**, 66–78 (1998)
5. Matsumoto, Y., Nishioka, T., Shiojiri, H., Matsuzaki, K.: Dynamic design of footbridges. In: IABSE Proceedings, No. P-17/78, pp. 1–15 (1978)
6. Živanovic, S.: Probability-based estimation of vibration for pedestrian structures due to walking. PhD Thesis, Department of Civil and Structural Engineering, University of Sheffield, UK (2006)
7. Kerr, S.C., Bishop, N.W.M.: Human induced loading on flexible staircases. Eng. Struct. **23**, 37–45 (2001)
8. Pedersen, L., Frier, C.: Sensitivity of footbridge vibrations to stochastic walking parameters. J. Sound Vib. **329**, 2683–2701 (2009). https://doi.org/10.1016/j.jsv.2009.12.022

9. Živanovic, S., Pavic, A., Reynolds, P.: Probability-based prediction of multi-mode vibration response to walking excitation. Eng. Struct. **29**, 942–954 (2007). https://doi.org/10.1016/j.engstruct.2006.07.004
10. Pedersen, L., Frier, C.: Predictions of footbridge vibrations and influencing load model decisions. In: Pakzad, S. (ed.) Dynamics of Civil Structures Proceedings of the 38th IMAC, 2020, vol. 2, pp. 151–157. Springer (2021)
11. Kramer, H., Kebe, H.W.: Man-induced structural vibrations. Der Bauingenieur. **54**, 195–199 (1979)
12. Wheeler, J.E.: Prediction and control of pedestrian induced vibration in footbridges. J. Struct. Div. **108**, 2045–2065 (1982)

Chapter 32
Assembling Uncertainty Effects on the Dynamic Response of Nominally Identical Motorbike Components

Elvio Bonisoli, Luca Dimauro, Simone Venturini, and Lorenzo Peroni

Abstract The chassis and swingarm are the main components of the motorbike frame. The dynamic response of these components strongly influences the frame flexibility and consequently the motorbike dynamics. However, there may be variability in nominally identical manufactured components. The uncertainty may arise from many sources including geometric tolerances, material properties, and variability in the manufacturing and assembling process, for example, adhesive bonding of hollow parts. The presence of uncertainties can significantly alter motorbike component dynamic response and modal properties, and thus their overall performance during a racing competition. Therefore, competitive riders test several components during the racing weekend to find the specific motorbike frame with which they are more comfortable.

In this chapter, experimental modal analyses have been carried out on the flexible components of a motorbike frame. The experimental campaign results have demonstrated significative differences in frequency response functions, natural frequencies and damping of motorbike components. Modal assurance criterion and other indexes have been used to compare mode shapes of the seemingly identical components and to assess possible crossing and veering phenomena, due to uncertainty.

Keywords Motorbike dynamics · Experimental modal analysis · Model uncertainty · Motorbike component design · FRF uncertainty

32.1 Introduction

In two-wheeler vehicle applications, handling and dynamic behaviour assume a more important role than in the well-known automotive environment, since motorcycles are inherently unstable due to their structure, hence roll and steer angles must be considered in the dynamic design to avoid troublesome consequences for the motorbike driveability and stability [1]. Relevant effects on motorbike dynamics [2] take place when motorbikes are cornering at large roll angles because the motion of the tyres [3] is no longer in plane with the suspension system. During this phase, the main motorbike components, like the chassis and swingarm, significantly affect the frame flexibility and consequently the global dynamic response of the system. Meanwhile, the joints connecting different motorbike components, such as the rear shock absorber, have a key role since their stiffness strongly influences motorbike stability, together with the durability and integrity of wheel subsystem [4] and tyres. Currently, the dynamic performance of a motorbike frame is evaluated during the pre-season tests or during the racing weekend tests, during which the rider runs with different release of the same components to have his feedback and to get the specific motorbike frame he is more comfortable with.

In the last decades, research on motorcycles has been based on the development of mathematical models with rigid bodies to analyse their dynamic behaviour, considering only the stiffness and damping of the connections [1, 5, 6], while recently, multibody models of motorcycles [7, 8] were improved, introducing also the component flexibility [9], for example, front [10] and rear swingarm [11] and chassis [12], to simulate the frame dynamic behaviour and the component effect on global dynamics. Moreover, it was highlighted in [13] that during high roll angle cornering the torsional stiffness given by the frame, flexibility works in series with the front fork: the front fork absorbs only the road irregularity component directed along the

---

E. Bonisoli · L. Dimauro · S. Venturini (✉) · L. Peroni
Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Torino, Italy
e-mail: simone.venturini@polito.it

shocks axis while the chassis absorbs the other components, working as an equivalent torsion spring. Hence chassis flexibility is very relevant for comfort and to guarantee adherence in the passages between low and high friction surfaces; at the same time, it acts as a skeleton and must be strong enough to support the engine and gearbox loads. In the past, several numerical and experimental modal analyses [14] were performed on motorcycle components with different purposes: in [15, 16] two models of two-wheeler chassis were investigated to reduce undesired vibrations and obtain a satisfying weight reduction improving material. Results of the modal analysis on components, subassemblies and motorbike assembly were used by the authors in [17, 18] to perform a components-to-assembly mode shapes tracing, evaluating the participation of every single component in the assembled multibody motorbike mode shapes. The tracing process is based on the concept of Dynamics for Modal Assembly (DMA) [19, 20], which is a tool to assess the influence of each component mode shape in the dynamics of a parametric assembled system changing the assembly adjustable parameters. The method has been adapted for the detection of component critical mode shapes for the high-frequency assembly vibrations and applied to the test case of a motorcycle frame [21] to have a method, based on an assembly-to-components modal approach, able to detect the component mode shapes responsible for undesired dynamics in the assembled multibody systems. In addition to this, the procedure reveals the critical components and their mode-shapes requiring modifications, offering a quick optimisation method for the optimal design in the field of vehicle for a racing competition where continuous developments and improvements are required to adapt the motorbike to the last innovations, fulfilling new competition rules.

In this challenging environment, motorbike components can be affected by uncertainty due to several factors like geometric tolerances, material properties, and variability in the manufacturing and assembling process, as analysed in [22] for a bike spoke wheel. The presence of uncertainties can become an issue when it could significantly alter motorbike component dynamic response and modal properties, and thus their overall performance during a racing competition. Different methodologies have been introduced in years; eventually, the eigenvalue sensitivity technique has been successfully used in [23, 24] to evaluate the stress-stiffening effect in aluminium frames. In other words, uncertainty can affect the modal properties of the chassis and swingarm, changing their natural frequencies.

The chapter is organised as follows: in Sect. 32.2, the approached case study is presented and the peculiarities of the structure in the analysis are described. In Sect. 32.3, the developed experimental setup is defined and details about excited nodes and geometry are supplied. In Sect. 32.4, the results of the experimental modal analysis are discussed, and several indexes are exploited to obtain a better understanding of the variability in identified dynamic behaviour. Finally, in Sect. 32.5, remarks about the performed activity and considerations on future developments are depicted.

32.2 Case Study

The motorbike studied in this research activity, reported in Fig. 32.1, is developed and produced by the student team 2WheelsPoliTO for the PreMoto3 category of the Italian Speed Championship (CIV) of 2022.

Fig. 32.1 2WheelsPoliTO motorbike

In this chapter, only chassis and swingarm, the two main structural components, are studied (Fig. 32.2); both are aluminium alloy modular structures, composed of tubular and boxed profiles (grey components) linked together by casted supports (blue components) through adhesive bonding. Moreover, the cavities in casted supports and between profiles are stuffed with closed cell aluminium foam (Fig. 32.3) [25] (green components). Additionally, aluminium plates (red components) are glued to protect the stuffed cavities.
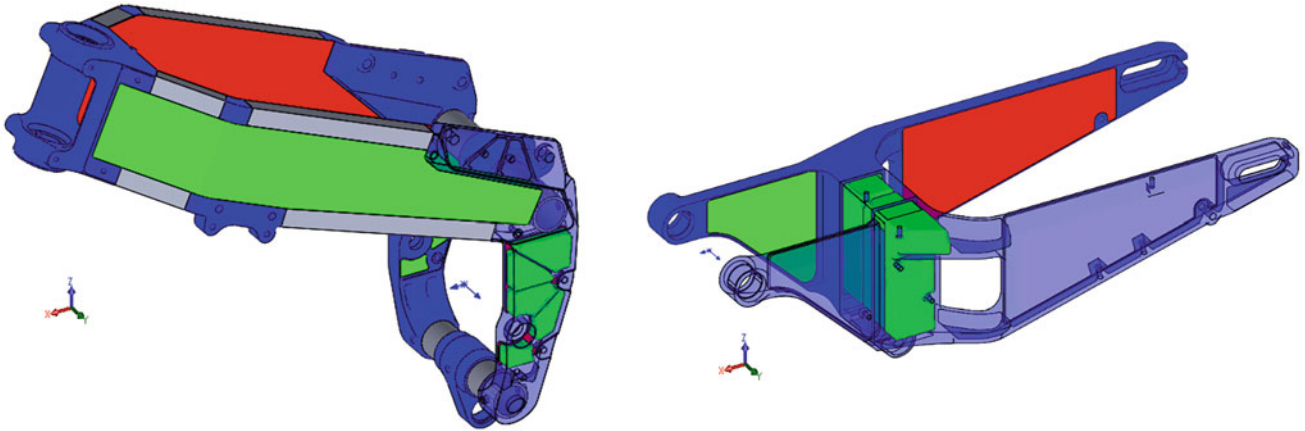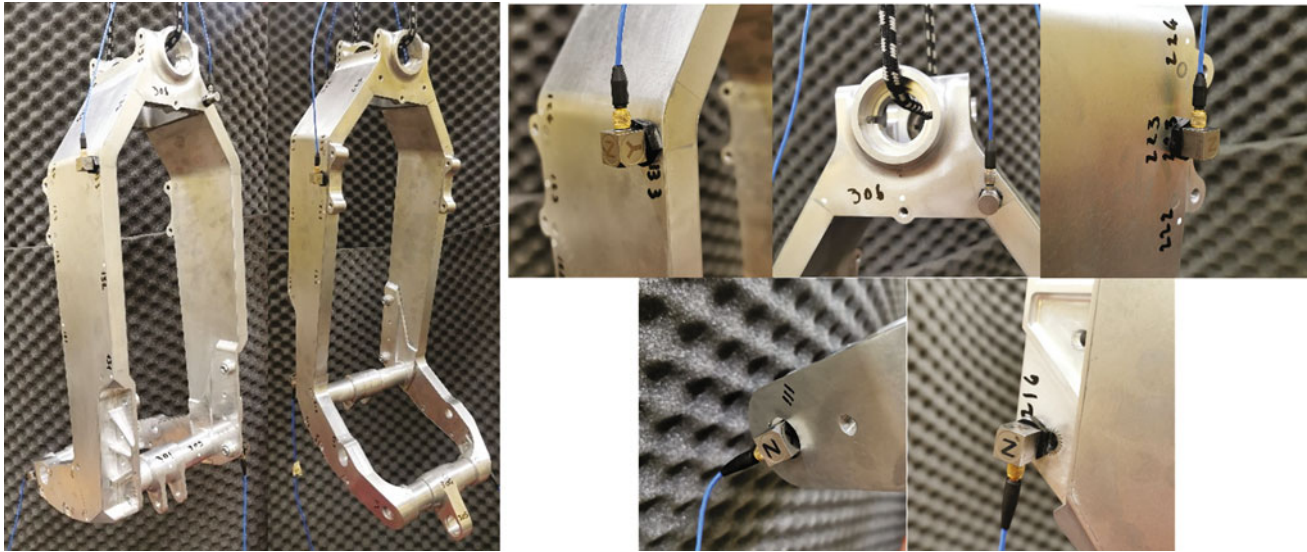

Fig. 32.2 Motorbike component architectures: chassis (*left*) and swingarm (*right*)

Fig. 32.3 Aluminium closed cell foam sample

Table 32.1 Component mass
properties

Component	Release	Weight (kg)
Chassis	C1	5.799
	C2	5.885
Swingarm	S1	3.442
	S2	3.409

Fig. 32.4 Overview of chassis experimental setup

The metal foam density roughly varies between 5% and 25% of the corresponding homogeneous material density. In this application, a density of 280 kg/m^3 is estimated. Furthermore, component masses are strongly affected by actual aluminium foam density and the amount of adhesive agents, both affected by uncertainties. Therefore, the masses of the tested components reported in Table 32.1 show wide variability.

32.3 Experimental Setup

The components are studied in free-free conditions, hanging the chassis and the swingarm by means of an elastic rubber band structure (Figs. 32.4 and 32.5), and exciting through a roving hammer. Each component is excited with an impact hammer PCB 086C03 using a hard plastic tip, and its response is measured using four tri-axial accelerometers PCB 356A15 and one mono-axial accelerometer PCB 352C33. The points, which are chosen to obtain a good compromise between geometry representation and expected modal behaviour, according to the MoGeSeC technique [26], are equally distributed between the left and right sides.

The two chassis are excited in 40 nodes along all the possible directions allowed by its geometry, obtaining an overall 91 excitations to the system. Few other points were selected on auxiliary components, such as linking shafts and suspension attachment. Chassis reference geometry for the EMA tests is shown in Fig. 32.6 (left). On the other hand, the swingarm was excited on 37 nodes with an overall of 86 excitations to the system. As done for the chassis, nodes were selected approximating the real geometry in the best possible way, obtaining the swingarm geometry for EMA tests reported in Fig. 32.6 (right).

The excitation and response signals are acquired using a 16 channels Siemens LMS SCADAS Mobile acquisition board. Acquisitions are performed using LMS Test.Lab software: time histories of 4 s (frequency resolution of 0.25 Hz) are acquired at a sampling frequency of 8192 Hz; the force-exponential window is used for the hammer input signal over 0.6% of samples, while an exponential window with a decay of 30% is applied to accelerometer output signal; the H_1 estimation of the system inertance Frequency Response Functions (FRFs) are linearly averaged on five repetitions [14].
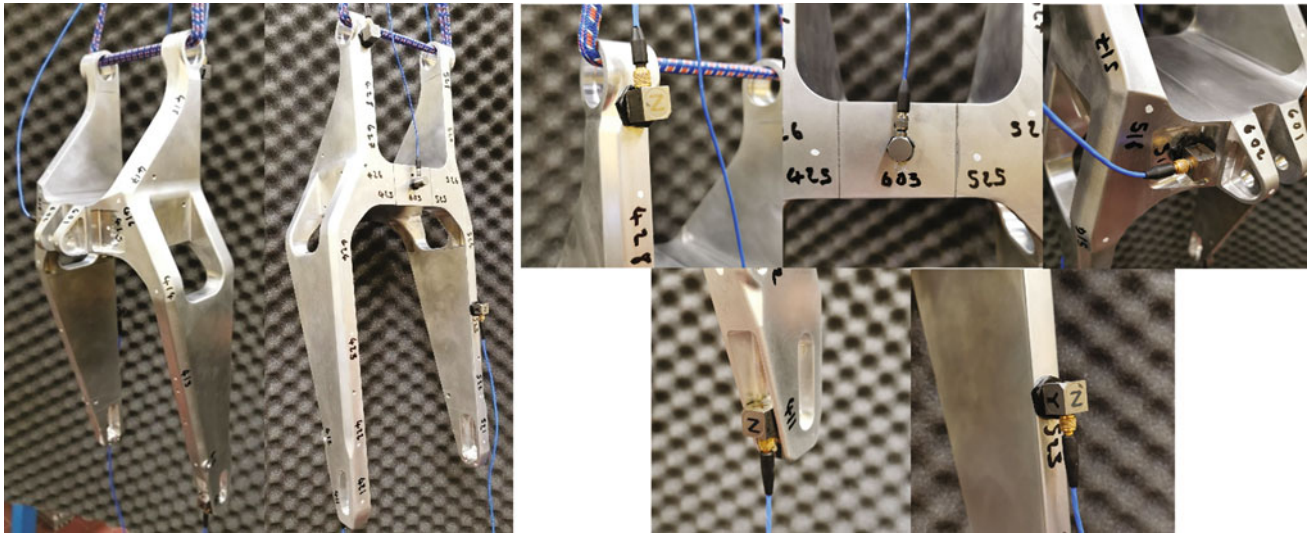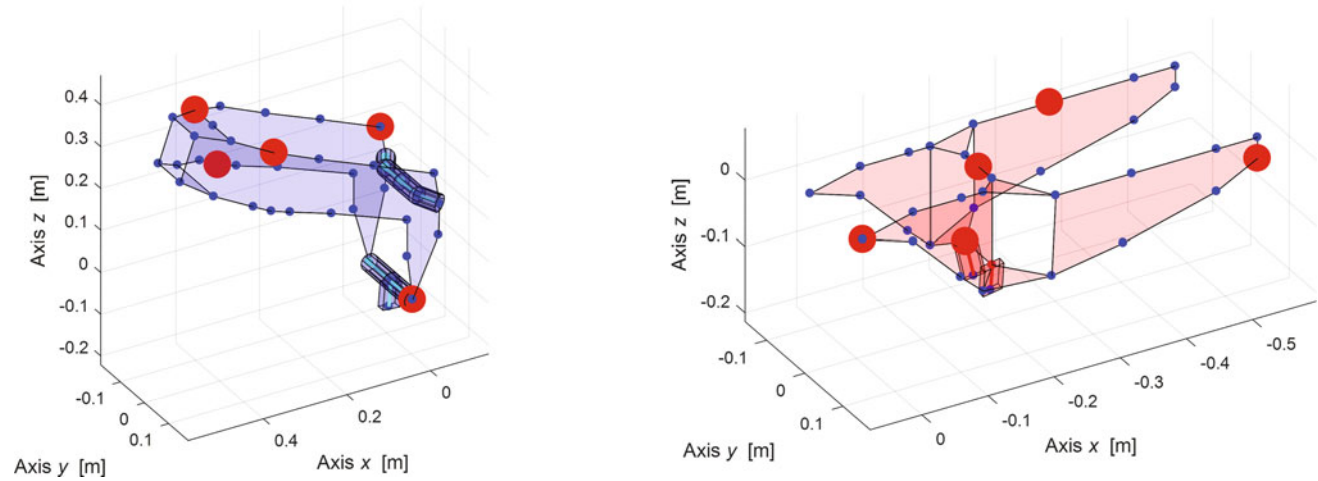
Fig. 32.5 Overview of swingarm experimental setup

Fig. 32.6 Reference geometry for experimental tests: chassis (*left*) and swingarm (*right*)

32.4   Experimental Modal Analysis

The Experimental Modal Analysis (EMA) has been carried out on two seemingly identical releases of the two analysed components. Modal properties of each system, namely, natural frequencies ω_r, damping ratios ζ_r and complex mode-shapes Ψ_r [27], are identified from the system FRFs using PolyMAX algorithm [28] in the frequency range $0 \div 2000$ Hz for chassis and $0 \div 1500$ Hz for swingarm. In Fig. 32.7 the comparison of FRF sum is shown: the nomenclature refers to Table 32.1 definition of component releases which is arbitrarily based on the order of experimental tests. Perhaps, C1 and S2 are the first chassis and swingarm subject to EMA. Figure 32.7 shows the reduced modal density in the bandwidth of interest: in both cases, the FRF sums show peaks at low frequencies corresponding to rigid body modes. Instead, at high frequencies, the behaviours of chassis and swingarm strongly differ even if they are nominally identical components.

Based on the results of PolyMAX identification, Modal Phase Collinearity (MPC) index [29, 30] is applied. MPC is a modal indicator that gives the idea of whether a mode shape used in its computation is a real or complex-valued vector. Its estimate inherits the statistical properties of the corresponding mode shape estimate giving as a result an estimation of the phase of the degree of freedom (DoF). An MPC value equal to 100% means the mode shape Ψ_j is real, while MPC equal to 0% means the Ψ_j mode shape is complex.
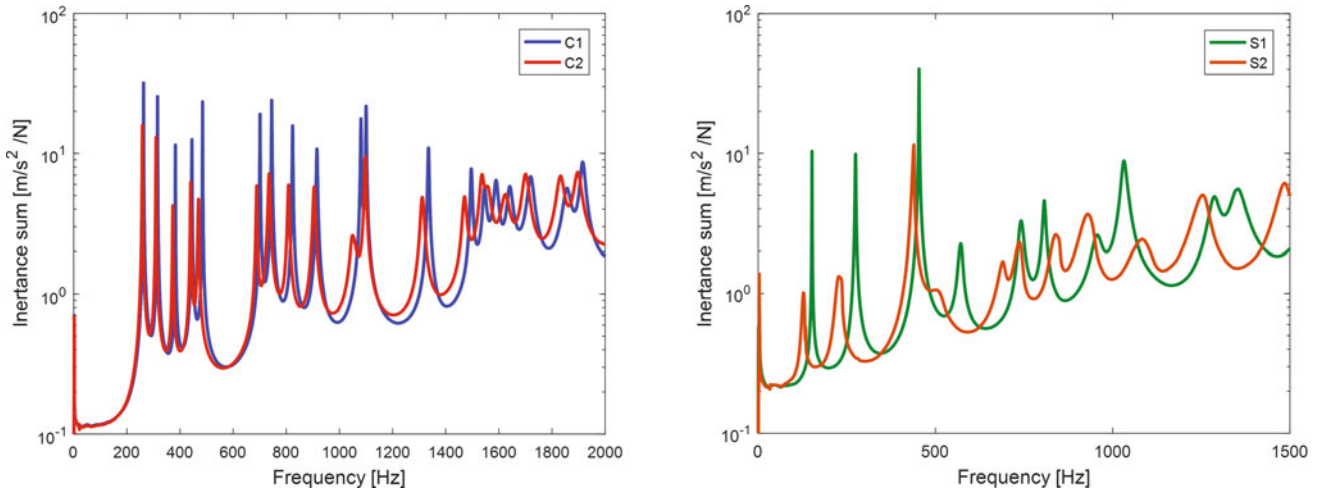
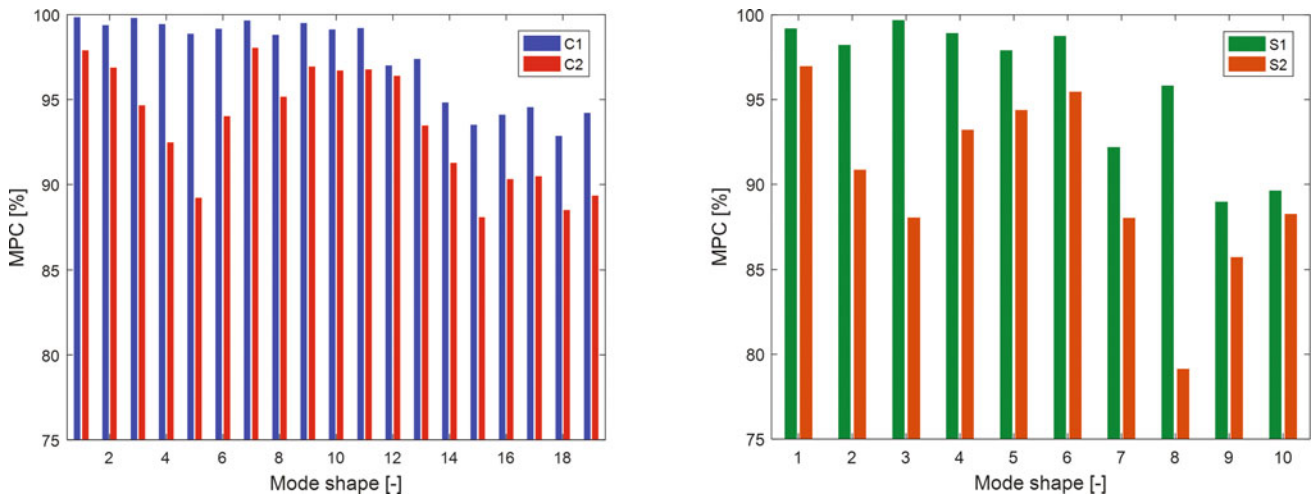**Fig. 32.7** FRF sum comparison: chassis (*left*) and swingarm (*right*)

Fig. 32.8 MPC comparison on component releases: chassis (*left*) and swingarm (*right*)

$$\mathrm{MPC} = \mathrm{MAC}_{j,j*} = \frac{\left|\boldsymbol{\Psi}_j^{\mathrm{H}}\boldsymbol{\Psi}_j^*\right|^2}{\left[\boldsymbol{\Psi}_j^{\mathrm{H}}\boldsymbol{\Psi}_j\right]\left[\boldsymbol{\Psi}_{j*}^{\mathrm{H}}\boldsymbol{\Psi}_{j*}\right]} = \frac{\left|\boldsymbol{\Psi}_j^{\mathrm{T}}\boldsymbol{\Psi}_j^*\right|^2}{\left[\boldsymbol{\Psi}_j^{\mathrm{H}}\boldsymbol{\Psi}_j\right]^2} = \cos^2\alpha_{j,j*} \qquad (32.1)$$

The results of the MPC index application on the EMA identification are shown in Fig. 32.8: chassis mode shapes are preponderantly real, that is, the minimum MPC value is above 85%, hence a real identification could be sufficient to describe the modal behaviour of the structure. Instead, swingarm mode shapes tend to be more complex and S2 component has lower collinearity. Since the variation of MPC is high between the two component releases, the effectiveness of real identification is not guaranteed in this case study.

Furthermore, Figs. 32.9 and 32.10 show a direct comparison of identified natural frequencies ω_r and damping ratios ζ_r. Figure 32.9 shows a systematic difference between the first and second release of the components: while the reduction in chassis natural frequencies between C1 and C2 could be justified by a reduction in mass, the same trend could not be confirmed on swingarm components. Moreover, second releases of both components show a strong increase in the damping ratio. In Fig. 32.10, the chassis damping ratio increases up to 1.3% and generally is more than twice; instead, the swingarm damping ratio increases up to four times on the fourth mode shape.

To carry out a comparison of the correlation of the mode shapes between component releases, the MACX (compleX Modal Assurance Criterion) [29] is adopted, allowing to measure the correlation level between two complex mode shapes $\boldsymbol{\Psi}_j$ and $\boldsymbol{\Psi}_k$, respectively. The MACX is defined as follows:
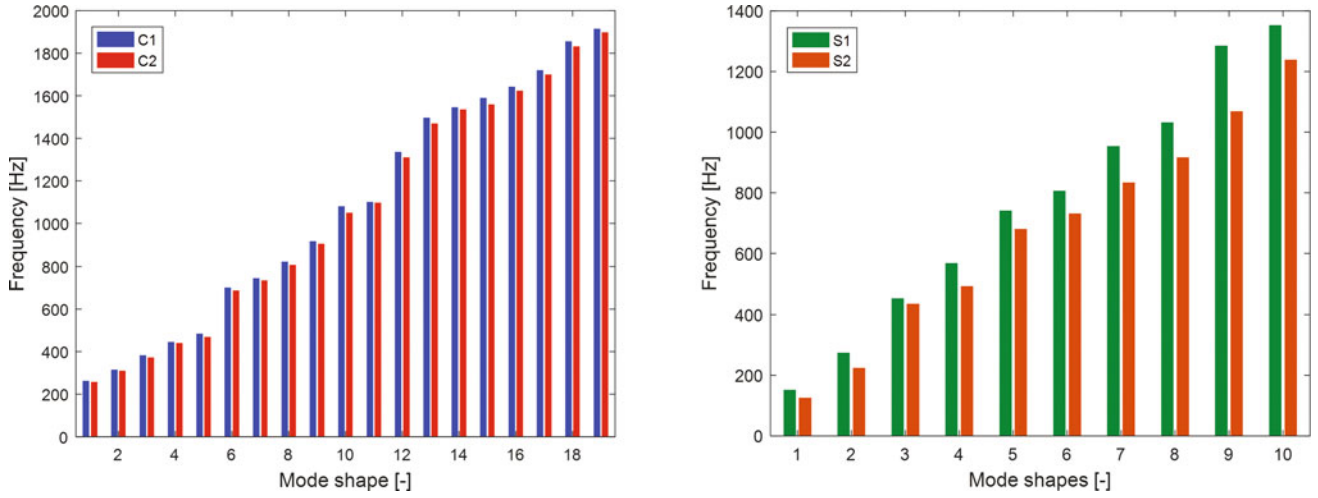
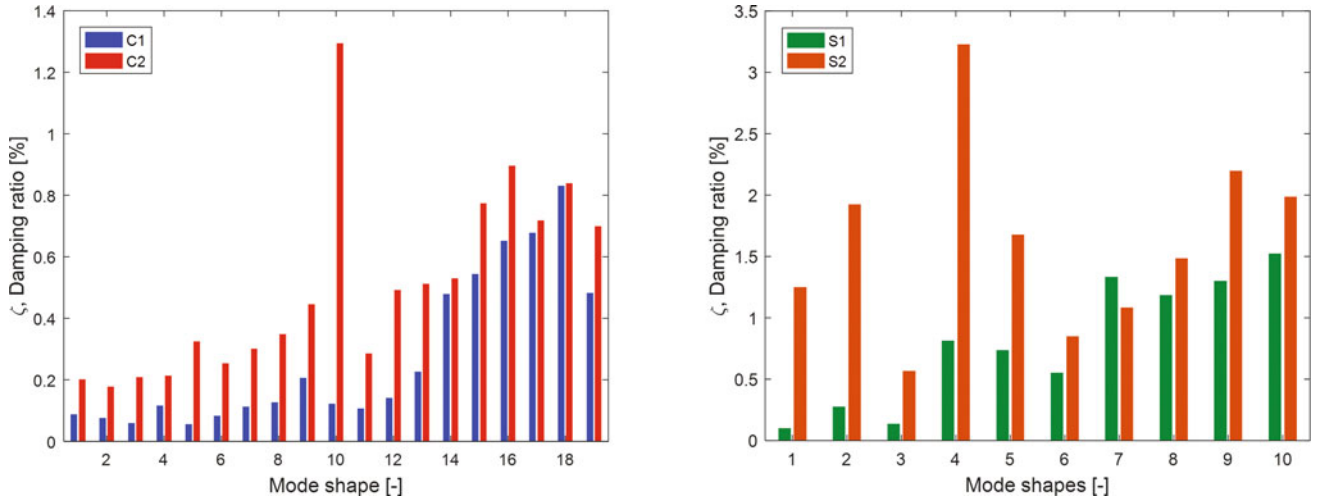**Fig. 32.9** Frequency comparison: chassis (*left*) and swingarm (*right*)

Fig. 32.10 Damping ratio comparison: chassis (*left*) and swingarm (*right*)

$$\text{MACX}_{j,k} = \frac{\left[\left|\boldsymbol{\Psi}_j^H \boldsymbol{\Psi}_k\right| + \left|\boldsymbol{\Psi}_j^T \boldsymbol{\Psi}_k\right|\right]^2}{\left[\boldsymbol{\Psi}_j^H \boldsymbol{\Psi}_j + \left|\boldsymbol{\Psi}_j^T \boldsymbol{\Psi}_j\right|\right]\left[\boldsymbol{\Psi}_k^H \boldsymbol{\Psi}_k + \left|\boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k\right|\right]} \tag{32.2}$$

where its value is always real and semi-positive, independently if the eigenvectors are real or complex; its value ranges from 0, corresponding to very far mode shapes or orthogonal, to 1, corresponding to equal or similar (scaled) shapes. The MACX application is shown in Fig. 32.11: in both cases, the correlation between component releases is optimal and significant interactions of modes cannot be identified. On chassis, the correlation is optimal up to 1300 Hz at which MACX starts to deprecate, and uncertainties on the structural parameters begin to act at this frequency, suggesting material structural properties could be affected by variability. Instead, swingarm suffers strong deviation in S2 frequencies which are basically lower than S1 even if the mode shape correlation is quite strict also at high frequencies.

This slight deviation in natural frequencies can be caused by uncertainty in the aluminium foam density and its usage fulfilling structure cavities. Tables 32.2 and 32.3 supply details of the experimentally identified natural frequencies and damping ratios of the component releases with description of the mode shapes, which are represented in Fig. 32.12 obtained in Matlab-based finite element environment LUPOS [31].
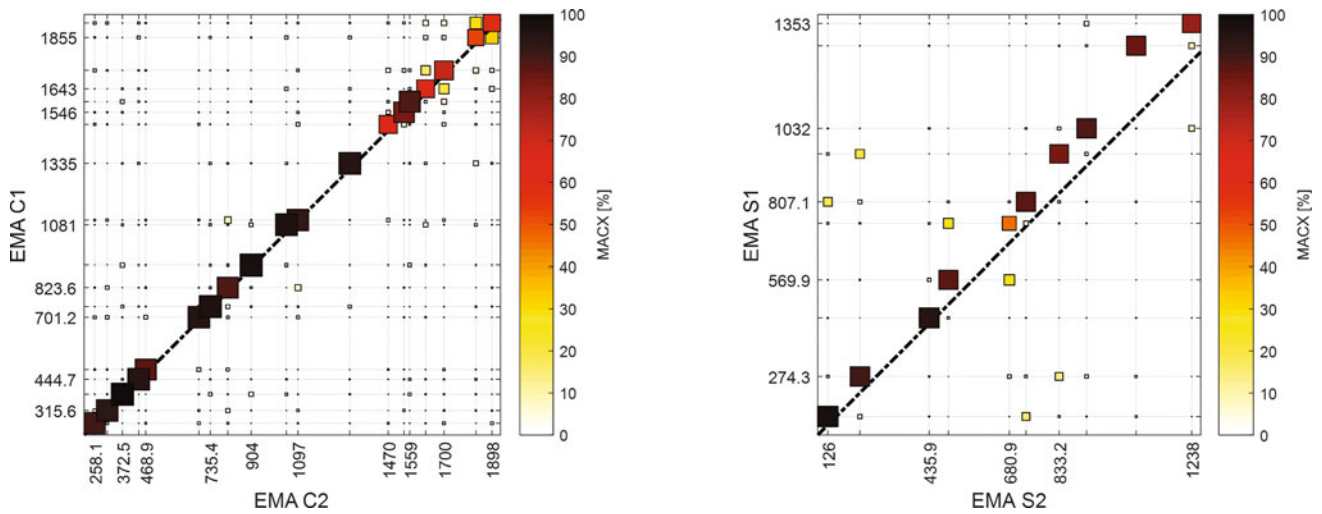
Fig. 32.11 MACX comparison: chassis (*left*) and swingarm (*right*)

Table 32.2 Experimental natural frequencies and damping ratios on chassis

Mode	Natural frequencies (Hz)		ζ, damping ratio (%)		Description
	C1	C2	C1	C2	
1	262.9	258.2	0.088	0.202	First torsional RY
2	315.6	309.6	0.076	0.177	First bending XY in-phase
3	382.8	372.5	0.059	0.208	First bending XY out-of-phase
4	444.7	439.7	0.117	0.213	First bending XZ in-phase
5	485	468.9	0.056	0.324	First shear XZ in-phase
6	701.2	688.0	0.083	0.254	Second bending XY in-phase
7	745.3	735.4	0.112	0.301	Torsional RX out-of-phase
8	823.6	807.7	0.127	0.349	Torsional RX in-phase
9	915.4	904.0	0.207	0.446	Second bending XZ in-phase

Table 32.3 Experimental natural frequencies and damping ratios on swingarm

Mode	Natural frequencies (Hz)		ζ, damping ratio (%)		Description
	S1	S2	S1	S2	
1	151.7	126.04	0.099	1.249	First bending XY out-of-phase
2	274.3	224.11	0.275	1.923	First bending XY in-phase
3	453.6	435.86	0.137	0.568	Torsional RY (first shear) out-of-phase
4	569.9	494.68	0.814	3.229	First local bending XY out-of-phase
5	741.3	680.99	0.736	1.677	First local bending XY in-phase
6	807.1	731.91	0.551	0.850	Second bending XY out-of-phase
7	953.9	833.25	1.334	1.084	Second bending XY in-phase
8	1032	917.05	1.184	1.483	Torsional RX out-of-phase
9	1285	1068.85	1.299	2.198	Torsional RX in-phase
10	1353	1238.57	1.523	1.985	First bending XZ in-phase

32.5 Conclusion

The proposed case study allows to evaluate the consequence of uncertainties on nominally identical motorbike components. The performed analyses show systematic differences in dynamic behaviour between releases of the same parts which are preliminarily attributed to uncertainty in aluminium foam material properties well documented in the state-of-the-art. Meanwhile, the difference at high frequencies on the chassis as the generalised difference in damping ratio leaves some open questions on the history of loading conditions the components were eventually subject to. Nevertheless, the strong
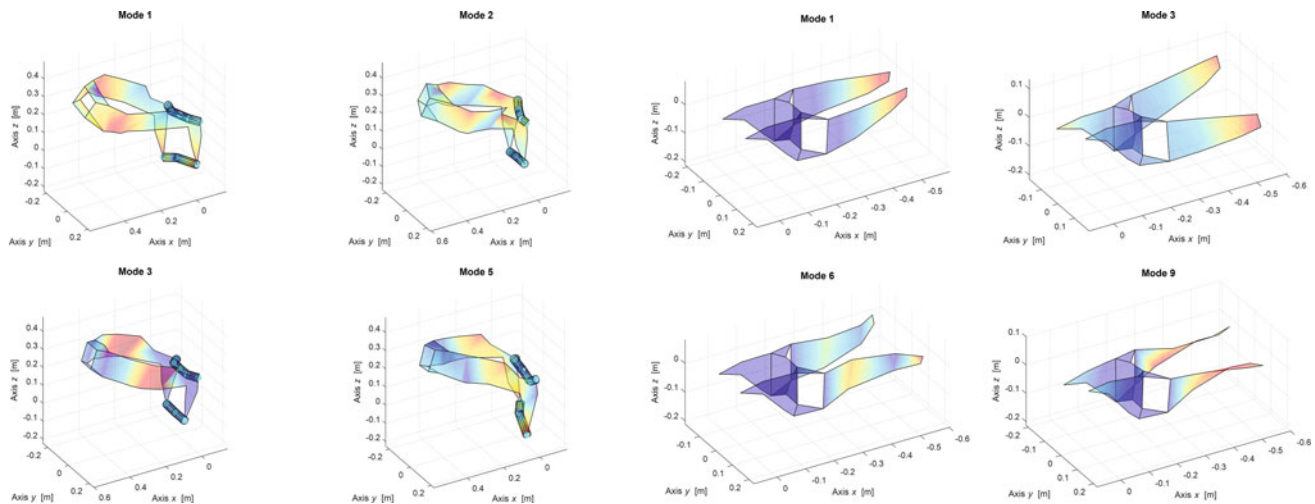
Fig. 32.12 Main mode shapes for chassis (*left*) and swingarm (*right*)

variation in dynamic behaviour on nominally identical motorbike components motivates a deep investigation of the causes of uncertainties for which a finite element model will be developed to discover the sources of variability.

References

1. Sharp, R.S., Limebeer, D.J.: A motorcycle model for stability and control analysis. Multibody Sys. Dyn. **6**, 123–142 (2001). https://doi.org/10.1023/A:1017508214101
2. Cossalter, V.: Motorcycle Dynamics. Lulu, Morrisville (2006)
3. Doria, A., Taraborrelli, L., Urbani, M.: A modal approach for the study of the transient behavior of motorcycle and scooter tires. In: Proceedings of ASME International Design Engineering Technical Conferences, August 17–20, 2014, Buffalo, USA, pp. 29–37 (2014). https://doi.org/10.1115/DETC2014-34023
4. Rovarino, D., Actis Comino, L., Bonisoli, E., Rosso, C., Venturini, S., Velardocchia, M., Baecker, M., Gallrein, A.: A methodology for automotive steel wheel life assessment. In: SAE 2020 World Congress, April 21–23, 2020, Detroit, Michigan, SAE Technical Paper 2020-01-1240, pp. 1–10 (2020). https://doi.org/10.4271/2020-01-1240
5. Cossalter, V., Lot, R.: A motorcycle multi-body model for real time simulations based on the natural coordinates approach. Veh. Syst. Dyn. **37**(6), 423–447 (2002). https://doi.org/10.1076/vesd.37.6.423.3523
6. Sharp, R.S., Evangelou, S., Limebeer, D.J.: Advances in the modelling of motorcycle dynamics. Multibody Sys. Dyn. **12**, 251–283 (2004). https://doi.org/10.1023/B:MUBO.0000049195.60868.a2
7. Sequenzia, G., Oliveri, S., Fatuzzo, G., Calì, M.: An advanced multibody model for evaluating rider's influence on motorcycle dynamics. Proc. Inst. Mech. Eng. K J. Multi-body Dyn. **229**(2), 193–207 (2015). https://doi.org/10.1177/1464419314557686
8. Barbagallo, R., Sequenzia, G., Oliveri, S., Cammarata, A.: Dynamics of a high-performance motorcycle by an advanced multibody/control co-simulation. Proc. Inst. Mech. Eng. K J. Multi-body Dyn. **230**(2), 207–221 (2016). https://doi.org/10.1177/1464419315602825
9. Olivieri, S.M., Calì, M., Catalano, L.: Dynamics of motorcycle using flexible elements. In: Proceedings of International Design Conference, Dubrovnik, Croatia, 2002, pp. 1227–1236 (2002)
10. Cossalter, V., Doria, A., Massaro, M., Taraborrelli, L.: Experimental and numerical investigation on the motorcycle front frame flexibility and its effect on stability. Mech. Syst. Signal Process. **60**, 452–471 (2015). https://doi.org/10.1016/j.ymssp.2015.02.011
11. Dixit, V.S., Nukulwar, M.R., Shinde, S.T., Pimpale, S.S.: Vibration response and optimization of swing arm through hardening. Int. J. Curr. Eng. Technol. **6**, 562–567 (2016)
12. Lake, K., Thomas, R., Williams, O.: The influence of compliant chassis components on motorcycle dynamics: an historical overview and the potential future impact of carbon fibre. Int. J. Veh. Mech. Mobil. **50**(7), 1043–1052 (2012). https://doi.org/10.1080/00423114.2011.647824
13. Bocciolone, M., Cheli, F., Pezzola, M., Viganò, R.: Static and dynamic properties of a motorcycle frame: experimental and numerical approach. WIT Trans. Model. Simul. **41**, 517–526 (2005)
14. Maia, N.M.M., Silva, J.M.M.: Theoretical and Experimental Modal Analysis. Wiley, New York (1997)
15. Harshal, B.B.: Study of experimental modal analysis on two wheeler frame. Int. J. Eng. Sci. Res. Technol. **4**(7), 2518–2520 (2015)
16. Hiremath, S., Kumar, N., Nagareddy, G., Rathod, L.: Modal analysis of two wheeler chassis. Int. J. Eng. Sci. Res. Technol. **5**(7), 68–73 (2016)

17. Bonisoli, E., Lisitano, D., Dimauro, L.: Experimental and numerical mode shape tracing from components to whole motorbike chassis. In: Proceedings of the 28th International Conference on Noise and Vibration Engineering ISMA2018, September 17–19, 2018, Leuven, Belgium, pp. 3597–3604 (2018)

18. Bonisoli, E., Lisitano, D., Dimauro, L., Peroni, L.: Chapter 14: A proposal of dynamic behaviour design based on mode shape tracing: numerical application to a motorbike frame. In: Dynamic Substructures Proceedings of the 37th IMAC, A Conference and Exposition on Structural Dynamics 2019, Conference Proceedings of the Society for Experimental Mechanics Series, vol. 4, 186 pp, pp. 149–158. Springer (2020). https://doi.org/10.1007/978-3-030-12184-6_14

19. Bonisoli, E., Marcuccio, G., Rosso, C.: Chapter 18: Crossing and veering phenomena in crank mechanism dynamics. In: Topics in Model Validation and Uncertainty Quantification Proceedings of the 31st IMAC, A Conference and Exposition on Structural Dynamics, Conference Proceedings of the Society for Experimental Mechanics Series 2013, vol. 5, 264 pp, pp. 175–187. Springer (2013). https://doi.org/10.1007/978-1-4614-6564-5_18

20. Bonisoli, E., Marcuccio, G., Venturini, S.: Interference fit estimation through stress-stiffening effect on dynamics. Mech. Syst. Signal Process. **160**, 107919–107942 (2021). https://doi.org/10.1016/j.ymssp.2021.107919

21. Bonisoli, E., Lisitano, D., Dimauro, L.: Detection of critical mode-shapes in flexible multibody systems dynamics: the case study of a racing motorcycle. Mech. Syst. Signal Process. **180**, 1–23, 109370 (2022). https://doi.org/10.1016/j.ymssp.2022.109370

22. Bonisoli, E., Vella, A.D., Venturini, S.: Chapter 13: Uncertainty effects on bike spoke wheel modal behaviour. In: Model Validation and Uncertainty Quantification Proceedings of the 40th IMAC, A Conference and Exposition on Structural Dynamics 2022, Conference Proceedings of the Society for Experimental Mechanics Series, vol. 3, 150 pp, pp. 111–123. Springer (2023). https://doi.org/10.1007/978-3-031-04090-0_13

23. Mottershead, J.E., Mares, C., Friswell, M.I., James, S.: Selection and updating of parameters for an aluminum space-frame model. Mech. Syst. Signal Process. **14**(6), 923–944 (2000). https://doi.org/10.1006/mssp.2000.1303

24. Kyprianou, A., Mottershead, J.E.: Uncertain systems: modelling and updating. In: Proceedings of the 18th International Congress on Sound and Vibration (ICSV 2011), July 10–14, 2011, Rio de Janeiro, Brazil, pp. 995–1002 (2011)

25. Banhart, J.: Manufacturing routes for metallic foams. J. Miner. Met. Mater. Soc. **52**(12), 22–27 (2000). https://doi.org/10.1007/s11837-000-0062-8

26. Bonisoli, E., Delprete, C., Rosso, C.: Proposal of a modal-geometrical-based master nodes selection criterion in modal analysis. Mech. Syst. Signal Process. **23**(3), 606–620 (2009). https://doi.org/10.1016/j.ymssp.2008.05.012

27. Ibrahim, S.R.: Existence and normalization of complex modes for post experimental use in modal analysis. In: Modal Analysis and Testing NATO Science Series (Series E: Applied Sciences), vol. 363, pp. 441–452. Springer, Dordrecht (1999). https://doi.org/10.1007/978-94-011-4503-9_21

28. Peeters, B., Van der Auweraer, H., Guillaume, P., Leuridan, J.: The PolyMAX frequency-domain method: a new standard for modal parameter estimation? Shock. Vib. **11**, 395–409 (2004). https://doi.org/10.1155/2004/523692

29. Vacher, P., Jacquier, B., Bucharles, A.: Extensions of the MAC criterion to complex modes. In: Proceedings of the 24th International Conference on Noise and Vibration Engineering ISMA2010, September 20–22, 2010, Leuven, Belgium, pp. 2713–2725 (2010)

30. Pappa, R.S., Elliott, K.B., Schenk, A.: Consistent-mode indicator for the eigensystem realization algorithm. J. Guid. Control. Dyn. **16**(5), 852–858 (1993). https://doi.org/10.2514/3.21092

31. Bonisoli, E., Dimauro, L., Venturini, S., Lupos: Open-source scientific computing in structural dynamics. Proceedings of the 41st IMAC, A Conference and Exposition on Structural Dynamics, Austin, TX, USA, pp. 13–16 (2023). https://doi.org/10.1007/978-3-031-34946-1_23

Correction to: Assembling Uncertainty Effects on the Dynamic Response of Nominally Identical Motorbike Components

Elvio Bonisoli, Luca Dimauro, Simone Venturini, and Lorenzo Peroni

Correction to:
Chapter 32 in: R. Platz et al. (eds.), *Model Validation and Uncertainty Quantification, Volume 3*,
Conference Proceedings of the Society for Experimental Mechanics Series,
https://doi.org/10.1007/978-3-031-37003-8_32

The original version of the chapter has been revised. In reference 31, the DOI was missing in the initially published version. It has been corrected.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-3-031-37003-8_32