

The role of surrogate models in the development of digital twins of dynamic systems

S. Chakraborty^{a,1}, S. Adhikari^{b,2}, R. Ganguli^{c,3,*}

^a*Department of Applied Mechanics, Indian Institute of Technology Delhi, India*

^b*College of Engineering, Swansea University, Swansea, UK*

^c*Department of Aerospace Engineering, Indian Institute of Science, India*

Abstract

Digital twin technology has significant promise, relevance and potential of widespread applicability in various industrial sectors such as aerospace, infrastructure and automotive. However, the adoption of this technology has been slower due to the lack of clarity for specific applications. A discrete damped dynamic system is used in this paper to explore the concept of a digital twin. As digital twins are also expected to exploit data and computational methods, there is a compelling case for the use of surrogate models in this context. Motivated by this synergy, we have explored the possibility of using surrogate models within the digital twin technology. In particular, the use of Gaussian process (GP) emulator within the digital twin technology is explored. GP has the inherent capability of addressing noisy and sparse data and hence, makes a compelling case to be used within the digital twin framework. Cases involving stiffness variation and mass variation are considered, individually and jointly along with different levels of noise and sparsity in data. Our numerical simulation results clearly demonstrate that surrogate models such as GP emulators have the potential to be an effective tool for the development of digital twins. Aspects related to data quality and sampling rate are analysed. Key concepts introduced in this paper are summarised and ideas for urgent future research needs are proposed.

Keywords: Digital twin, vibration, response, frequency, surrogate

1. Introduction

A digital twin is a virtual model of a physical system which exists in the computer cloud. Increasingly, this physical system is called a physical twin [1]. Attempts to emulate the behaviour of physical systems are a key component of engineering and science. However, the difference between the digital twin and a computer model is that the digital twin updates itself to track the physical twin through the use of sensors, data analysis, machine learning, and the internet of things. The digital twin may also direct changes to

*Corresponding author. Tel: +91-80-22933017, Fax: +91-80-22933043

Email addresses: souvik@am.iitd.ac.in (S. Chakraborty), S.Adhikari@swansea.ac.uk (S. Adhikari), ganguli@iisc.ac.in (R. Ganguli)

¹Assistant Professor

²Chair of Aerospace Engineering, FRAeS

³Satish Dhawan Chair Professor, FRAeS

the physical twin through signals sent to actuators on the physical twin. Ideally, physical and digital twins must achieve temporal synchronization.

The widespread connectivity of physical systems to the internet allows the tracking of millions of physical twins in real-time. However, the challenge is to ensure that the digital twin mimics the physical system as closely as possible with the evolution of time. In other words, the digital twin must evolve with the physical twin over its life cycle. At the end of life, both digital and physical twins will terminate. The physical twins exist in the real world and the digital twin typically exists in the cloud [2]. Important problems in the development of digital twin technology are related to modelling and simulation of the physical twin, accuracy and speed of data transmission between the physical twin and the cloud where the digital twin resides, storage and processing of big data created by sensors through the life cycle, efficient computer processing of the digital twin model in the cloud, and possibly the actuation of the physical twin through signals sent by the digital twin from the cloud to actuators on the physical twin.

Several authors have considered the concept of digital twin over the last two decades. An early paper on digital twins in the context of life prediction of aircraft structures was published by Tuegel and his co-workers [3]. Their objective was to use an ultrahigh fidelity model of an individual aircraft identified by its tail number to predict its structural life. This research was motivated by the growing power of computers and the evolution of high-performance computing. The key idea introduced in this paper was the need to track each individual aircraft by its tail number, which is a unique identifier. Thus, every physical aircraft in the fleet would have a separate digital twin, which will evolve differently compared to the other twins as each aircraft faces unique combinations of missions, loads, environment, pilot behaviour, sensor situation, etc. This concept of digital twins was later expanded to any physical system such as cars, computer servers, locomotives, turbines, machine tools, etc. The primary applications of digital twin have been in production, product design, and prognostics and health management [4–8].

Digital twin technology has found many industrial applications, as summarized by Tao et al in their recent review paper [9]. They mention that digital twin is an enabling technology for realizing smart manufacturing and Industry 4.0. In Industry 4.0, factories are envisaged with wireless connectivity and sensors which allow the production line to function automatically. As physical systems become embellished with sensors, data transmission technology allows the collection of data throughout the life stage of the physical system. This data set is typically very large, and big data analytics is needed to find failure causes, streamline supply chains and enhance production efficiency. According to this review paper, prognostics and health management is the area which has seen the maximum application of digital twin technology. Li et al [10] created a digital twin based on the dynamic Bayesian network to monitor the operational state of aircraft wings. A probabilistic model was used to replace the deterministic physical model. Digital twins have also been used in prognostics and health management of cyber-physical systems and additive manufacturing processes. The fusion of the physical and the cyber systems remains the key challenge for the use of digital twins and one approach to address this problem is to make the modelling and simulation tasks more efficient. Surrogate modelling could be one approach to increase the efficiency of the simulation process.

In [11], the authors proposed a physics-based digital twin model for a dynamical

system. The development was based on the assumption that the properties of the physical system typically evolve much more slowly than real-time. This evolution thus takes place in slow time as compared to real-time. A single degree of freedom system model was created to explain the concept of digital twin and study the effect of changes in system mass and stiffness properties. Closed-form solutions to obtain the evolution of mass and stiffness with the slow time was proposed. In case, the data collected is clean, the proposed closed-form solutions yield exact estimates of the mass and stiffness evolution. However, the proposed physics-based digital twin has two major limitations. First, the physics-based digital twin proposed only yields the mass and stiffness at the slow time-step, i.e., the time-steps at which we have sensor measurements. In other words, the physics-based digital twin is not capable of providing the mass and stiffness at the intermediate and future time. Second, the physics-based digital twin can fail if the data collected is contaminated by noise. This is a major limitation as in real-life, data collected is always corrupted by noise.

From this discussion, it is clear that there is a lack of clarity and specificity regarding digital twins. Although there exists a broad consensus about what a digital twin is, there is almost no detailed methodology on how to develop one for a given system. In this work, we introduce the concept of surrogate models within the digital twin framework. This is motivated by the fact that a physics-driven digital twin, such as the one proposed in [11, 12] can only provide estimates at certain discrete time-steps (the time-steps corresponding to sensor measurements). Moreover, such digital twins are likely to yield erroneous results when the data collected is corrupted by complex noise. By definition, a surrogate model can be considered to be a *proxy* to the actual high-fidelity model. Popular surrogate models available in the literature include analysis-of-variance decomposition [13], polynomial chaos expansion [14, 15], support vector machines [16–18], Neural networks [19–21] and Gaussian process (GP) [22–24]. Successful use of surrogate models can be found in various domains including, but not limited to, stochastic mechanics [25, 26], reliability analysis [27, 28] and optimisation [29–31]. The objective of this paper is to illustrate how surrogate models can be integrated with the digital twin technology and what possible advantages can be achieved from this integration.

Although all the surrogate models discussed above can be used within the digital twin framework, we explore the use of GP. In essence, GP is a probabilistic machine learning technique that tries to infer a distribution over functions and then utilise the same to make predictions at some unknown points. The advantage of GP over other surrogate models is two-fold. Firstly, GP being a probabilistic surrogate model is immune to over-fitting. This is extremely important for the digital twin as the data collected are corrupted by complex noise and any form of over-fitting will result in erroneous predictions. Secondly, GP is able to quantify the uncertainty due to limited and noisy data. This, in turn, can be used in the decision-making process using a digital twin. In this paper, we have illustrated the use of *vanilla* GP as a surrogate with the digital twin model.

The paper is organized as follows. In Section 2, the equation of motion of an SDOF digital twin is introduced using multiple time-scales. The fundamentals of GP are discussed in Section 3. The development of a digital twin using only the mass evolution is considered in Section 4. Three separate cases with (a) only mass evolution, (b) only stiffness evolution and (c) both mass and stiffness evolution are considered. Numerical

examples are given to illustrate the proposed ideas. Some critical discussion on the proposed methodology as well as the overall development of digital twin is given in Section 5 and concluding remarks are given in Section 6.

2. The dynamic model of the digital twin

A single degree of freedom dynamic system is considered to explore the concept of a digital twin. A key idea is that a digital twin starts its journey from a ‘nominal model’. The nominal model is therefore, the ‘initial model’ or the ‘starting model’ of a digital twin. For engineering dynamic systems, the nominal model is often a physics based model which has been verified, validated and calibrated. For example, this can be a finite element model of a car or an aircraft when the product leaves the manufacturer facility and is ready to go into service. Another key characteristics of a digital twin is its connectivity. The recent development of the Internet of Things brings forward numerous new data technologies and consequently drives the development of digital twin technology. This enables connectivity between the physical SDOF system and its digital counterpart. The basis of digital twins is based on this connection, without it, digital twin technology cannot exist. This connectivity is created by sensors on the physical system which obtain data and integrate and communicate this data through various integration technologies.

We consider a physical system which can be well approximated by a single degree of freedom spring, mass and damper system as before [11]. It is reasonable to assume that the sensors sample data intermittently, typically, t_s represents discrete time points. It is assumed that changes in $k(t_s)$, $m(t_s)$ and $c(t_s)$ as so slow that the dynamics of the system is effectively decoupled from these functional variations. The equation of motion can be written as

$$m(t_s)\frac{\partial^2 u(t, t_s)}{\partial t^2} + c(t_s)\frac{\partial u(t, t_s)}{\partial t} + k(t_s)u(t, t_s) = f(t, t_s) \quad (1)$$

Here t and t_s are the system time and a “slow time”, respectively. $u(t, t_s)$ is a function of two variables and therefore the equation of motion is expressed in terms of the partial derivative with respect to the time variable t . The slow time or the service time t_s can be considered as a time variable which is much slower than t . For example, it could represent the number of cycles in an aircraft. Thus, mass $m(t_s)$, damping $c(t_s)$, stiffness $k(t_s)$ and forcing $F(t, t_s)$ change with t_s , for example due to system degradation during its service life. The forcing is also a function of time t and slow time t_s , as is the system response $x(t, t_s)$. Eq. (1) is considered as a digital twin of a SDOF dynamic system. When $t_s = 0$, that is at the beginning of the service life of the system, the digital twin Eq. (1) reduces to the nominal system

$$m_0\frac{d^2 u_0(t)}{dt} + c_0\frac{du_0(t)}{dt} + k_0 u_0(t) = f_0(t), \quad (2)$$

where m_0 , c_0 , k_0 and f_0 are respectively the mass, damping, stiffness and force at $t = 0$. It is assumed that sensors are deployed on the physical system and take measurements at locations of time defined by t_s . The functional form of the relationship of mass, stiffness and forcing with t_s is unknown and needs to be estimated from measured sensor data. We assume that damping is small so that the effect of variations in $c(t_s)$ is negligible.

In effect, only variations in the mass and stiffness are considered. Without any loss of generality, the following functional forms are considered

$$\begin{aligned} k(t_s) &= k_0(1 + \Delta_k(t_s)) \\ \text{and } m(t_s) &= m_0(1 + \Delta_m(t_s)) \end{aligned} \quad (3)$$

In general $k(t_s)$ is expected to be a decaying function over a long time to represent a loss in the stiffness of the system. On the other hand, $m(t_s)$ can be an increasing or a decreasing function. For example, in the context of aircraft, it can represent the loading of cargo and passengers and also represent the use of fuel as the flight progresses. The following representative functions have been chosen as examples

$$\Delta_k(t_s) = e^{-\alpha_k t_s} \frac{(1 + \epsilon_k \cos(\beta_k t_s))}{(1 + \epsilon_k)} - 1 \quad (4)$$

$$\text{and } \Delta_m(t_s) = \epsilon_m \text{SawTooth}(\beta_m(t_s - \pi/\beta_m)) \quad (5)$$

Here $\text{SawTooth}(\bullet)$ represents a sawtooth wave with a period 2π . In Fig. 1 overall variations in stiffness and mass properties arising from these function models have been plotted as function of time normalised to the natural time period of the nominal model. Numeri-

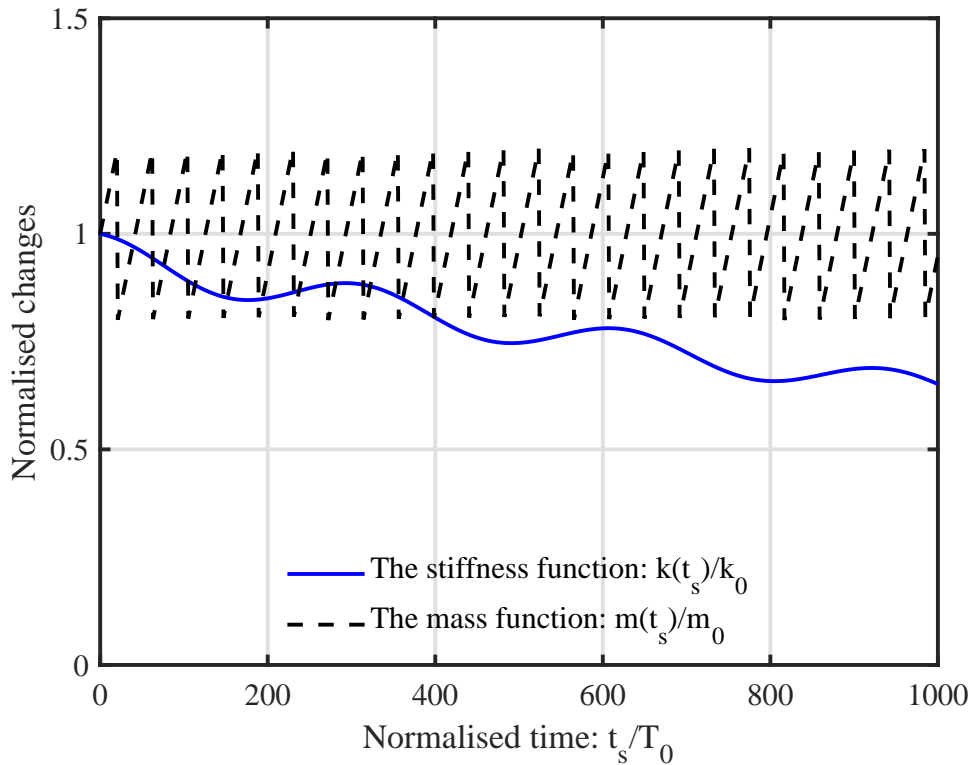


Fig. 1. Examples of model functions representing long term variabilities in the mass and stiffness properties of a digital twin system.

cal values used for these examples are $\alpha_k = 4 \times 10^{-4}$, $\epsilon_k = 0.05$, $\beta_k = 2 \times 10^{-2}$, $\beta_m = 0.15$ and $\epsilon_m = 0.25$. The choice of these functions is motivated by the fact that the stiffness degrades over time in a periodic manner representing a possible fatigue crack growth in

an aircraft over repeated pressurisation. On the other hand, the mass increases and decreases over the nominal value due to re-fuelling and fuel burn over a flight period. The key consideration is that a digital twin of the dynamical system should track these types of changes by exploiting sensor data measured on the system.

3. Overview of Gaussian process emulators

One of the most popular machine learning techniques is the Gaussian process (GP) [22, 23]. Gaussian process has been successfully used in the context of structural dynamic analysis [32] and finite element method [33]. Unlike frequentist machine learning techniques such as the analysis-of-variance decomposition [13], support vector machine [17] and neural networks [34], GP adopts an optimal approach and tries to infer a *distribution over functions* and then utilise the same to make predictions at some unknown points [35]. It is worthwhile to note that in literature, GP is also known as the Kriging [36, 37]. Different variants of GP has been exists in the literature. This includes fully Bayesian GP [22], sparse GP [38], multi-fidelity GP [39] among other. In this work, *vanilla* GP has been used and hence, the discussion hereafter is focused on the same. For details on other types of GP, interested readers may refer [40].

We consider, $\ell \in \mathbb{R}^d$ as the input variable and

$$y = g(\ell) + \nu \quad (6)$$

to be a set of noisy measurements of the response variable, where ν represents the noise. With this setup, the objective is to estimate the latent (unobserved) function $g(\ell)$ that will enable prediction of the response variable, \hat{y} at new values of ℓ . In GP based regression, we define a GP defined over $g(\ell)$ with the mean $\mu(\ell)$ and covariance function $\kappa(\ell, \ell'; \boldsymbol{\theta})$

$$g(\ell) \sim \mathcal{GP}(\mu(\ell), \kappa(\ell, \ell'; \boldsymbol{\theta})), \quad (7)$$

In the above equation

$$\begin{aligned} \mu(\ell) &= \mathbb{E}[g(\ell)], \\ \kappa(\ell, \ell'; \boldsymbol{\theta}) &= \mathbb{E}[(g(\ell) - \mu(\ell))(g(\ell') - \mu(\ell'))]. \end{aligned} \quad (8)$$

and $\boldsymbol{\theta}$ denotes the hyperparameters of the covariance function κ . The choice of the covariance function κ allows encoding of any prior knowledge about $g(\ell)$ (e.g., periodicity, linearity, smoothness), and can accommodate approximation of arbitrarily complex functions [41]. The notation in Eq. (7) implies that any finite collection of function values has a joint multivariate Gaussian distribution, that is $(g(\ell_1), g(\ell_2), \dots, g(\ell_N)) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K})$, where $\boldsymbol{\mu} = [\mu(\ell_1), \dots, \mu(\ell_N)]^T$ is the mean vector and \mathbf{K} is the covariance matrix with $\mathbf{K}(i, j) = \kappa(\ell_i, \ell_j)$ for $i, j = 1, 2, \dots, N$. If no prior information is available about the mean function, it is generally set to zero, i.e. $\mu(\ell) = 0$. However, for the covariance function, any function $\kappa(\ell, \ell')$ that generates a positive, semi-definite, covariance matrix \mathbf{K} is a valid covariance function. With this set-up, the objective of GP is to estimate the hyperparameters, $\boldsymbol{\theta}$, based on the observed input-output pairs, $\{\ell_j, y_j\}_{j=1}^{N_t}$, where N_t is the number of training samples. In general, this is achieved by maximising the likelihood of the data [24]. Alternatively, one can choose to adopt a Bayesian approach to compute the posterior of the hyperparameters, $\boldsymbol{\theta}$ [22]. Once $\boldsymbol{\theta}$ have been computed, the predictive distribution of $g(\ell^*)$ given the dataset ℓ, \mathbf{y} , hyperparameters $\boldsymbol{\theta}$ and new inputs, ℓ^* is

represented as

$$p(g(\ell^*) | \mathbf{y}, \boldsymbol{\ell}, \boldsymbol{\theta}, \ell^*) = \mathcal{N}(g(\ell^*) | \mu_{\mathcal{GP}}(\ell^*), \sigma_{\mathcal{GP}}^2(\ell^*)), \quad (9)$$

where

$$\begin{aligned} \mu_{\mathcal{GP}}(\ell^*) &= \mathbf{k}^T(\boldsymbol{\ell}, \ell^*; \boldsymbol{\theta}) [\mathbf{K}(\boldsymbol{\ell}, \boldsymbol{\ell}; \boldsymbol{\theta}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y}, \\ \sigma_{\mathcal{GP}}^2(\ell^*) &= k(\ell^*, \ell^*; \boldsymbol{\theta}) - \mathbf{k}^T(\boldsymbol{\ell}, \ell^*; \boldsymbol{\theta}) [\mathbf{K}(\boldsymbol{\ell}, \boldsymbol{\ell}; \boldsymbol{\theta}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{k}(\boldsymbol{\ell}, \ell^*; \boldsymbol{\theta}). \end{aligned} \quad (10)$$

For further details on GP, interested readers may refer [42]. Next, we employ GP to obtain digital twin of a single degree of freedom damped dynamic systems.

4. Gaussian process based digital twin

The development of digital twin concept hinges on the advances in the sensor technology. Using modern sensors, it is possible to collect different types of data such as acceleration time-history, displacement time-history *etc.* In this paper, we assume that using sensors, the natural frequency of the system described in Eq. (1) have been measured at some distinct time-steps. Wireless sensor technology is making such measurements and transmission possible [43]. For example, Wang *et al.* [44] showed that the natural frequency of a system can be measured by using sensor technology, a fact that we use in this paper.

4.1. Digital twin via stiffness evolution

4.1.1. Formulation

In the first case, we assume that the change in the natural frequency with time is due to the deterioration of the stiffness of the system only. With this, the equation of motion of the digital twin is represented as

$$m_0 \frac{d^2 u(t)}{dt^2} + c_0 \frac{du(t)}{dt} + k(t_s) u(t) = f(t), \quad (11)$$

where

$$k(t_s) = k_0 (1 + \Delta_k(t_s)). \quad (12)$$

Eq. (11) is a special case of Eq. (1). Note that in a practical scenario, we generally have no prior information about how the stiffness will deteriorate. We postulate that the deterioration of stiffness can be represented by using a GP

$$\Delta_k(t_s) \approx \Delta_{\hat{k}}(t_s) \sim \mathcal{GP}(\mu_k(t_s), \kappa_k(t_s, t'_s; \boldsymbol{\theta})). \quad (13)$$

However, at this stage, we have no measurements corresponding to $\Delta_{\hat{k}}(t_s)$ and hence, it is not possible to estimate the hyperparameters, $\boldsymbol{\theta}$ of the GP. We substitute Eqs. (12) and (13) into Eq. (11). With this, the governing differential equation of the digital twin can be represented as

$$m_0 \frac{d^2 u(t)}{dt^2} + c_0 \frac{du(t)}{dt} + k_0 (1 + \Delta_{\hat{k}}(t_s)) u(t) = f(t). \quad (14)$$

Solving Eq. (14), the natural frequency of the system is obtained as

$$\lambda_{s1,2}(t_s) = -\zeta_0 \omega_0 \pm i \omega_0 \sqrt{1 + \Delta_{\hat{k}}(t_s) - \zeta_0^2}, \quad (15)$$

where ω_0 and ζ_0 are the natural frequency and damping ratio of the system at $t_s = 0$. Rearranging Eq. (15), we have

$$\lambda_{s1,2}(t_s) = - \underbrace{\frac{\zeta_0}{\sqrt{1 + \Delta_{\hat{k}}(t_s)}}}_{\zeta_s(t_s)} \underbrace{\omega_0 \sqrt{1 + \Delta_{\hat{k}}(t_s)}}_{\omega_s(t_s)} \pm i \underbrace{\omega_0 \sqrt{1 + \Delta_{\hat{k}}(t_s)} \sqrt{1 - \left(\frac{\zeta_0}{\sqrt{1 + \Delta_{\hat{k}}(t_s)}} \right)^2}}_{\omega_{d_s}(t_s)}, \quad (16)$$

where $\omega_s(t_s) = \omega_0 \sqrt{1 + \Delta_{\hat{k}}(t_s)}$ is the evolution of the natural frequency, $\zeta(t_s) = \zeta_0 / \sqrt{1 + \Delta_{\hat{k}}(t_s)}$ is the evolution of the damping factor and $\omega_{d_s}(t_s) = \omega_s(t_s) \sqrt{1 - \zeta_s^2(t_s)}$ is the evolution of the damped natural frequency with slower time-scale t_s . Since most of the natural frequency extraction techniques extract the damped natural frequency of the system, we assume that the data obtained using the sensor to be the same. Following Ganguli and Adhikari [11], it can be shown that

$$\Delta_{\hat{k}}(t_s) = -\tilde{d}_1(t_s) \left(2\sqrt{1 - \zeta_0^2} - \tilde{d}_1(t_s) \right), \quad (17)$$

where

$$\tilde{d}_1(t_s) = \frac{d_1(\omega_{d_0}, \omega_{d_s}(t_s))}{\omega_0}. \quad (18)$$

The function $d_1(\omega_{d_0}, \omega_{d_s}(t_s))$ in Eq. (18) is the distance between ω_{d_0} and $\omega_{d_s}(t_s)$

$$d_1(\omega_{d_0}, \omega_{d_s}(t_s)) = \|\omega_{d_0} - \omega_{d_s}(t_s)\|_2. \quad (19)$$

Now given the fact that the initial damped frequency of the system, ω_{d_0} is known and we have sensor measurements for $\omega_{d_s}(t_s)$, one can easily compute $\tilde{d}_1(t_s)$ and $\Delta_{\hat{k}}(t_s)$ by using Eq. (18) and substituting it into Eq. (17). Having said that, it is only fair to assume that the sensor measurements, $\omega_{d_s}(t_s)$, will be corrupted by some noise and hence, the estimates of $\Delta_{\hat{k}}(t_s)$ are also noisy. Nonetheless, by considering these noisy estimates of $\Delta_{\hat{k}}(t_s)$ to be training outputs corresponding to the inputs t_s , we obtain the hyperparameters (θ) of the GP described in Eq. (13). In this work, we use Bayesian optimisation [45] to maximise the likelihood of the data. Within the Bayesian optimization framework, L-BFGS optimizer is used with a tolerance of 10^{-5} . To avoid local convergence, multiple starting points were used. The hyperparameter θ completely describes the digital-twin in Eq. (14). One important question associated with GP is the form of the covariance function and the mean function. In this work, we have considered a pool for the mean function and the covariance function and then selected the most suitable mean function and covariance function based on Bayesian information criteria. The possible candidates for the mean function and the covariance function are shown in Table 1. Functional form of these covariance kernels can be found in [41]. Details on Bayesian information criteria are presented in Appendix Appendix A.

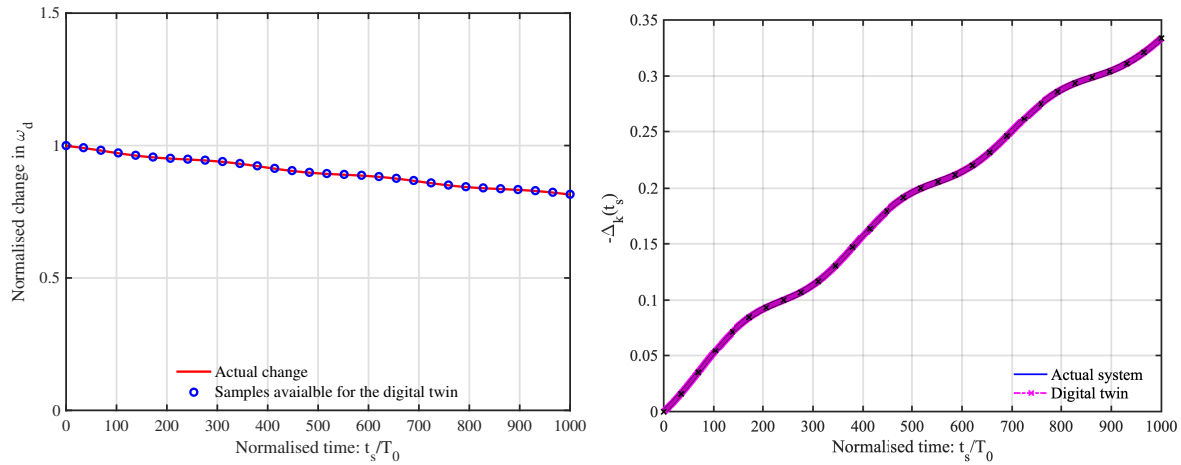
4.1.2. Numerical illustration

To illustrate the applicability of GP based digital twin with only stiffness evolution, an SDOF system with nominal damping ratio $\zeta_0 = 0.05$ is considered. We assume that the sensor data is transmitted intermittently with a certain regular time interval. For simulating the variation in the natural frequency, we consider the change in the stiffness

Table 1. Candidate functions for mean and covariance functions in GP

Functions	Candidates
Mean function	Constant, Linear, Quadratic
Covariance function	Exponential, Squared Exponential, Matern 3/2, Matern 5/2, Rational Quadratic, ARD Exponential, ARD Squared Exponential, ARD Matern 3/2, ARD Matern 5/2, ARD Rational Quadratic

property of the system shown in Fig. 1. Fig. 2(a) shows the actual change in the damped natural frequency of the system with time. The data-points available for the digital twin are also shown. At this stage, it is important to emphasise that the frequency of data availability depends on several factors such as the bandwidth of the wireless transmission system, cost of data collection, *etc.* In this case, we have considered 30 data-points. Fig. 2(b) shows the GP based digital twin model with clean data. We observe that the GP based digital twin perfectly captures the time evolution of Δ_k with utmost confidence. However, we will like to emphasise that this is an imaginary scenario as it is almost impossible to have measurements with no noise.



(a) Changes in the damped natural frequency with time

(b) GP assisted digital twin with clean data

Fig. 2. Changes in the (damped) natural frequency and the GP based digital twin obtained using the exact data plotted as a function of the normalised slow time t_s/T_0 . Using Bayesian information criteria, ‘linear’ based function and ‘ARD Rational quadratic’ covariance functions are selected. The estimated hyperparameters from Bayesian optimisation are $\beta = [0.1804, 0.1009]$ and $\theta = [0.3729, 3.48 \times 10^8, 0.0128]$.

Now we consider a more realistic case where the data collected is corrupted by noise. For illustration purposes, a zero-mean Gaussian white noise with three different standard deviations is considered: (a) $\sigma_\theta = 0.005$, (b) $\sigma_\theta = 0.015$ and (c) $\sigma_\theta = 0.025$. Figs. 3(a)–(c) show the GP based digital twin model corresponding to the three cases. For $\sigma_\theta = 0.005$, the GP based digital twin successfully captures the evolution of stiffness with time. With the increase in the noise level, a slight deterioration in the performance of the GP based digital twin is observed. Nonetheless, even when $\sigma_\theta = 0.025$, the GP based digital twin

captures the time-evolution with high accuracy. An additional advantage of GP based digital twin resides in its capability to capture the uncertainty arising due to limited data and noise. This is illustrated by a shaded plot in Fig. 3. We observe that with increase in noise level, the uncertainty also increases.

To illustrate the importance of having more data, a case with 100 data-points is shown in Fig. 3(d). We observe that with the increase in the number of observations, the GP based digital twin can capture the evolution of stiffness in a more accurate manner even for the highest level of noise in the data.

4.2. Digital twin via mass evolution

4.2.1. Formulation

In the second case, we consider that the time-evolution of the natural frequency is due to the change in the mass of the system. Subsequently, the equation of motion of the digital twin can be represented as

$$m_s(t_s) \frac{d^2u(t)}{dt^2} + c_0 \frac{du(t)}{dt} + k_0u(t) = f(t), \tag{20}$$

where

$$m_s(t_s) = m_0(1 + \Delta_m(t_s)). \tag{21}$$

Similar to case 1, we assume

$$\Delta_m(t_s) \approx \Delta_{\hat{m}}(t_s) \sim \mathcal{GP}(\mu_m(t_s), \kappa_{\hat{m}}(t_s)), \tag{22}$$

substitute Eqs. (21) and (22) into Eq. (20) and solve it to obtain the natural frequency of the system

$$\lambda_{s1,2}(t_s) = -\omega_s(t_s) \zeta_s(t_s) \pm i\omega_{d_s}(t_s), \tag{23}$$

where

$$\omega_s(t_s) = \frac{\omega_0}{\sqrt{1 + \Delta_{\hat{m}}(t_s)}}, \tag{24a}$$

$$\zeta_s(t_s) = \frac{\zeta_0}{\sqrt{1 + \Delta_{\hat{m}}(t_s)}} \text{ and} \tag{24b}$$

$$\omega_{d_s}(t_s) = \omega_s(t_s) \sqrt{1 - \zeta_s^2(t_s)}. \tag{24c}$$

are the evolution of natural frequency, damping ratio and damped natural frequency of the digital twin. Again, following a procedure similar to case 1, we obtain

$$\begin{aligned} \Delta_{\hat{m}}(t_s) = & \frac{-2\tilde{d}_2(t_s)^2 + 4\tilde{d}_2(t_s) \sqrt{1 - \zeta_0^2} - 1 + 2\zeta_0^2}{2 \left(-\tilde{d}_2(t_s) + \sqrt{1 - \zeta_0^2} \right)^2} \\ & + \frac{\sqrt{1 - 4\tilde{d}_2(t_s)^2 \zeta_0^2 + 8\tilde{d}_2(t_s) \sqrt{1 - \zeta_0^2} \zeta_0^2 - 4\zeta_0^2 + 4\zeta_0^4}}{2 \left(-\tilde{d}_2(t_s) + \sqrt{1 - \zeta_0^2} \right)^2} \end{aligned} \tag{25}$$

where $\tilde{d}_2(t_s)$ is the equivalent of \tilde{d}_1 for the mass evolution case. Similar to case 1, $\Delta_{\hat{m}}(t_s)$ obtained using Eq. (25) will be noisy in nature. We consider this noisy estimates of $\Delta_{\hat{m}}(t_s)$ to be the training outputs and t_s to be the training inputs for estimating the

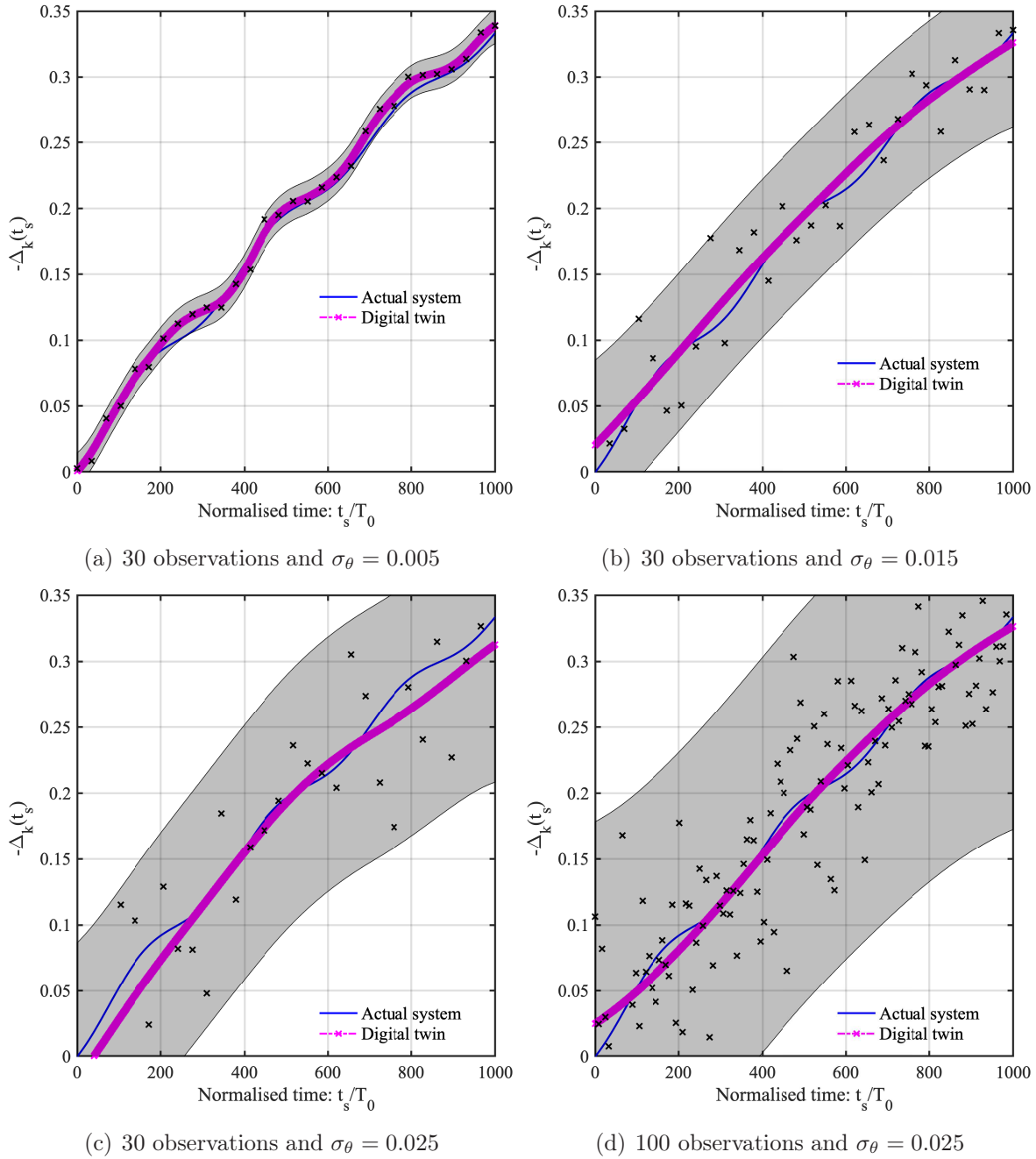


Fig. 3. GP based digital twin constructed with noisy data as a function of the normalised slow time t_s/T_0 . For (a) - (c), Bayesian information criteria selects ‘constant’ basis function. ‘ARD Rational quadratic’, ‘ARD squared exponential’ and ‘ARD Matern 3/2’ covariance kernels are selected for cases (a) to (c) respectively. For (d), Bayesian information criteria yields ‘Constant’ basis with $\beta = 0.1768$ and ‘ARD Matern 3/2’ covariance kernel with $\theta = [2.7298 \times 10^3, 0.3402]$. The shaded plot depicts the 95% confidence interval.

hyperparameters, θ , of the GP defined in Eq. (22). This is achieved by maximising the likelihood of the data. The parameter settings for solving the optimization problem are kept similar as before. For determining the best mean function and covariance function, Bayesian information criteria as before has been adopted. Once the hyperparameter, θ ,

has been estimated, the digital twin model defined in Eq. (20) is completely deciphered.

4.2.2. Numerical illustration

To illustrate the applicability of the GP based digital twin model for mass evolution, we revisit the SDOF example considered for stiffness evolution. For simulating the variation in the natural frequency, we consider the change in the mass of the system shown in Fig. 1. Fig. 4(a) shows the actual change in the damped natural frequency of the system with time. The data-points available for the digital twin are also shown. Note that the evolution of mass with time is more complex and hence, unlike the stiffness evolution case, only 30 observations is inadequate. Therefore, we have assumed access to more data points in this case. Fig. 4(b) shows the GP based digital twin constructed from clean data. It is observed that the GP based digital twin is able to capture the time evolution of mass with high accuracy. Also the uncertainty due to limited data is adequately captured. However, as already stated, this is an unrealistic case as for a practical scenario, data collected will always be corrupted by some form of noise.

Next, we consider a more realistic case where the data collected is corrupted by noise. Fig. 5 shows the GP based digital twin model trained with 100 noisy observations and $\sigma_\theta = 0.005$. Some discrepancy between the GP based digital twin and the physical twin is observed. For improved performance, we increase the number of observations to 150. The GP based digital twin corresponding to the three noise level are shown in Fig. 6(a) – (c). In this case, the GP based digital twin yields accurate result. Lastly, Fig. 7 shows GP based digital twin with 200 samples and $\sigma_\theta = 0.025$. This yields the best result with accurate time evolution of mass and uncertainty quantification. Overall, this example illustrates the importance of sampling rate and denoising of data for digital twin technology.

4.3. Digital twin via mass and stiffness evolution

4.3.1. Formulation

As the last case, we consider simultaneous evolution of the mass and stiffness. The governing differential equation for this case is represented as

$$m_s(t_s) \frac{d^2 u(t)}{dt^2} + c_0 \frac{du(t)}{dt} + k_s(t_s) u(t) = f(t), \quad (26)$$

where $k_s(t_s)$ and $m_s(t_s)$ are represented by Eqs. (12) and (21), respectively. We represent $\Delta_m(t_s)$ and $\Delta_k(t_s)$ by a multi-output GP [22, 38],

$$[\Delta_m(t_s), \Delta_k(t_s)] \approx [\Delta_{\hat{m}}(t_s), \Delta_{\hat{k}}(t_s)] \sim \mathcal{GP}(\boldsymbol{\mu}_{t_s}, \boldsymbol{\kappa}(t_s, t'_s; \boldsymbol{\theta})), \quad (27)$$

Substituting Eqs. (27), (12) and (21) in Eq. (26) and solving, we obtain

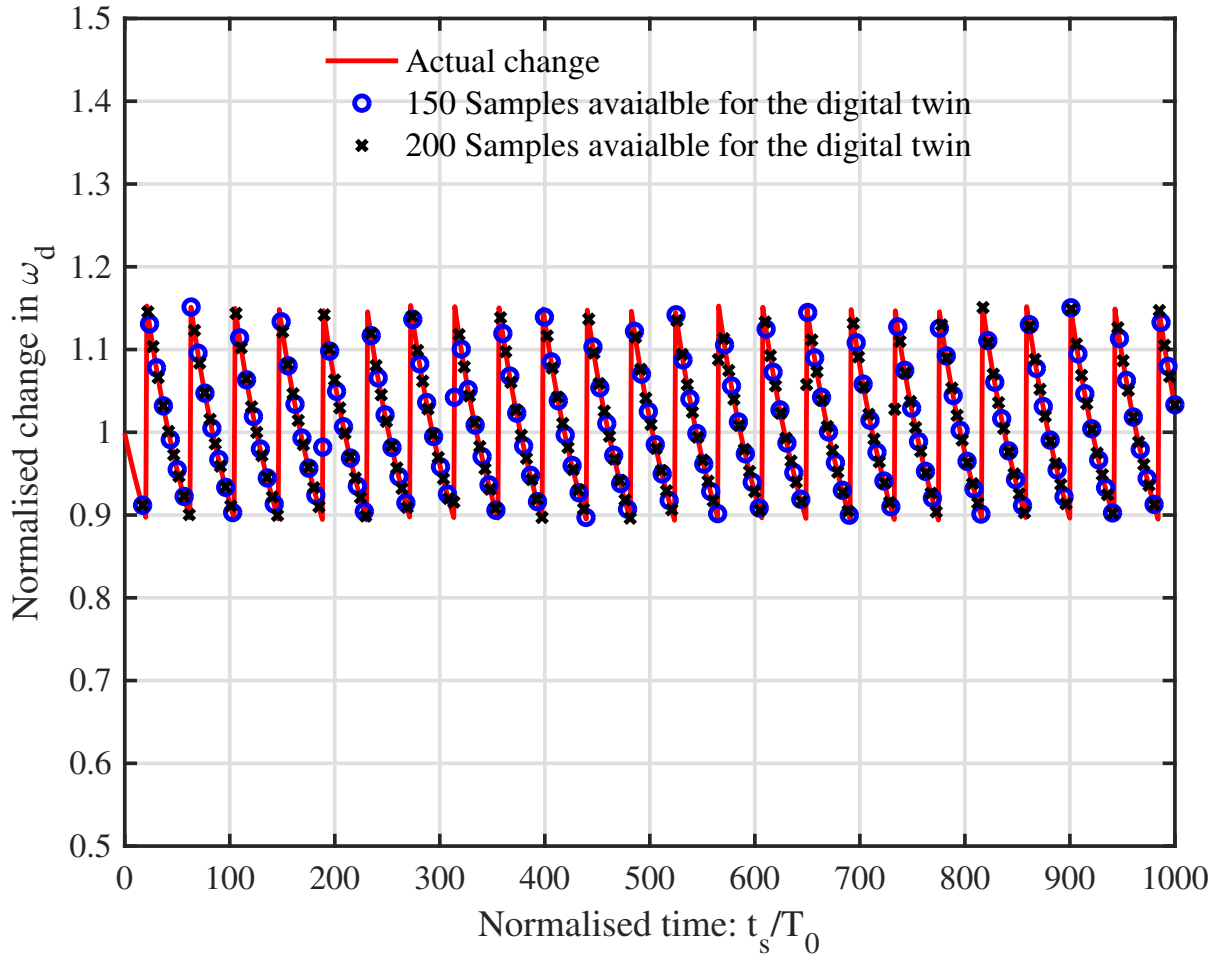
$$\lambda_{s1,2} = -\omega_s(t_s) \zeta_s(t_s) \pm i\omega_{d_s}(t_s), \quad (28)$$

where

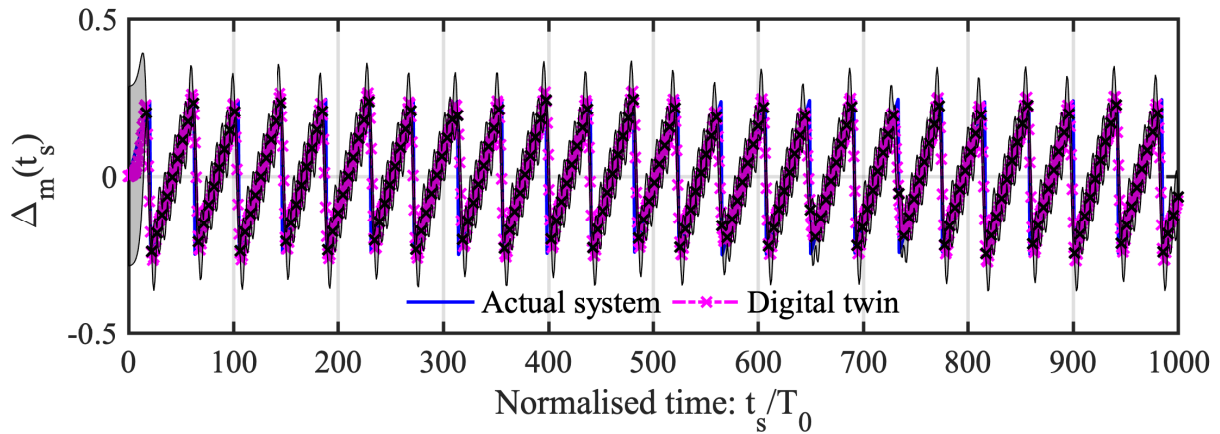
$$\omega_s(t_s) = \omega_0 \frac{\sqrt{1 + \Delta_{\hat{k}}(t_s)}}{\sqrt{1 + \Delta_{\hat{m}}(t_s)}} \quad (29a)$$

$$\zeta_s(t_s) = \frac{\zeta_0}{\sqrt{1 + \Delta_{\hat{m}}(t_s)} \sqrt{1 + \Delta_{\hat{k}}(t_s)}} \quad \text{and} \quad (29b)$$

$$\omega_{d_s}(t_s) = \omega_s(t_s) \sqrt{1 - \zeta_s^2}. \quad (29c)$$



(a) Changes in the (damped) natural frequency



(b) GP based digital twin

Fig. 4. Changes in the (damped) natural frequency and the GP based digital twin obtained using the exact data plotted as a function of the normalised slow time t_s/T_0 . Using Bayesian information criteria, ‘Constant’ basis function and ‘ARD Matern 5/2’ covariance functions are selected. The estimated hyperparameters from Bayesian optimisation are $\beta = 0.0$ and $\theta = [4.1532, 0.1429]$. The shaded plot depicts the 95% confidence interval.

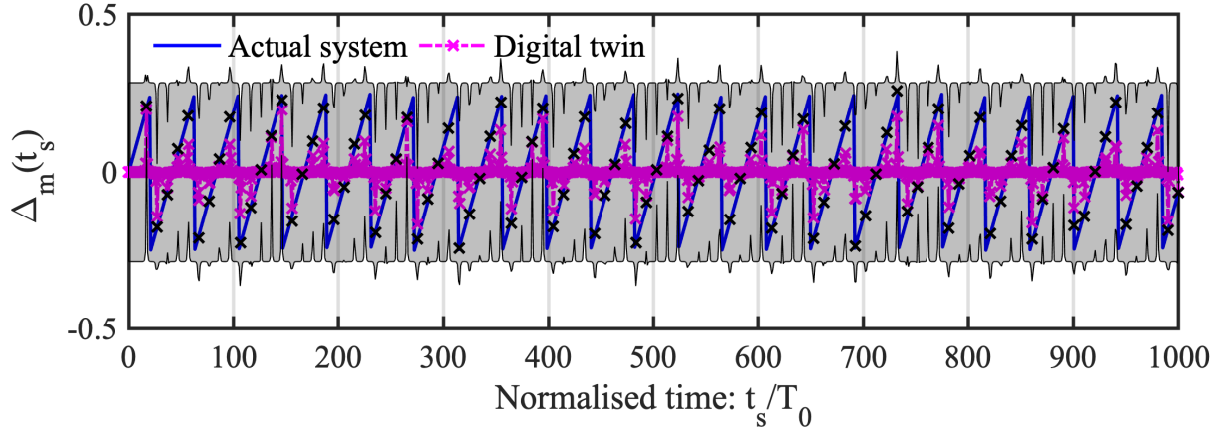


Fig. 5. GP based digital twin obtained using 100 noisy data with $\sigma_\theta = 0.005$, plotted as a function of the normalised slow time t_s/T_0 . The shaded plot depicts the 95% confidence interval.

Note that unlike the previous two cases, it is not possible to compute the hyperparameters of the GPs based on measurements of damped natural frequencies only. Therefore, we take a different route by considering the real and imaginary part in Eq. (28) separately. Following Ganguli and Adhikari [11], it can be shown that

$$\Delta_{\hat{m}}(t_s) = -\frac{\tilde{d}_{\mathcal{R}}(t_s)}{\zeta_0 + \tilde{d}_{\mathcal{R}}(t_s)}, \quad (30a)$$

$$\Delta_{\hat{k}}(t_s) = \frac{\zeta_0 \tilde{d}_{\mathcal{R}}^2(t_s) - (1 - 2\zeta_0^2) \tilde{d}_{\mathcal{I}}(t_s) + \zeta_0^2 \tilde{d}_{\mathcal{I}}^2(t_s)}{\zeta_0 + \tilde{d}_{\mathcal{R}}(t_s)}, \quad (30b)$$

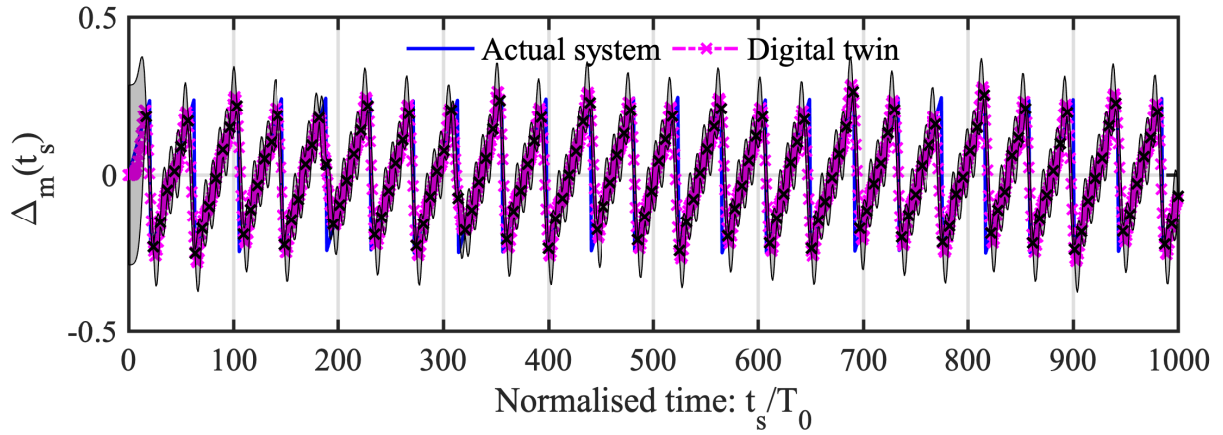
where $\tilde{d}_{\mathcal{R}}(t_s)$ and $\tilde{d}_{\mathcal{I}}(t_s)$, as before, are distance measures

$$\tilde{d}_{\mathcal{R}}(t_s) = \frac{d_{\mathcal{R}}(t_s)}{1 + \Delta_{\hat{m}}(t_s)}, \quad \tilde{d}_{\mathcal{I}}(t_s) = \sqrt{1 - \zeta_0^2} - \frac{\sqrt{(1 + \Delta_{\hat{k}}(t_s))(1 + \Delta_{\hat{m}}(t_s)) - \zeta_0^2}}{1 + \Delta_{\hat{m}}(t_s)}. \quad (31)$$

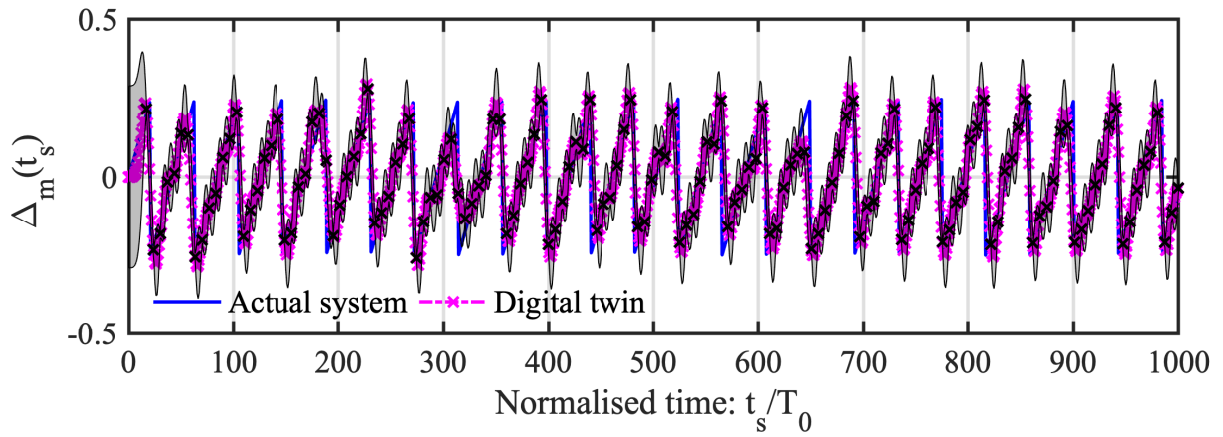
We emphasise that Eq. (30) provides estimates $\Delta_{\hat{m}}(t_s)$ and $\Delta_{\hat{k}}(t_s)$ based on noisy measurements, and hence, are also noisy. Using this noisy estimates as training outputs and t_s as the training inputs, we estimate the hyperparameters, θ of the multi-output GP defined in Eq. (27). This is achieved by maximising the likelihood of the data as described in Section 3. Same parameter settings as before is considered. For determining the best mean function and covariance function, Bayesian information criteria is adopted as before. The hyperparameters, θ once estimated completely defines the digital twin described in Eq. (26).

4.3.2. Numerical illustration

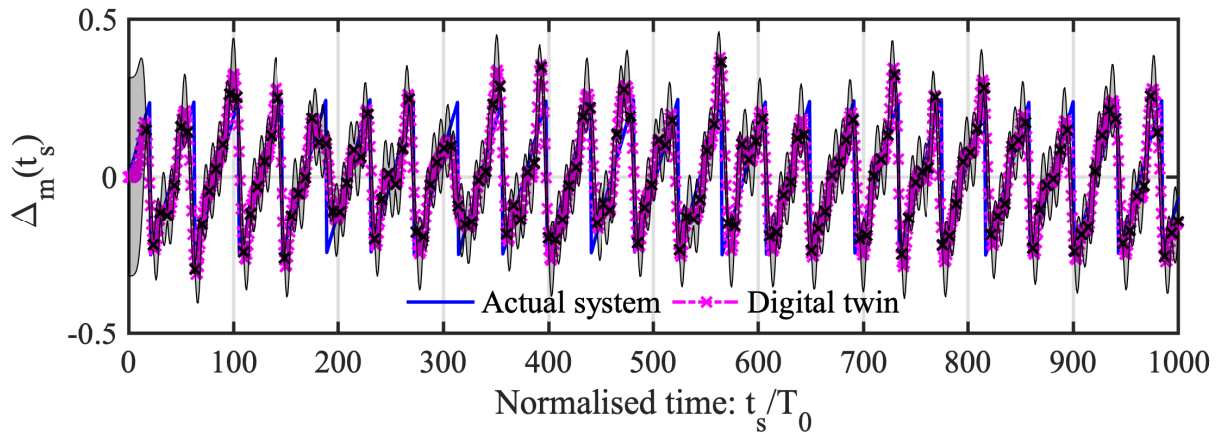
We revisit the previously studied SDOF system to illustrate the applicability of GP based digital twin for simultaneous mass and stiffness evolution. For simulating the variation in the natural frequency, we consider the change in the mass of the system shown in Fig. 1. Fig. 8 shows the actual change in real and imaginary parts of the natural frequency of the system over time. Similar to previous cases, the damping ratio is assumed to be 0.05. Fig. 9 shows the GP based digital twin constructed from clean data. The



(a) $\sigma_\theta = 0.005$



(b) $\sigma_\theta = 0.015$



(c) $\sigma_\theta = 0.025$

Fig. 6. GP based digital twin obtained using 150 noisy data plotted as a function of the normalised slow time t_s/T_0 . For (a) - (c), Bayesian information criteria selects ‘Constant’ basis and ‘ARD Matern 5/2’ covariance kernel. The shaded plot depicts the 95% confidence interval.

trained GP based digital twin is able to capture the time evolution of mass and stiffness.

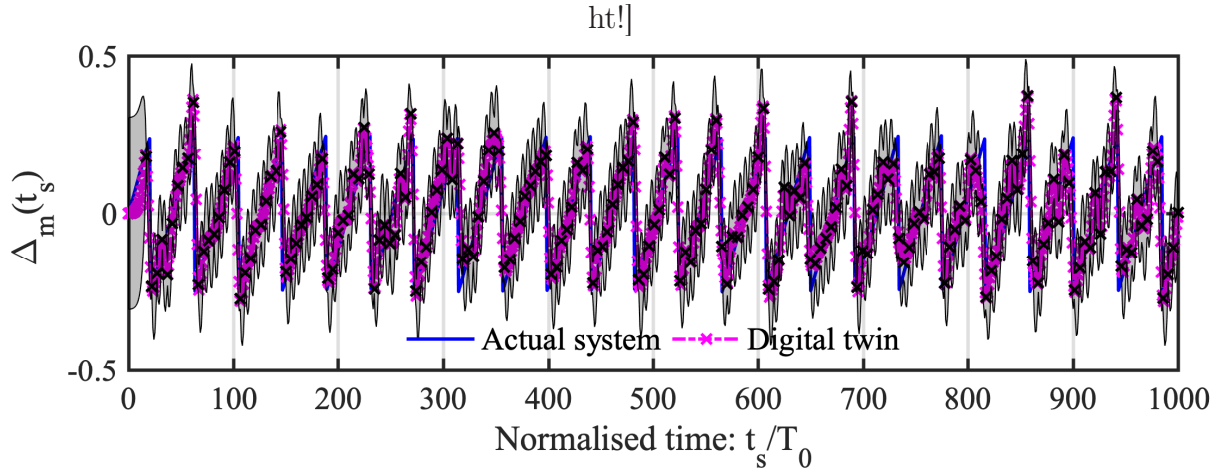


Fig. 7. GP based digital twin obtained using 200 noisy data with $\sigma_\theta = 0.025$ plotted as a function of the normalised slow time t_s/T_0 . Bayesian information criteria yields ‘Constant’ basis with $\beta = 0.0$ and ‘ARD Matern 5/2’ covariance kernel with $\theta = [4.1242, 0.1527]$.

However, data with no noise is an unrealistic scenario as, even with the most advanced sensors, the data collected will always be noisy [46].

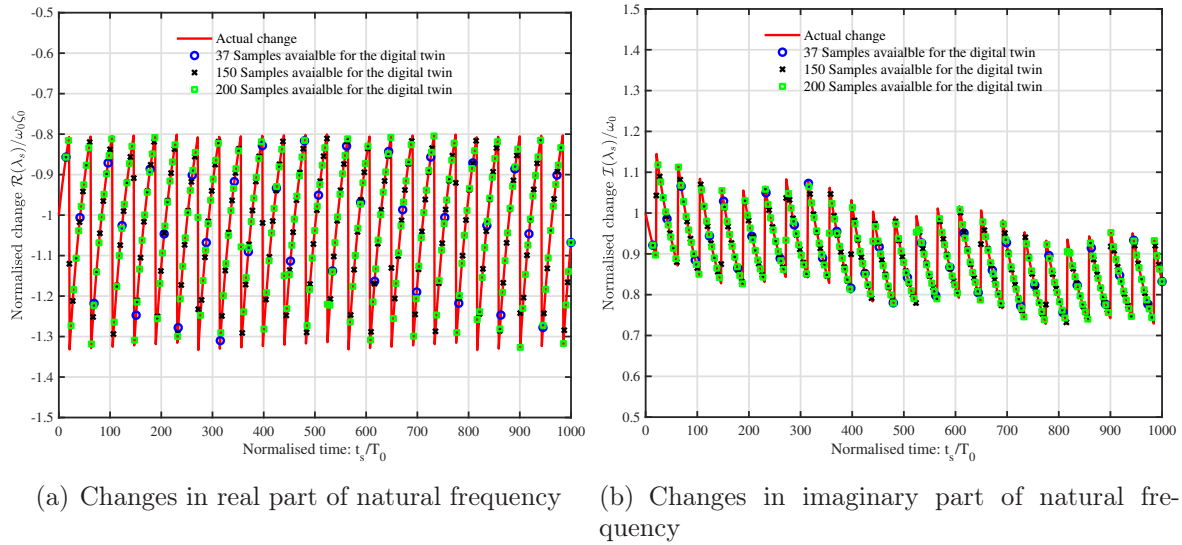
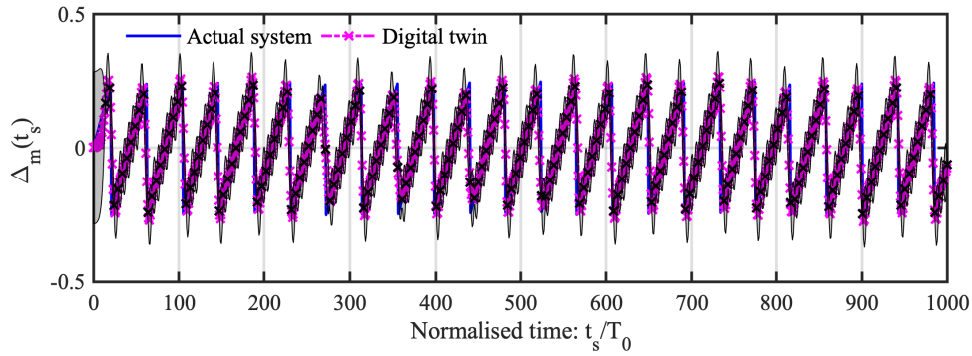
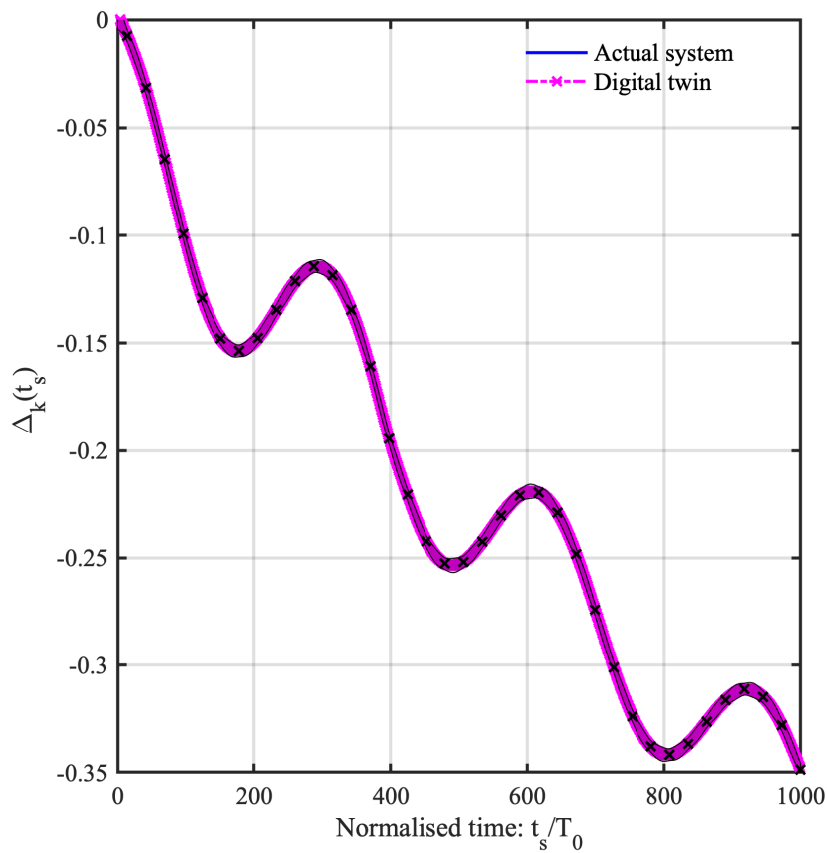


Fig. 8. Normalised changes in the real and imaginary parts of the natural frequency as a function of the normalised slow time t_s/T_0 . The shaded plot depicts the 95% confidence interval.

Next, we consider a more realistic case where the sensor data is corrupted by noise. Similar to the previous two cases, we consider three noise levels. Fig. 10 shows the mass and stiffness evolution of the GP based digital twin trained with 37 noisy observations. The observations are corrupted by white Gaussian noise with $\sigma_\theta = 0.005$. It is observed that the GP based digital twin is able to capture the time evolution of stiffness with high accuracy. However, it fails to capture the mass evolution in an adequate manner. This is expected as the mass evolution function is quite complex and it is unreasonable to expect that with so little data, GP will be able to track the mass evolution. Note that



(a) mass function

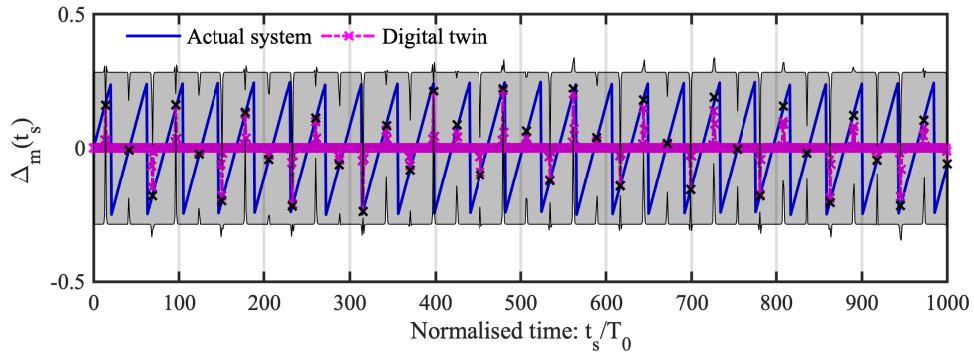


(b) stiffness function

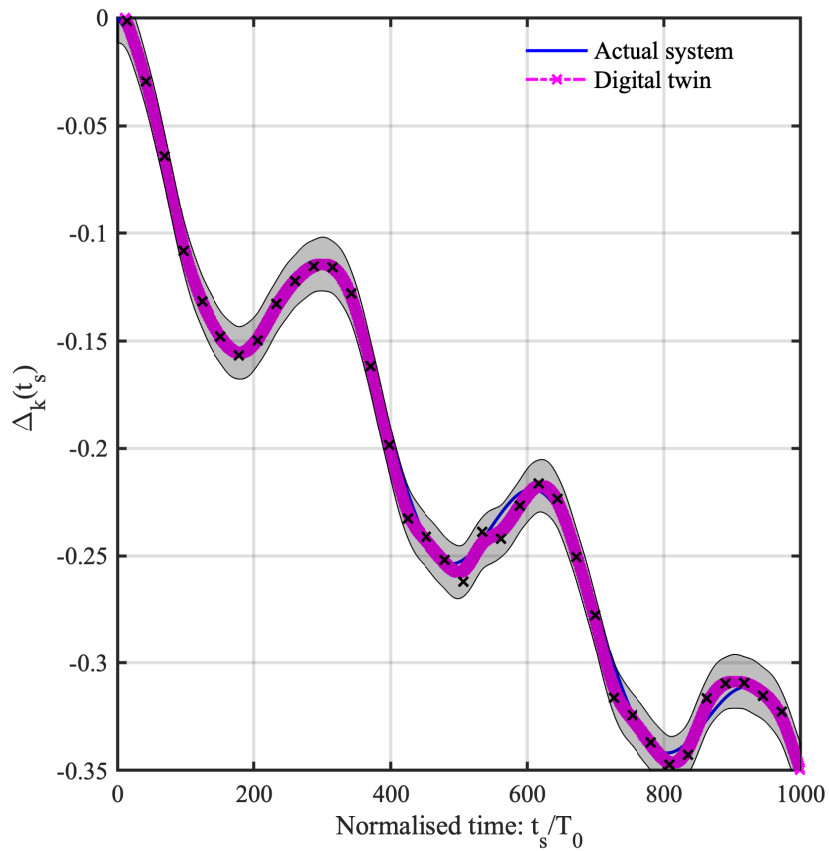
Fig. 9. GP based digital twin obtained from exact data via simultaneous mass and stiffness evolution as a function of the normalised ‘slow time t_s/T_0 . Bayesian information criteria yields ‘Linear’ basis and ‘squared exponential’ covariance kernel.

the uncertainty due to limited and noisy data is perfectly captured and hence, the true solution resides within the shaded portion.

Figs. 11 and 12 show the evolution of mass and stiffness of the GP based digital twin trained with 150 noisy observations. For mass evolution, results corresponding to all three noise levels are shown. With the increase in the number of observations, we observe a dramatic improvement in the GP assisted digital twin. The evolution of stiffness is shown



(a) mass function



(b) stiffness function

Fig. 10. GP based digital twin obtained from 37 noisy data with $\sigma_\theta = 0.005$ via simultaneous mass and stiffness evolution as a function of the normalised slow time t_s/T_0 . Bayesian information criteria yields ‘Linear’ basis and ‘squared exponential’ covariance kernel. The shaded plot depicts the 95% confidence interval.

only for $\sigma_\theta = 0.025$ and results obtained are highly accurate. Lastly, evolution of mass with 200 noisy observations and $\sigma_\theta = 0.025$ is shown in Fig. 13. As expected, this yields the best results. The evolution of stiffness for 200 observations are found to be similar to those with 150 samples and hence, the same is not shown in this paper.

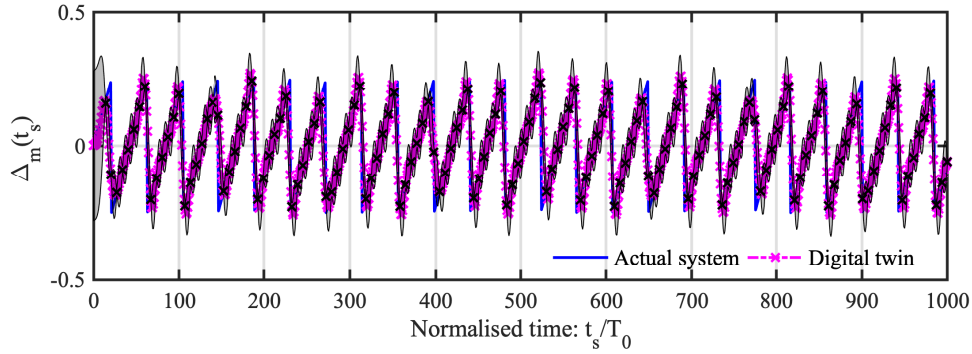
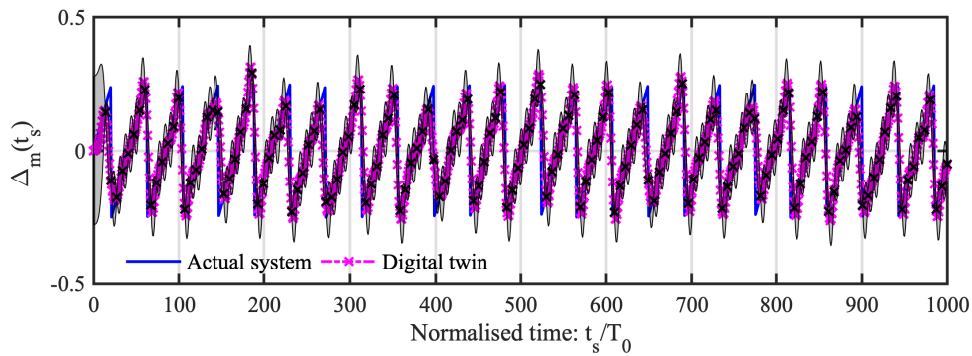
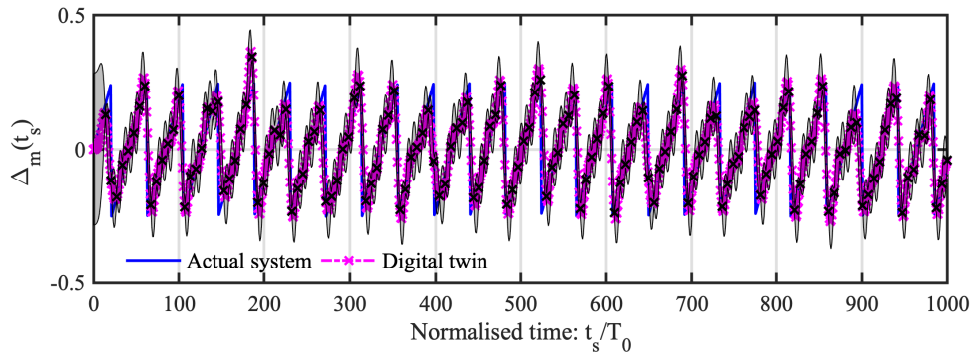
(a) $\sigma_\theta = 0.005$ (b) $\sigma_\theta = 0.015$ (c) $\sigma_\theta = 0.025$

Fig. 11. Mass evolution as a function of the normalised slow time t_s/T_0 for GP based digital twin (simultaneous mass and stiffness evolution) obtained from 150 noisy data. Noise levels of $\sigma_\theta = 0.005$, $\sigma_\theta = 0.015$ and $\sigma_\theta = 0.025$ are considered. Bayesian information criteria yields ‘Linear’ basis and ‘ARD Matern 5/2’ covariance kernel. The shaded plot depicts the 95% confidence interval.

5. Discussion

Although in theory, a digital twin of a physical system can be achieved in several ways, most of the existing works on digital twin have focused on the broader conceptual aspect. In this paper, we take a different path and focus on a specific case of a structural dynamical system. More specifically, we have considered surrogate models such as GP with a single degree of freedom system. The key ideas proposed in this paper include:

- We introduce the concept of a surrogate model within the digital twin technology. In

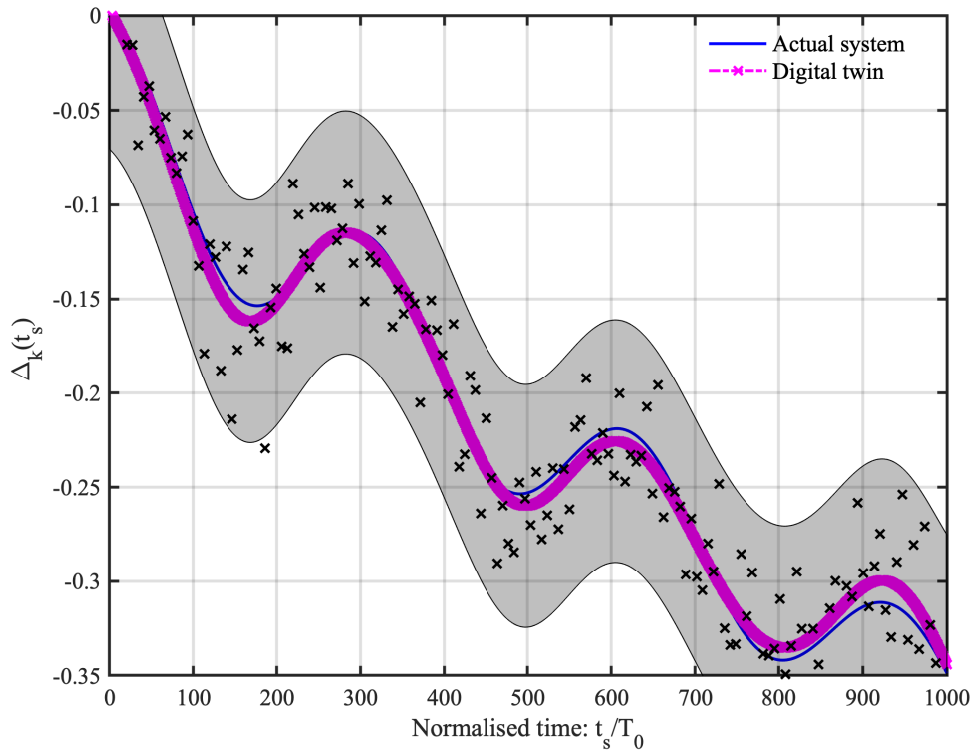


Fig. 12. Stiffness evolution as a function of the normalised slow time t_s/T_0 for GP based digital twin (simultaneous mass and stiffness evolution) trained with 150 noisy data with $\sigma_\theta = 0.025$. Bayesian information criteria yields ‘Linear’ basis and ‘ARD Matern 5/2’ covariance kernel. The shaded plot depicts the 95% confidence interval.

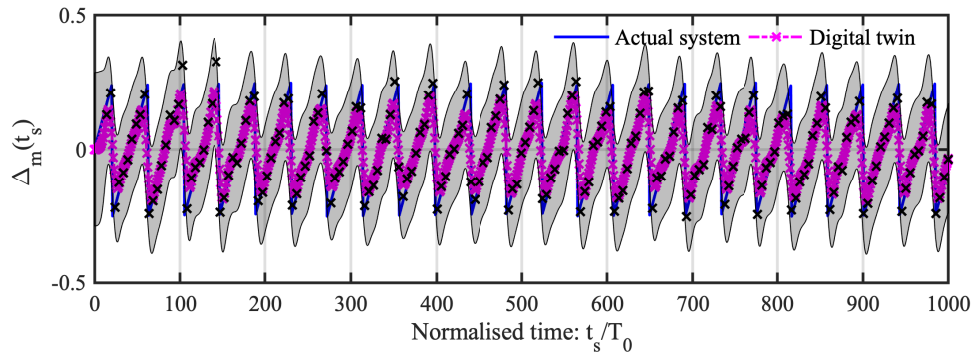


Fig. 13. Mass evolution as a function of the normalised slow time t_s/T_0 for GP based digital twin (simultaneous mass and stiffness evolution) trained with 200 noisy data with $\sigma_\theta = 0.025$. Bayesian information criteria yields ‘Linear’ basis and ‘ARD Matern 5/2’ covariance kernel. The shaded plot depicts the 95% confidence interval.

particular, we advocated the use of GP within the digital twin technology. The GP was trained based on observations on a slow time-scale and was used for predicting the model parameters in the fast time-scale.

- The importance of collecting more data during the system life-cycle is illustrated in this study. A dramatic improvement in the performance of the digital twin was

observed with an increase in the number of data collected. In other words, over the same window of time, a higher sampling rate is needed.

- The importance of cleaner data in the context of a digital twin is illustrated in this paper. With less noise in the data, the digital twin can quickly track the time evolution of the model parameters. Digital signal processing algorithms can be used for data cleaning [47, 48].
- GP, being a Bayesian surrogate model, can quantify the uncertainty in the system due to limited data and noise. These uncertainties can be used for deciding if there is a need to increase the frequency of data collection, i.e. increase the measurement sampling rate.
- Overall, GP was able to capture both mass and stiffness variation from limited noisy data; although, for capturing the time evolution of mass, more observations are needed as compared to the stiffness evolution case.

The system studied here is a simple single-degree-of-freedom dynamical system governed by second-order ordinary differential equations. The same framework, as it is, applies to other physical systems (e.g., simple electrical systems involving a resistor, capacitor and an inductor) governed by this kind of equations. The framework presented here can be extended for more rigorous investigations encompassing a wider variety of practical problems. Some of the future possibilities are highlighted below:

1. *Surrogate assisted digital twin for big data:* In this paper, natural frequency measurements are used for establishing the GP based digital twin. However, it will be more useful if the GP based digital twin can be trained directly from the time-history of measured responses. This is essentially a big-data problem and *vanilla* GP is unable to tackle such problems. More advanced versions of GP, such as the sparse GP [49] and convolution GP [50] may be explored in that case.
2. *Surrogate based digital twin for continuum system:* For many engineering problems, a continuum model represented by the partial differential equation is the preferred choice. Developing a surrogate-based digital twin for such a system will be of great use. This is primarily a sparse data scenario as sensor responses will only be available at few spatial locations. Methods such as convolution neural networks [51] can be used for such systems.
3. *Digital twin for systems with unknown/imperfect physics:* In literature, there exists a wide range of problems for which the governing physical law is not well defined. It will be interesting and extremely useful to learn/develop the digital twin purely based on data. Works on discovering physics from data by using machine learning can be found in the literature [52]. Genetic programming can also be used to fit mathematical functions to data [53].
4. *Surrogate based digital twin models with machine learning:* Over the past few years, the machine learning community has witnessed rapid progress with developments of techniques such as deep neural networks [19–21], convolution neural networks

[51], among others. Using these techniques within the digital twin framework can possibly push this technology to new heights. The required sampling rate and data quality could be reduced by using machine learning methods.

5. *Multiple-degrees-of-freedom (MDOF) digital twins using surrogate models:* In this paper, we focused on a single-degree-of-freedom (SDOF) model, which can be considered to be a simple idealisation of complex multiple-degrees-of-freedom (MDOF) systems. However, for more effective digital twin with realistic predictive capabilities, MDOF digital twin should be considered. To that end, it is necessary to develop surrogate-based MDOF digital twins.
6. *Surrogate models for nonlinear digital twins:* The model considered in this paper is a linear ordinary differential equation. However, many physical systems exhibit nonlinear behaviour. A classical example of a nonlinear system is the Duffing oscillator [54]. For such a non-linear system, it is unlikely that a closed-form solution will exist and hence, developing a surrogate-based digital twin model will be immensely useful.
7. *Forecasting using surrogate models for digital twins:* One of the primary tasks of a digital twin is to provide a future prediction. With the physics-based digital twin proposed previously [11], it is not possible to predict the future responses. The GP based digital twin model, on the contrary, can predict the short term future response. It is necessary to develop surrogate-based digital twin capable of forecasting the long-term response of the system.
8. *Predicting extremely low-probability catastrophic events for digital twins with surrogate models:* Till date, application of digital twin technology is mostly limited to maintenance, prognosis and health monitoring. Other possible applications of digital twin involve computing probability of failure, rare event probability and extreme events.
9. *High-dimensional surrogate models for multi time-scale digital twins:* In this paper, we assumed one time-scale for the evolution of the digital twin. However, there is no physical or mathematical reason as to why this must be restricted to only one time scale. It is possible that the parameters may evolve at different time-scales. Such systems are known as multi-scale dynamical system [55]. Developing surrogate-based digital twin for such multi-scale dynamical systems needs further investigation.
10. *Hybrid surrogate models for digital twins:* In this paper only one type of surrogate model, namely, the Gaussian process emulator was used. Several types of surrogate models have been employed to solve a wide range of complex problems with different numbers of variables [24, 53]. It is well known that some surrogate models perform better than the others in certain situations. It is therefore perfectly plausible to employ multiple surrogate models simultaneously by taking the benefit of their relative strengths. Such hybrid surrogate digital twins are expected to perform in a superior manner compared to single surrogate digital twins.

6. Conclusions

Surrogate models have been developed over the past four decades with increasing computational efficiency, sophistication, variety, depth and breadth. They are a class of machine learning methods which thrive on the availability of data and superior computing power. As digital twins also expected to exploit data and computational methods, there is a compelling case for the use of surrogate models in this context. Motivated by this synergy, we have explored the possibility of using a particular surrogate model, namely, the Gaussian process (GP) emulator, for the digital twin of a damped single-degree-of-freedom dynamic system. The proposed digital twin evolves at a time-scale which is much slower than the dynamics of the system. This makes it possible to identify crucial system parameters as a function of the ‘slow-time’ from continuously measured data. Closed-form expressions derived considering the dynamics of the system in the fast timescale have been employed. The Gaussian process emulator is employed in the slow timescale, where the impact of lack of data (sparse data) and noise in the data have been explored in detail.

The results arising from applying the Gaussian process emulator show that sparsity of data may prevent the GP from accurately capturing the evolution of mass and stiffness. This is particularly true for the mass evolution case where a higher sampling rate is necessary for accurately capturing the time evolution of mass. It is to be noted that the uncertainty in the GP based digital twin model results from both sparsity in the data and noise in the measurements. Consequently, only increasing the number of observations may not necessarily reduce the uncertainty in the system. On the other hand, collecting clean data (i.e., data with lower noise level) helps GP with tracking the evolution of mass and stiffness evolution more accurately. Nonetheless, even for the cases with a relatively larger noise level, GP is found to yield an accurate result. Moreover, GP also captures the uncertainty due to limited and sparse data.

Although only the Gaussian process emulator with single-degree-of-freedom dynamic systems are considered, several conceptual extensions directly followed from this work are described in detail. The approach proposed here exploits closed-form expressions based on the physics of the system and a surrogate model for the measured data. However, the proposed surrogate model based approach is perfectly suitable when more advanced computational simulation to be used for the physical problem (for example, when multi-degree-of-freedom system is considered). Therefore, the underlying formulation is not restricted to the single-degree-of-freedom example used here. The overall framework provides a paradigm for a hybrid physics-based and data-driven digital twin approach. The physics-based approach is related to the ‘fast time’, while the data-driven approach is operational on the ‘slow time’. The separation of scientific approaches based on the inherently different time-scales will allow the integration of multi-disciplinary approaches in the future development of the digital twin technology.

Appendix A. Bayesian information criteria

In this paper, the optimal order of basis function and the optimal covariance kernel for GP are determined by using the Bayesian information criteria. The Bayesian information

criteria corresponding to the m -th model is defined as [35]

$$bic_m = k_m \log(n) - \mathcal{L}(\hat{\theta}_m), \quad (\text{A.1})$$

where k_m is the number of parameters and n represents the number data-point available to the m -th model. $\mathcal{L}(\theta)$ in Eq. (A.1) represents the data-likelihood for the m -th model. The first term in Eq. (A.1) penalizes a complex model whereas the second term ensures that the model that best explains the model is selected. For GP, $\mathcal{L}(\theta_m)$ follows a multi-variate Gaussian distribution [41].

For determining the optimal model using the BIC criteria, models corresponding to all possible combinations from mean function and covariance function is considered. The possible candidates for mean and covariance functions are shown in Table 1. There are three possible mean function and 10 possible covariance functions. Therefore, total 30 models were considered in this study. The BIC score is evaluated for each of the 30 models and the one with the minimum BIC score is selected. The steps involve are shown in algorithm 1.

Algorithm 1: Model selection using Bayesian information criteria

- 1 **Initialize:** Provide the pool of mean function and covariance function. Also provide the training data-set, \mathcal{D} .
 - 2 Formulate all possible models (M) by using the pool of mean function and covariance function.
 - 3 **for** $m = 1, \dots, M$ **do**
 - 4 $\hat{\theta}_m \leftarrow \arg \max_{\theta} \mathcal{L}(\theta)$; ▷ Train GP
 - 5 Use $\hat{\theta}_m$ to compute $bic[m]$; ▷ Eq. (A.1)
 - 6 **end**
 - 7 Select the model with the lowest bic score.
-

Acknowledgements

RG acknowledges the financial support from The Royal Academy of Engineering through the award of a Distinguished Visiting Fellowship, grant number DVFS21819/9/5. SA acknowledges the financial support from The Engineering Physical Science Research Council (EPSRC) through a programme grant EP/R006768/1.

- [1] V. Souza, R. Cruz, W. Silva, S. Lins, V. Lucena, A digital twin architecture based on the industrial internet of things technologies, in: 2019 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2019, pp. 1–2.
- [2] P. D. U. Coronado, R. Lynn, W. Louhichi, M. Parto, E. Wescoat, T. Kurfess, Part data integration in the shop floor digital twin: Mobile and cloud technologies to enable a manufacturing execution system, *Journal of Manufacturing Systems* 48 (2018) 25–33.
- [3] E. Tuegel, A. Ingraffea, T. Eason, S. Spottswood, Reengineering aircraft structural life prediction using a digital twin, *International Journal of Aerospace Engineering* 2011 (Art. 154768).
- [4] F. Tao, F. Sui, A. Liu, Q. Qi, M. Zhang, B. Song, Z. Guo, S. C.-Y. Lu, A. Nee, Digital twin-driven product design framework, *International Journal of Production Research* 57 (12) (2019) 3935–3953.

- [5] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, F. Sui, Digital twin-driven product design, manufacturing and service with big data, *International Journal of Advanced Manufacturing Technology* 2018 (94).
- [6] S. Haag, R. Anderl, Digital twin – proof of concept, *Manufacturing Letters* 15 (2018) 64–66.
- [7] H. Millwater, J. Ocampo, N. Crosby, Probabilistic methods for risk assessment of airframe digital twin structures, *Engineering Fracture Mechanics* 221 (2019) 106674.
- [8] M. Zhou, J. Yan, D. Feng, Digital twin framework and its application to power grid online analysis, *CSEE Journal of Power and Energy Systems* 5 (3) (2019) 391–398.
- [9] F. Tao, H. Zhang, A. Liu, A. Nee, Digital twin in industry: state of the art, *IEEE Transactions on Industrial Informatics* 15 (4) (2018) 2405–2415.
- [10] C. Li, S. Mahadevan, Y. Ling, S. Choze, L. Wang, Dynamic bayesian network for aircraft health monitoring digital twin, *AIAA Journal* 55 (3) (2017) 930–941.
- [11] R. Ganguli, S. Adhikari, The digital twin of discrete dynamic systems: Initial approaches and future challenges, *Applied Mathematical Modelling* 77 (2020) 1110–1128. doi:10.1016/j.apm.2019.09.036.
- [12] D. Guivarch, E. Mermoz, Y. Marino, M. Sartor, Creation of helicopter dynamic systems digital twin using multibody simulations, *CIRP Annals* 68 (1) (2019) 133–136. doi:10.1016/j.cirp.2019.04.041.
- [13] S. Chakraborty, R. Chowdhury, Polynomial correlated function expansion, IGI global, 2016. doi:10.4018/978-1-5225-0588-4.ch012.
- [14] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, *Reliability Engineering & System Safety* 93 (7) (2008) 964–979.
- [15] D. Xiu, G. E. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM Journal on Scientific Computing* 24 (2) (2002) 619–644.
- [16] W. Zhao, T. Tao, E. Zio, W. Wang, A novel hybrid method of parameters tuning in support vector regression for reliability prediction: particle Swarm optimization combined With analytical selection, *IEEE Transactions on Reliability* (2016) 1–13.
- [17] S. R. Gunn, Support vector machines for classification and regression, Technical report, image speech and intelligent systems research group, Tech. rep., University of Southampton, Southampton, U.K. (1997).
- [18] J.-M. Bourinet, F. Deheeger, M. Lemaire, Assessing small failure probabilities by combined subset simulation and support vector machines, *Structural Safety* 33 (6) (2011) 343–353.
- [19] S. Goswami, C. Anitescu, S. Chakraborty, T. Rabczuk, Transfer learning enhanced physics informed neural network for phase-field modeling of fracture, *Theoretical and Applied Fracture Mechanics* 106 (2020) 102447.
- [20] S. Chakraborty, Simulation free reliability analysis: A physics-informed deep learning based approach, arXiv preprint arXiv:2005.01302.
- [21] S. Chakraborty, Transfer learning based multi-fidelity physics informed deep neural network, arXiv preprint arXiv:2005.10614.
- [22] I. Billionis, N. Zabaras, B. A. Konomi, G. Lin, Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification, *Journal of Computational Physics* 241 (2013) 212–239.
- [23] I. Billionis, N. Zabaras, Multi-output local Gaussian process regression: Applications to uncertainty quantification, *Journal of Computational Physics* 231 (17) (2012) 5718–5746.
- [24] S. Chakraborty, R. Chowdhury, Graph-theoretic-approach-assisted Gaussian process for nonlinear stochastic dynamic analysis under generalized loading, *Journal of Engineering Mechanics* 145 (12) (2019) 04019105. doi:10.1061/(ASCE)EM.1943-7889.0001685.
- [25] F. Wu, W. X. Zhong, A modified stochastic perturbation method for stochastic hyperbolic heat conduction problems, *Computer Methods in Applied Mechanics and Engineering* 305 (2016) 739–758.
- [26] F. A. DiazDelaO, S. Adhikari, Gaussian process emulators for the stochastic finite element method, *International Journal of Numerical Methods in Engineering* 87 (6) (2011) 521–540.
- [27] S. Chakraborty, R. Chowdhury, Hybrid framework for the estimation of rare failure event probability, *Journal of Engineering Mechanics* 143 (5) (2017) 04017010.

- doi:10.1061/(ASCE)EM.1943-7889.0001223.
- [28] V. Dubourg, Adaptive surrogate models for reliability analysis and reliability-based design optimization, Ph.D. thesis, Universite Blaise Pascal, Clermont-Ferrand, France (2011).
- [29] S. Goswami, S. Chakraborty, R. Chowdhury, T. Rabczuk, Threshold shift method for reliability-based design optimization, *Structural and Multidisciplinary Optimization* doi:10.1007/s00158-019-02310-x.
- [30] S. Rahman, D. Wei, Reliability-based Design Optimization by a Univariate Decomposition Method, in: 13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2010, pp. AIAA 2010-9037:1 – AIAA 2010-9037:14. doi:10.2514/6.2010-9037.
- [31] M. G. Kapteyn, K. E. Willcox, A. B. Philpott, Distributionally robust optimization for engineering design under uncertainty, *International Journal for Numerical Methods in Engineering*.
- [32] F. A. DiazDelaO, S. Adhikari, Structural dynamic analysis using gaussian process emulators, *Engineering Computations* 27 (5) (2010) 580–605.
- [33] F. A. DiazDelaO, S. Adhikari, Bayesian assimilation of multi-fidelity finite element models, *Computers and Structures* 92-93 (2) (2012) 206–215.
- [34] H. Dai, H. Zhang, W. Wang, A multiwavelet neural network-based response surface method for structural reliability analysis, *Computer-Aided Civil and Infrastructure Engineering* 30 (2) (2015) 151–162.
- [35] K. P. Murphy, *Machine learning: a probabilistic perspective*, MIT Press, 2012.
- [36] S. Biswas, S. Chakraborty, S. Chandra, I. Ghosh, Kriging-based approach for estimation of vehicular speed and passenger car units on an urban arterial, *Journal of Transportation Engineering, Part A: Systems* 143 (3) (2017) 04016013.
- [37] A. Saha, S. Chakraborty, S. Chandra, I. Ghosh, Kriging based saturation flow models for traffic conditions in indian cities, *Transportation Research Part A: Policy and Practice* 118 (2018) 38–51.
- [38] M. Alvarez, N. D. Lawrence, Sparse convolved gaussian processes for multi-output regression, in: *Advances in neural information processing systems*, 2009, pp. 57–64.
- [39] P. Perdikaris, D. Venturi, J. O. Royset, G. E. Karniadakis, Multi-fidelity modelling via recursive co-kriging and gaussian-markov random fields, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471 (2179) (2015) 20150018.
- [40] T. Mukhopadhyay, S. Chakraborty, S. Dey, S. Adhikari, R. Chowdhury, A critical assessment of kriging model variants for high-fidelity uncertainty quantification in dynamics of composite shells, *Archives of Computational Methods in Engineering* 24 (3) (2017) 495–518.
- [41] C. K. Williams, C. E. Rasmussen, *Gaussian processes for machine learning*, Vol. 2(3), MIT press Cambridge, MA, 2006.
- [42] C. E. Rasmussen, H. Nickisch, Gaussian processes for machine learning (gpml) toolbox, *Journal of machine learning research* 11 (Nov) (2010) 3011–3015.
- [43] S. Sony, S. Laventure, A. Sadhu, A literature review of next-generation smart sensing technology in structural health monitoring, *Structural Control and Health Monitoring* 26 (3) (2019) e2321.
- [44] Q. Wang, J. Huang, Q. Liu, Z. Zhou, Dynamic strain measurement of hydraulic system pipeline using fibre bragg grating sensors, *Advances in Mechanical Engineering* 8 (4) (2016) 1687814016645069.
- [45] M. Pelikan, D. E. Goldberg, E. Cantú-Paz, Boa: The bayesian optimization algorithm, in: *Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1*, Morgan Kaufmann Publishers Inc., 1999, pp. 525–532.
- [46] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, *Sensors* 17 (2) (2017) 425.
- [47] N. Roy, R. Ganguli, Helicopter rotor blade frequency evolution with damage growth and signal processing, *Journal of Sound and vibration* 283 (3-5) (2005) 821–851.
- [48] R. Ganguli, Noise and outlier removal from jet engine health signals using weighted fir median hybrid filters, *Mechanical Systems and Signal Processing* 16 (6) (2002) 967–978.
- [49] E. Snelson, Z. Ghahramani, Sparse gaussian processes using pseudo-inputs, in: *Advances in neural information processing systems*, 2006, pp. 1257–1264.
- [50] M. A. Álvarez, N. D. Lawrence, Computationally efficient convolved multiple output gaussian processes, *Journal of Machine Learning Research* 12 (May) (2011) 1459–1500.

- [51] Y. Kim, Convolutional neural networks for sentence classification, arXiv preprint arXiv:1408.5882.
- [52] H. Xu, H. Chang, D. Zhang, Dl-pde: Deep-learning based data-driven discovery of partial differential equations from discrete and noisy data, arXiv preprint arXiv:1908.04463.
- [53] A. P. Singh, V. Mani, R. Ganguli, Genetic programming metamodel for rotating beams, *Computer Modeling in Engineering and Sciences* 21 (2) (2007) 133.
- [54] A. H. Nayfeh, *Introduction to Perturbation Techniques*, John Wiley & Sons, New York, NY, 1993.
- [55] S. Chakraborty, N. Zabaras, Efficient data-driven reduced-order models for high-dimensional multi-scale dynamical systems, *Computer Physics Communications* doi:10.1016/j.cpc.2018.04.007.