K. Worden
Professor
Dynamics Research Group,
Department of Mechanical Engineering,
University of Sheffield,
Sheffield S1 3JD, UK
e-mail: k.worden@sheffield.ac.uk

E. J. Cross
Dynamics Research Group,
Department of Mechanical Engineering,
University of Sheffield,
Sheffield S1 3JD, UK
e-mail: e.j.cross@sheffield.ac.uk

R. J. Barthorpe
Dynamics Research Group,
Department of Mechanical Engineering,
University of Sheffield,
Sheffield S1 3JD, UK
e-mail: r.j.barthorpe@sheffield.ac.uk

D. J. Wagg
Professor
Dynamics Research Group,
Department of Mechanical Engineering,
University of Sheffield,
Sheffield S1 3JD, UK
e-mail: david.wagg@sheffield.ac.uk

P. Gardner
Dynamics Research Group,
Department of Mechanical Engineering,
University of Sheffield,
Sheffield S1 3JD, UK
e-mail: p.gardner@sheffield.ac.uk

On Digital Twins, Mirrors, and Virtualizations: Frameworks for Model Verification and Validation

A powerful new idea in the computational representation of structures is that of the digital twin. The concept of the digital twin emerged and developed over the last decade, and has been identified by many industries as a highly desired technology. The current situation is that individual companies often have their own definitions of a digital twin, and no clear consensus has emerged. In particular, there is no current mathematical formulation of a digital twin. A companion paper to the current one will attempt to present the essential components of the desired formulation. One of those components is identified as a rigorous representation theory of models; most importantly, governing how they are verified and validated, and how validation information can be transferred between models. Unlike its companion, which does not attempt detailed specification of any twin components, this paper will attempt to outline a rigorous representation theory of models, based on the introduction of two new concepts: mirrors and virtualizations. The paper is not intended as a passive wish list; it is intended as a rallying call. The new theory will require the active participation of researchers across a number of domains including: pure and applied mathematics, physics, computer science, and engineering. The paper outlines the main objects of the theory and gives examples of the sort of theorems and hypotheses that might be proved in the new framework. [DOI: 10.1115/1.4046740]

1 Introduction

The *digital twin* has emerged in the last two decades as a highly sought-after generalization of the computation models routinely used by industry and academia in attempts to understand the behavior of real structures, systems, and processes and to make predictions in previously unseen circumstances [1–3]. There is currently no real consensus on what the necessary and sufficient ingredients of a digital twin are, although a sister paper to this one [4] will attempt to bring some order to the subject. What is inarguable is that because the digital twin extends the concept of a computational model, such a model must be a core ingredient. Furthermore the model must be *validated*; it must be demonstrated to be in correspondence with reality, at least in the context of immediate engineering importance. Because of the problems which a digital twin will be required to address, it will also potentially need to extrapolate or generalize to predictions on different structures or the same structure in different contexts. This paper will argue that, in order to ensure the correct operation of digital twins, a mathematical framework is needed in order to quantify the likely fidelity of validated models when used to generalize or extrapolate. This paper will propose that what is needed is a type of *algebra* of models, which can be used in order to extend current concepts of verification and validation (V&V).

For the purposes of this paper, the fundamental problem of V&V will be regarded as the need to answer two questions:

(1) What is the lowest-cost model that will allow predictions of the required accuracy for the structure of interest in the context of interest?
(2) What is the lowest-cost program of experimental testing that will validate the model with prescribed confidence?

Note that in answering these questions, one does not need a model that represents the whole structure across its entire range of possible behaviors; one only needs a model that matches in the *context of interest*.[1] In a machine learning context, the question is essentially of generalization; having learned from model data, can one say something meaningful about the structure twinned with the model?

In order to establish an overarching mathematical framework, one needs to be precise and meaningful in one's terminology. The use of the term "twin" is inconsistent with this goal for two reasons; the first is that there is already widespread and disparate use of the term in the engineering community; the second is that it does not really make sense as an analogy anyway (most twins are not identical). The view taken in this paper will be that a more meaningful term is provided by the word *mirror* (this terminology within a digital modeling context was similarly proposed by Tao et al. in Ref. [5]). A mirror is an instrument that faithfully reflects

Manuscript received April 30, 2019; final manuscript received March 2, 2020; published online May 12, 2020. Assoc. Editor: Sifeng Bi.

[1]Some would argue that a true "digital twin" has to match the structure of interest in *all* contexts. This viewpoint does not make complete sense, as the physics of a given structure is unlikely to be known at all scales and in all contexts; this means that modeling would not be possible. Furthermore, a lot of the motivation for digital twins comes from industry, and it is not conceivable that a profit-driven enterprise would require a model to function outside the immediate context of interest if that extended functionality came at an increased cost.

reality in terms of the aspects of an object that are *mirror-facing*; it provides no "information" about aspects that are not mirror-facing. The idea of "mirror-facing" will be formalized in the following as a *context*. Finally, if the object moves, the movement will be reflected perfectly, in the mirror—at least as far as those aspects that are mirror-facing. This paper then will attempt to motivate a mathematical basis for understanding *mirrors*.[2] As such, it will have the opportunity to develop independently of current conceptions as to what a 'digital twin' is, but leaving the possibility for engineers to adopt the technology in developing whatever their favored definition of a digital twin actually is.

Everything here is motivated by the desire to construct meaningful validated models of structures and systems; if one were to do nothing more than rearrange the terminology and dress the problem in pretty mathematical trappings, then that would be ultimately empty. This paper is motivated by the belief that a general mathematical theory of models and their validation will be of value; however, this paper will not be able to go beyond development of the basic terminology and theory and some attempts to convince the reader of the ultimate possibilities. One might argue that general frameworks have already been proposed in terms of the formulation and evaluation of models, and that there is no need to propose another one until the existing ones have been fairly evaluated. This is a fair point; however, the authors here would argue that the current proposal is more sympathetic to the needs of the digital twin concept, because of the explicit attention given to context and environment. There is no intention here to play down any previous works on general methodologies; the assumption is that the tools already proposed will play important roles. One example of a general framework for V&V is provided in Ref. [6]. That publication provides a methodology for estimating the uncertainty in system-level predictions, where system-level parameters are estimated in terms of lower-level experiments. The paper is largely concerned with calibration and uncertainty propagation, and introduces tools for estimating the reliability of models. Perhaps more importantly for the current discussion, the paper introduces a concept of "relevance" which quantifies the relationship between the system-level model and lower-level models, and potentially allows a "confidence" measure in terms of extrapolating from lower levels to the system level. The paper by Nagel and Sudret [7] proposes a Bayesian unified framework which provides a "… toolkit for statistical model building. It forms some kind of superstructure that embeds a variety of stochastic inverse problems as special cases." (There are of course, many other papers one could cite; however, there is no intention here to provide a survey.) Another fair criticism of this paper is that the new term "mirror" is not needed either, it refers simply to a validated model; however, it is introduced here because it refers to a specific class of models and because, as discussed previously, there is a need to distinguish the idea from the more overarching digital twin. The relationships and distinctions between "mirrors" and "twins" will be highlighted throughout the paper.

The layout of the paper is as follows: Section 2 will make the main series of definitions of the important concepts in the framework: contexts, mirrors, etc. The section will also define the concepts of *environments* and *virtualizations* which are central to the idea of a digital twin. Section 3 will discuss a number of example problems in which the idea of a mirror would be fruitful, assuming that the appropriate mathematical underpinnings of the theory can be provided. Section 4 ends the paper with some discussion and conclusions.

[2]The term *digital mirror* is already in use to define an item of technology; the items being exactly what one might imagine them to be. One could use the term with complete confidence that the two meanings are unlikely to be confused; however, for simplicity the objects of interest will just be referred to as "mirrors," although different kinds of mirrors will be introduced.

2 Mirrors

2.1 Basic Definitions. To start with the *simplest* situation, the discussion will initially consider only physics-based models; data-based and hybrid models[3] will be brought in later.

One must begin with a structure (or system) S; this is the *physical* object of interest. It will be interpreted as having an objective reality independent of its surroundings, i.e., it is possible to think of it in a vacuum remote from any other matter. Temporal changes in the confirmation and behavior of the structure will be summarized in a *state vector* $\underline{s}(t) = \{s_1(t), \ldots s_{N_S}(t)\}$, which consists of a set of N_S instantaneous measurements (at time t) which completely characterize its state.

Now, the environment of the structure could be considered as all physical reality exterior to it; however, that is too general. Considering the fact that the environment could also be characterized by a state vector, the *environment* E of S will be defined as the set of environmental variables that can actually affect S, i.e., a change in variable will evoke a change in the state $\underline{s}(t)$. With this in mind, one will have an environmental state vector $\underline{e}(t) = \{e_1(t), \ldots e_{N_E}(t)\}$.

Recognizing that one will generally only wish to model some aspects of the behavior of S, a *context* C for S will be defined as a set of environmental state variables $C = \{e_i^C \in E, s_j^C \in \underline{s}; i, j\}$. The subset $\{e_i^C\}$ will be referred to as the *environmental context*, and the subset $\{s_j^C\}$ as the *response* or *predictive context*.

Now, a *schedule* W_C for the context C will be a set of time series $\{\underline{e}_W^C(t_i); i = 1, \ldots N_t; t_i \in [0, T]\}$. (In principle, the set $\{t_i\}$ could be continuous or discrete.) The response $\underline{r}_W^C(t)$ to a schedule W_C is defined as the measurement sequence resulting from testing the structure and imposing the schedule as inputs. As the process will generally be dynamic, it will be denoted by the functional,

$$\underline{r}_W^C(t) = S[\underline{e}_W^C(t) \equiv W_C] \tag{1}$$

where the notation S is used again to indicate that the functional is identified with the physical structure of interest.

One can now define the *test* T_W^C associated with the schedule W_C in the context C, as the set $T_W^C = \{\underline{e}_W^C, \underline{r}_W^C\}$. In general, tests will be carried out for multiple purposes; for the moment, it will be observed that data are captured for training of models and for testing of models. For this reason, it is useful to divide data accordingly. Supposing that tests have been carried out multiple times, one can define the *training schedule* (respectively *testing schedule*) as the set of schedules associated with acquiring data for training (respectively testing); the set being denoted by D_{tr} (respectively D_t). (Of course, these sets are specific to a context and a schedule, but the notation will become too unwieldy if this is made explicit.)

Now, a *model* of S for a context C will be defined as a mathematical function M^C which attempts to predict the behavior of S for any schedule specific to the context C. Depending on the environmental and predictive variables, this may be a multiscale and/or multiphysics model, and it will almost always be implemented in computer code in some appropriate language.[4] A *simulation* for a context C under a schedule W_C is then defined as,

$$\underline{m}_W^C(t) = M^C[\underline{e}_W^C(t) \equiv W_C] \tag{2}$$

Now, it is clear that one can obtain the simulation $m_i^C(t)$ corresponding to a test $T_i^C = \{\underline{e}_i^C, \underline{r}_i^C\}$ (with i now a schedule label), so

[3]Hybrid models are also referred to in the literature as *gray-box* or *data-augmented* models. In the statistics literature, the addition of a data-based model in order to correct a physics-based model is commonly called model *bias correction* or *model discrepancy* correction; the most influential framework is probably that proposed in Ref. [8].

[4]In fact, it may be the case that different models are needed in order to completely cover the context of interest. For notational simplicity, it is assumed here that M^C represents the set of relevant models, returning the values required by the overall context C; there is no overall loss of generality at this point.

that one can attempt to assess the fidelity of the model by comparing its predictions to reality.

A *metric* on a given context C will be defined here simply as a function $d^C(\underline{x},\underline{y})$ such that $d^C(\underline{x},\underline{y}) \geq 0$, with the zero only if $\underline{x} = \underline{y}$. (This is only one of the conditions for a true mathematical *metric*, but it will do here for now).

Finally, the main definitions of the paper are possible

DEFINITION 2.1. *(ε-Mirror) A model M_ε^C for a given context C is an ε-mirror if and only if,*

$$d^C(\underline{m}^C(t), \underline{r}^C(t)) \leq \varepsilon \tag{3}$$

for all scheduled tests in D_t.

DEFINITION 2.2. *(Fitness-for-purpose) A model M_ε^C is fit-for-purpose in a given context C iff it is an ε-mirror for C and $\varepsilon \leq \varepsilon_T$ where ε_T is a critical threshold based on engineering judgment and/or context requirements.*[5]

2.2 Hybrid Models and Uncertainty. So far, only pure physics-based models have been considered; models sometimes termed *white-box models*. At the other end of the modeling spectrum are *black-box models* which are formed by taking a model basis with a universal approximation property, and tuning the parameters of the model to a set of observed data; examples of such models are artificial neural networks or support vector machines [9,10]. One can also make use of *hybrid* or *gray-box* models, which combine some element specified by physics with an element of learning from data.

Suppose that it is desirable or necessary to form or update a model based on data. The model will be established using data acquired from a training schedule D_{tr} and tested on data from a test schedule D_t.[6] The resulting model $M^{hC}(D_{tr})$ is then an ε-mirror if it satisfies the conditions of Definition 2.1 on D_t. The model $M^{hC}(D_{tr})$ is adapted to the measured data D_{tr}, and is thus now a hybrid model as indicated by the symbol h; the context does not change.

There is no distinction here on how $M^{hC}(D_{tr})$ is obtained. One might start with a white-box model and learn the parameters via system identification, or one might adopt a gray-box structure where a physics-based model is augmented with a nonparametric machine learner [11]

As the use of machine learning has been raised, it would seem to be an appropriate point to discuss *uncertainty*; this is because many modern machine learning algorithms are probabilistic and accommodate uncertainty directly. For example, Bayesian approaches to parameter estimation can characterize the entire density functions of parameters, rather than simply producing point estimates [12,13]. Furthermore, nonparametric learners like Gaussian process regression can produce a natural confidence interval on predictions [14].

So, under the circumstances, one might allow the possibility that the model $M^{hC}(D_{tr})$ is a function that returns a random variable, i.e., the simulation responses are stochastic processes,

$$M_t^C = M^{hC}[\underline{e}_W^C(t)](D_{tr}) \tag{4}$$

The simulation might provide the whole density function for M_t^C, or just low-order moments. In the first case, suppose that the model returns the predictive mean of the process $\bar{m}^C(t) = E[M_t^C]$ (where E is an expectation), then $\bar{m}^C(t)$ can be used to determine whether $M^{hC}(D_{tr})$ is an ε-mirror *in the mean*.

Alternatively, suppose that the model returns enough information to determine confidence intervals on the prediction. In this case, then if $\underline{r}^C(t) \in [\bar{m}^C(t) - \alpha\sigma_r^C(t), \bar{m}^C(t) + \alpha\sigma_r^C(t)]$ with probability determined by α, and for all schedules in D_t, then one can define $M^{hC}(D_{tr})$ as an *α-mirror*. Note that a given stochastic model can be both an ε-mirror and an α-mirror.

It would be possible to define various metrics for comparison in the uncertain case; the one based on low-order moments described previously is related to the reliability metric discussed in Ref. [6], or using the Mahalanobis distance as in Ref. [15], which is in turn related to a formulation of validation as an outlier analysis problem, as discussed in Ref. [16]. If the comparison were made on the whole predictive or parameter density functions, i.e., the scenario in which the predictive distribution is compared to observational distribution (often via a finite sample set), one might define a statistical distance (or divergence) measure [15,17], for example, a Hellinger distance, leading to the definition of an α-mirror as a Hellinger-mirror, etc.

2.3 The Environment and Virtualization. Raising the question of uncertainty means that one must reconsider the status of the *environment*.

Recall that the environment is comprised of all those variables which can have a causal influence on S, the structure of interest. In general, this set will be composed of variables that can be controlled (e.g., forces applied to the structure) and variables that cannot (or cannot be controlled with any precision). In an operational modal analysis context, for example, even the forces may not be controllable. It is therefore necessary to separate the variables (in context) accordingly into $\underline{e}_u^C$ and $\underline{e}_c^C$ (uncontrolled and controlled, respectively). This distinction is very important if one wishes to use the model to make true predictions, i.e., to determine what the structure might do at some point in the future, under a given (controlled) forcing, but when the $\underline{e}_u^C$ are unknown.

In this situation, what is needed is a generative model M_u^{EC}, that will make some best estimate of $\underline{e}_u^C(t)$,

$$\hat{\underline{e}}_u C = M_u^{EC}(t) \tag{5}$$

This model itself will need to be validated appropriately, as far as possible. Given training data for the M_u^{EC}, it might be possible to establish a nonparametric black-box model that is an ε- or α-mirror, or one could substitute mean values for the variables and treat variations as uncertainty that needs to be propagated. In any case, one can now make predictions (in the given context),

$$\underline{p}^C(t) = M[\underline{e}_c^C(t), \hat{\underline{e}}_u C = M_u^{EC}(t)] \tag{6}$$

It is now possible to make another important definition: a *virtualization* for a given context C is a pair,

$$V^C = (M_{\varepsilon_1}^{hC}, M_{\varepsilon_2}^{EC}) \tag{7}$$

where the two models concerned are ε-mirrors with the fidelities specified. The importance of the virtualization is that it can be used to examine what-if scenarios for the structure of interest in previously unseen circumstances. Of course, one can make a similar definition with α-mirrors. Finally, it is important to note that a virtualization is itself a model, and as such can also be an ε- or α-mirror; this will prove to be of interest later, when the use of virtualizations for design is discussed.

The problem of the "environment" is discussed in Ref. [7]; however, there it appears to have been condensed into the estimation/calibration of a further parameter set.

2.4 The Turing Mirror. One can also think of a semiphilosophical means of defining a mirror; this parallels the *Turing test* in the field of artificial intelligence, which is a test of the ability of a machine to perform in a manner indistinguishable from a human [18].

[5]The standard mathematical notation is adopted here, where "iff" is taken to mean "if and only if."

[6]Following the best practice in machine learning, different datasets are potentially required in order to fit parameters and establish hyperparameters [9]. In order to keep the notation simpler here and avoid confusion about the term "validation," it is assumed that the modeler simply partitions D_{tr} appropriately.

The test will involve two protagonists: an *interrogator* and an *oracle*. The two people can only interact in a very limited way, the interrogator is allowed to present questions to the oracle about the structure of interest via a remote interface. The oracle is equipped with a model of the structure of interest, which is the candidate mirror, and also has facilities for carrying out physical testing on the structure. The interrogator is allowed to present the oracle with a set of schedules $\underline{e}_W^C$ from some given context, and the oracle is required to return either the test responses of the structure $\underline{r}_W^C$, or simulations from the model $\underline{m}_W^C$.[7] If the interrogator is unable to decide which option the oracle has taken in any case, then the model in question is a *Turing-mirror* or *T-mirror*.

While this may seem like nothing more than an amusing digression, there is the possibility that the work over the years in terms of implementing the Turing test could be used in order to derive rigorous methods of testing mirrors.[8]

2.5 Transfer Learning and Mirrors. The problem of generating mirrors lends itself to being formulated in terms of transfer learning problems. Although there are various techniques that could provide solutions to the mathematical framework proposed here, transfer learning provides a potential approach for addressing these challenges. Throughout the example sections in this paper, each problem will also be formulated using transfer learning. For this reason, general definitions about transfer learning are provided [20–22].

First, one must define two key quantities: a *domain* and a *task*.

DEFINITION 2.3. *A domain $\mathcal{D}$, consists of a feature space $\mathcal{X}$ and a marginal probability distribution $\mathcal{D} = \{\mathcal{X}, p(X)\}$ where $X = \{\underline{x}_i\}_{i=1}^N \in \mathcal{X}$,* i.e., *a finite sample set from $\mathcal{X}$.*

DEFINITION 2.4. *A task $\mathcal{T}$, for a given domain, defines an output space*[9] *$\mathcal{Y}$ and predictive function $f(\cdot)$ learnt from a training* dataset *$\{\underline{x}_i, y_i\}_{i=1}^N$, where $\underline{y} \in \mathcal{Y}$,* i.e., *$\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$.*

With these definitions, transfer learning can be defined as:

DEFINITION 2.5. *… the process of utilizing knowledge about a source domain $\mathcal{D}_s$ and task $\mathcal{T}_s$ to improve the learning of a target predictive function $f_t(\cdot)$ for a given target domain $\mathcal{D}_t$ and corresponding task $\mathcal{T}_t$, where $\mathcal{D}_s \neq \mathcal{D}_t$ and $\mathcal{T}_s \neq \mathcal{T}_t$.*

Transfer learning methodologies then attempt to solve problems where different information is available or not, i.e., $\mathcal{X}$, $p(X)$, $\mathcal{Y}$ or $p(\underline{y}|X)$ are consistent across the source and target [21].

To illustrate transfer learning concepts, a descriptive example is provided (although further illustrations are presented throughout the paper). Typically within model validation, a computer model M may be established as an ε-mirror for some context C_1, given some measured response $\underline{r}^{C_1}(t)$ from the physical structure S. Often an engineer wishes to repurpose the model for some new context C_2, which differs from the original context C_1. In this scenario, the model simulation $\underline{m}^{C_1}(t)$ and structural response $\underline{r}^{C_1}(t)$ for context C_1 form a source domain and task, as the predictive function from the model simulation $\underline{m}^{C_1}(t)$ to the structural response $\underline{r}^{C_1}(t)$ has been established in calculating that it is an ε-mirror. The target domain and task are the model simulation $\underline{m}^{C_2}(t)$ and structural response $\underline{r}^{C_2}(t)$ for context C_2, where typically $\underline{r}^{C_2}(t)$ is not known (at least before experimental testing has been performed). Transfer learning in this scenario seeks to leverage the knowledge from context C_1 to estimate the expected response $\underline{r}^{C_2}(t)$ in context C_2, therefore establishing a bound on ε for C_2.

3 Examples

3.1 Examples Concerning Context Change. One of the simpler problems one can imagine in the context of mirrors is how to analyze the performance of a given model, when asked to make predictions outside its original context C. This problem is interesting because it can be made to include the case of *extrapolation*, although that will not be discussed in great detail here. Extrapolation for a data-based or hybrid model occurs, when the model $M^{hC}(D_{tr})$ is used to make predictions outside the range of data encompassed by the training set D_{tr}. Even if the model $M^{hC}(D_{tr})$ is an ε-mirror on schedules in the training set, this may not hold if the model extrapolates. Likewise, for a white-box model the inferred parameters in context C may not be optimal for a new context C'. One simple way to make the problem of context change encompass the problem of extrapolation would be to extend the definition of context C, so that it not only specifies the variables under investigation, but also the ranges of those variables encountered in training data.

This example will consider a different problem, where a model M_ε^C is required to make predictions on different variables to its context C. Suppose the model is modified in order to predict in a context C', with the new model denoted $M'^{C'}$. Furthermore, assume that there are no training or test data available for the context C'. The interesting question is:

Given that a model M^C is an ε-mirror for the context C; following modification to $M'^{C'}$, is the new model an ε'-mirror for C' for any ε', and if so, what is the minimum value of ε' for which this holds? (Note that, with the extended definition of context discussed above, this is the extrapolation problem if $M = M'$).

Consider a simple example. Suppose one has constructed a finite element (FE) model M^C, of a cantilever beam (as in Fig. 1). The model has been validated on test data measured as the acceleration responses $\ddot{y}_i(t)$ at points $i = 1, 4$, so that the predictive context is $\{\ddot{y}_1, \ddot{y}_4\}$. Suppose that M^C has been established as an ε-mirror on the context C. Now, further suppose that one wishes to make predictions of the response at points 2, 3, 5, and 6, so the predictive context for C' is $\{\ddot{y}_2, \ddot{y}_3, \ddot{y}_5, \ddot{y}_6\}$. In this situation, there are two simple ways to establish M':

The trivial approach is to simply change the output deck of M^C, so that the model outputs the required variables (if it didn't before).

One can add a numerical interpolation step to the process in order to estimate the variables in C' from those in C.

In the first case, it should be a fairly straightforward matter to establish that the model is an ε'-mirror based on the existing theory of error estimates for FE models [23,24], and one would expect that $\varepsilon' \approx \varepsilon$. In the second case, one should be able to use error estimates from the numerical analysis of interpolation, combined with some reasonable assumptions about the continuity of the beam profile. One could also bound the errors based on much coarser assumptions, e.g., one could estimate how far $\ddot{y}_3$ could get from $\ddot{y}_1$ and $\ddot{y}_4$ before the induced stresses in the beam exceeded the yield stress. Although the latter approach would likely work, it would probably yield an $\varepsilon' \gg \varepsilon$, so conservative that one would find the value impractical in terms of model trust. In an exercise like this, the objective would be to find the lowest bound on ε' possible.

Another viewpoint for solving this problem, in which one wishes to known the ε' for $M'^{C'}$ where the outputs are the local stresses σ_1 and σ_4 instead of $\ddot{y}_1$ and $\ddot{y}_4$ from the validated ε-mirror, is to think of it in the context of transfer learning. Here the objective would be to use knowledge about the ε-mirror M^C, and the structure S^C, to create a mapping to the unknown stress outputs for $S'^{C'}$. In a transfer learning setting, the source domain would be the acceleration outputs from M and S, where the known

[7] Clearly, there are subtleties. For example, if the necessary test program in a given case were to take 10 days, while running the model would only take 10 hours, the oracle would only return the results after the greater time.

[8] A very close variant of this Turing test is proposed in Ref. [19]; however, the "Grieves test" as it is called there, fails to make precise the details of how computer models are incorporated.

[9] An output space $\mathcal{Y}$ traditionally refers to a label space within the machine learning transfer learning literature [20–22]. In this context, transfer learning is used to aid classification tasks where the output space is the set of possible labels from the feature data. In this paper, an output space will typically refer to output quantities from a model, e.g., the output space for a dynamical system could be a space of frequency response functions.

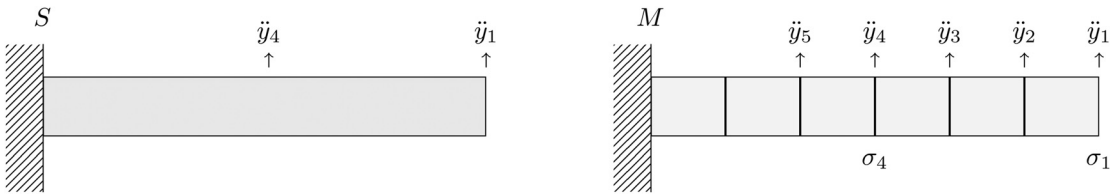

Fig. 1 Output context change illustration using a cantilever beam, where the structure *S* is left and the FE model *M* right

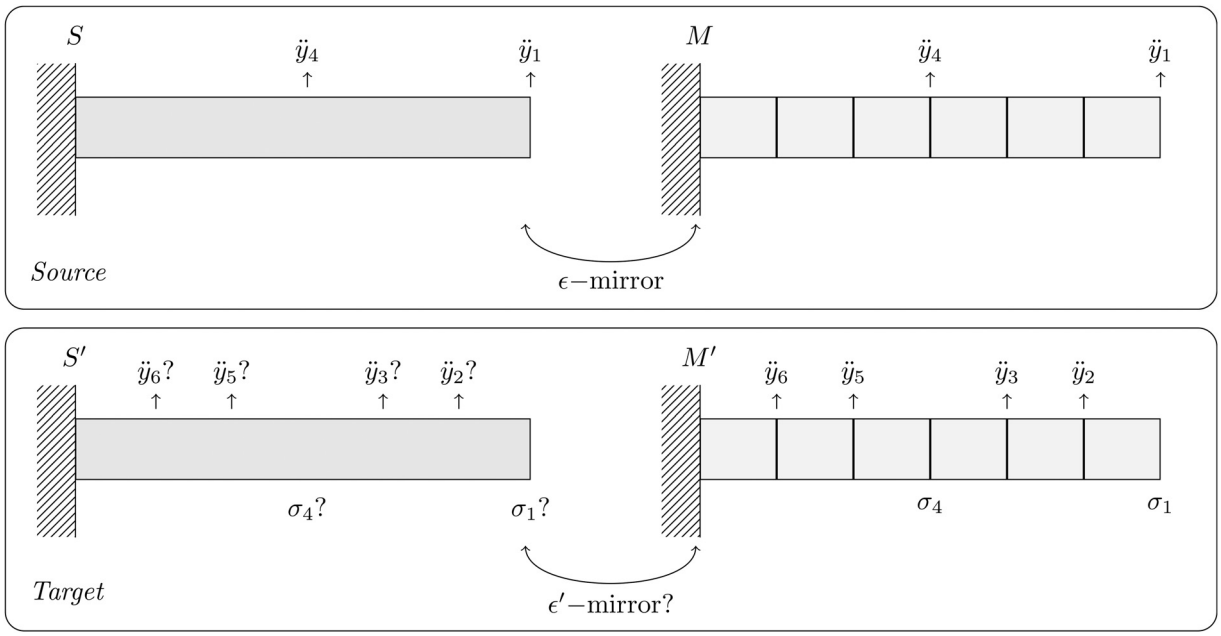

Fig. 2 Output context change illustration using cantilever beam. Example of transfer learning problem setup for predicting acceleration outputs at different locations and stress outputs.

information in the target domain is the stress output data from M', as shown in Fig. 2. By learning the nonlinear mapping to the stress outputs from S', it would be possible to find a bound on ε'.

A more interesting problem arises in the case of the extended definition of context to account for input changes. Suppose C covered points 1 and 4 at low levels of excitation, and C' covered points 2, 3, 5, and 6 at a higher level of excitation; there would be two different answers to this question, depending on whether M^C was linear or nonlinear.

3.2 An Example Concerning Assembly. This example concerns a very important objective of any program of "virtualization". Suppose one could validate a model of a full-scale assembled structure using only test data acquired from substructure testing. The cost savings in the design/production cycle would be potentially very high. It is important that the "algebra" of models being developed covers this situation, and this will entail an understanding of how to model joints and joining processes.

For the sake of simplicity, consider the case of two substructures (but note that this is not a real restriction, as the substructure assembly can be considered recursively). The substructures, denoted S_1 and S_2, will be assumed to have individual contexts C_1 and C_2, respectively. It will be assumed that the substructures will be joined using some technology, which can itself be modeled; in the general case, one assumes that the joint may itself be a substructure S_J. With a small abuse of mathematical notation, the assembled structure S_A will be denoted by,

$$S_A = S_1 \underset{S_J}{\oplus} S_2 \tag{8}$$

For simplicity, it will be assumed that all the responses from the substructures can still be measured; in this case, one can denote the new context by $C_A = C_1 \oplus C_2$. (Here, the $\oplus$ is largely just a direct sum with some reordering of symbols and deletion of copies of symbols that appear in the environment context twice.) In general, one would have to allow for the fact that the joining process might eliminate a possible measurement point on the substructure, and thus change the context by removing a variable.

It is assumed that each substructure S_i has a model M_i^C associated with it, and that the models have been validated using test data from the individual structures, and it has been established that M_i^C is an ε_i-mirror in each case. Furthermore, assume that the joint/joining process has a model M_J, and that this model may or may not have been validated. The model of the assembled structure is denoted,

$$M_A = M_1 \underset{M_J}{\oplus} M_2 \tag{9}$$

where the appropriate contexts C_A, C_1, and C_2 have been omitted to improve clarity of the expression.

The key question is now:

Given the assumptions stated, is it possible to show that there exists any ε_A such that M_A is an ε_A-mirror for S_A in the context C_A,

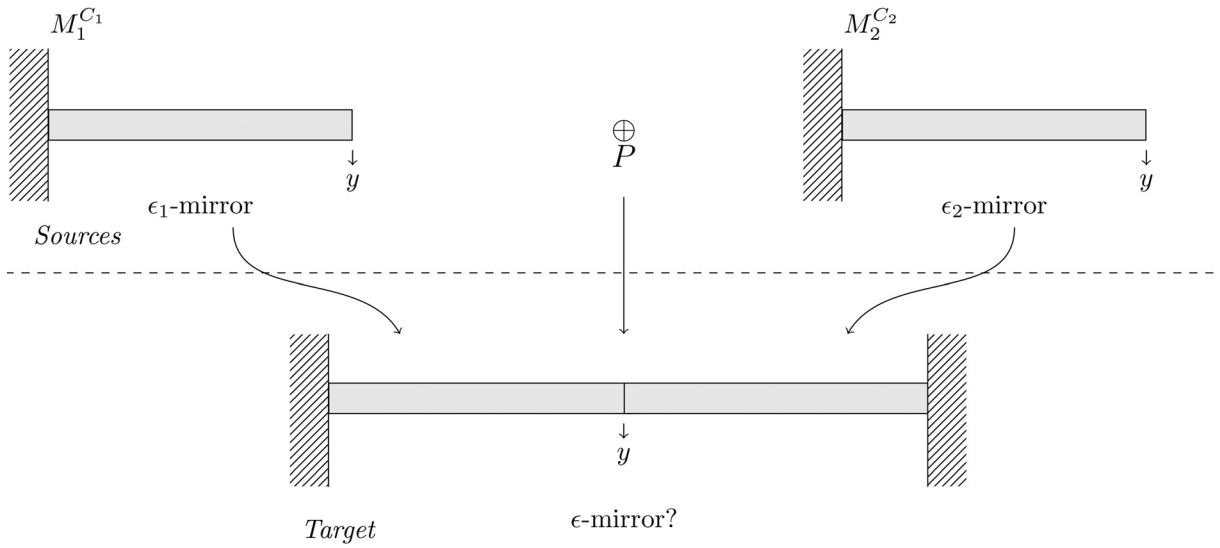

Fig. 3 Encastré beam as sum of two cantilevers and a perfect joint

in the absence of any test data for the assembly S_A? If so, then what is the smallest ε_A for which this is true?

Of course, one could also attempt to accommodate uncertainty, and frame the question in terms of α-mirrors (as discussed in Sec. 2.2). This is the most difficult question so far, but it also offers the highest returns, if it can be answered. The problem also depends on whether a validated model for M_J is available. For example, consider the case when the joint is a weld, and that coupon tests have established some of the material properties of the weld material (perhaps with a high degree of uncertainty). Even allowing for the fact that the issue is not just about material properties, one would expect ε_A to be a monotonically increasing function of the weld parameter uncertainties. One might also model the weld as a hybrid model, given that the physics of the joint are not perfectly understood. From first principles, one might approach the problem from the same viewpoint as before; one could make reasonable/trusted assumptions about the real joint and the model joint, and try to determine how far they can diverge.

In a general theory, one would hope to prove theorems that were general, perhaps across particular classes of joint models; consider, for example, the reasonable conjecture:

> Suppose that given models $M_i^{C_i}$ $(i=1,\ 2)$ are ε_i-mirrors for structures S_i in contexts C^i, then $M^A = M_1^{C^1} \overset{\oplus}{M_J} M_2^{C^2}$ is an ε_A-mirror for the structure $S_1 \overset{\oplus}{S_J} S_2$ in the context $C_1 \oplus C_2$ with $\varepsilon_A \geq \max(\varepsilon_1, \varepsilon_2)$ (where $\delta = \varepsilon_A - \max(\varepsilon_1, \varepsilon_2) \geq 0$ is defined as the *joining deficit*).

Finally, it is important to mention another use of the idea of joining models. One might simply wish to represent a complex structure in terms of substructures, even if there is no physical joining process involved (a situation that arises in hybrid testing [25]). A simple example will suffice. Suppose one wished to model a fixed–fixed beam, and to validate the model. However, suppose that one had no validation data for the beam, but one did possess a validated model for a cantilever beam; in fact the cantilever model had been established as an ε-mirror. Clearly, one can regard the fixed–fixed beam as two cantilevers joined *perfectly* at their tips, as depicted in Fig. 3. One could now attempt to answer the question above, as to whether joining two copies of the cantilever beam is an ε_A-mirror for the fixed–fixed beam. In this case, one might assume that the joint model M_J is *perfect*; in practice a perfect joint when joining two FE models would be accomplished by seamlessly merging the meshes at the joint so that material continuity is as good at the joint as anywhere else. Perfect or idealized joints of this nature will be denoted by the symbol $\overset{\oplus}{P}$.

Even in the case of a perfect joint, one should be aware of a *caveat*, and this relates to *context*. Suppose that the cantilever model was *linear* and had been validated on test data showing small or moderate deflections of the cantilever tip. When the cantilevers are joined, and the cantilever tips become the midpoint of the beam, the response of the real beam will become nonlinear for much smaller values of midpoint displacement than the values measured at the cantilever tip.

It is possible that this problem is achievable via transfer learning, where the scenario would become multisource transfer learning [26]. For the encastré beam example the sources would become the two $M_i^{C_i}$ $(i=1,\ 2)$ which are known to be ε_i-mirrors and the perfect joint model. The challenge here is obtaining the information about the perfect joint model, and knowing that it is some form of ε-mirror. This in turn could be inferred from multiple perfect joint models that may have been validated for different geometry and boundary condition scenarios, the idea being that the mapping for a perfect joint can be learnt from this set. If this is the case, then the three models could be used as source data in order to obtain the target deflections for the encastré beam.

Many of the ideas discussed here are covered by the *multilevel* framework discussed in Ref. [6], and it may be that the ideas of *reliability* and *relevance* applied in that framework can be adopted in order to prove hypotheses like those pointed out in this paper.

3.3 An Example Concerning Structural Health Monitoring. One of the major problems with data-based Structural Health Monitoring (SHM) is that data from damaged structures is scarce. Although damage detection is possible even if one only has data from the structure of interest, using unsupervised learning [27], higher-level diagnostics like locating damage or assessing its type or severity can only be accomplished if one has data from all the damage states of interest. It is inconceivable that one might carry out a test program that systematically involved damaging numbers of high-value structures, so one has to turn toward modeling as a means of providing the necessary data, as defined in forward model-driven (SHM) [28,29]—where models (potentially inferred using inverse methods) are utilized to perform forward simulations under various damage scenarios.

The context responses in an SHM problem are usually going to be features for machine learning. Given the importance of the specific context, new notation will be introduced; the SHM context will be denoted F.

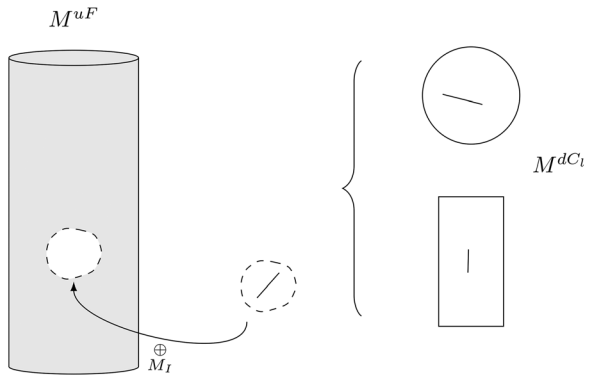

Fig. 4 Structural health monitoring illustration using pressure vessel and crack models. An example of insertion of a local damage model into an undamaged structure model.

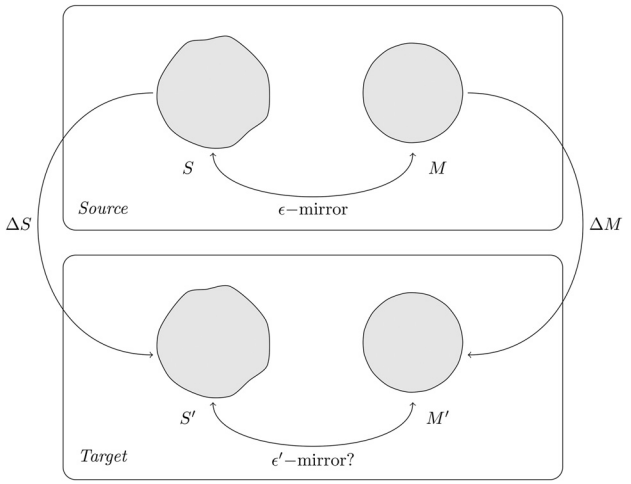

Fig. 5 A depiction of the validation problem in the context of design and the transfer learning problem setup

Assume two ingredients: the first is a validated model of the undamaged structure of interest S^u, denoted by M^{uF}. Further assume a set of data $\{D^u_{Tr}, D^u_T\}$ which has been used to validate the data. Further assume that M^{uF} is an ε^u-mirror, according to some appropriate metric.

The second ingredient is a *local* damage model M^d, which has been validated in a context C_l using data from coupon tests. The model may have been updated on the basis of test data and may well be a hybrid (gray-box) model. Assume that under the circumstances M^{dC_l} is an ε^d-mirror for the context C_l according to some appropriate metric. Finally, we assume that there are *no validation data* for the damaged structure S^d.

The problem is essentially a joining problem; however, it is of a specific type and merits a little more new notation. An *insertion* model M_I is defined as an algorithm or prescription for embedding the model M^{dC_L} in M^{uF}, in such a way that the result is a model for S^d. This differs from the previous joint definitions in that there is no new physics associated with the join. M_I could be a very simple process, i.e., if the two component models are FE models, insertion will only really mean harmonizing the two meshes along the boundary of the join, or using a super-element approach. One can think of the process as a type of *surgery*,[10] i.e., one cuts out a healthy region of M^{uF} and replaces in with M^{dC_l}, as in Fig. 4, and then harmonizes the meshes at the boundary.[11] Clearly this means that there will need to be compatibility conditions which guarantee some degree of smoothness/continuity across the boundary.[12]

There is another compatibility condition required here by the theory; the models M^u and M^d must exchange information in such a way that the dynamics evolves appropriately, i.e., the response context of C_l must overlap with the environmental context of F, i.e., $C_l \cap F \neq \phi$. In fact, in a general assembly model $M^{C_1} \oplus_{M_I} M^{C_2}$, it will usually be necessary that $C_1 \cap C_2 \neq \phi$ and $C_1 \cap C_2 \neq \phi$ (where ϕ represents the empty set here).

As a fairly simple example, consider the problem of modeling a crack in a pressure vessel (Fig. 4). The undamaged model M^{uF} represents the vessel; the damage model M^{dC_l}, represents a through crack in a section of plate. By joining the two models, one can embed a crack of arbitrary location, length or orientation in the vessel (the process might require some care near the boundaries). A subtlety here is that the crack model might have been validated for flat specimens, or for a range of different plate assumptions (as seen in Fig. 4), in which case a modification might be needed for compatibility with the curved surface of the vessel. A more important issue is the following. The behavior of the structure will usually be modeled using macroscopic physics, while the detailed crack model will require microscopic physics; this means that the features have to be chosen very carefully so that the behavior of the crack is communicated over the boundary effectively.

The mathematical question of interest is:

Given all of the above, is $M^{C_1} \oplus_{M_I} M^{C_2}$ an ε-mirror for S^d, and if so, what is the smallest value of ε for which this is true?

This will usually be a probabilistic problem where the metrics are quantities like probability of misclassification or probability of detection, in which case it will probably be more appropriate to frame the problem in terms of α-mirrors.

The insertion model M_I could also be seen as the output from transfer learning; there the aim would be to transfer knowledge from various validated coupon crack models that are α-mirrors (source) to learn the target predictive function which maps to some damage feature space in the pressure vessel model. This would be suited to a multisource transfer learning problem [26].

3.4 An Example Concerning Design. This is one of the potential applications of digital twin technology that would produce large cost savings for industry.

Suppose one has an existing structure S and a context C; further suppose that a virtualization $V^C = (M^{hC}_{\varepsilon_1}, M^{EC}_{\varepsilon_2})$ exists which has been validated and shown to be an ε-mirror for S^C.

Imagine now that one wished to design a new structure S', and thus wanted to know how it would behave either in the old context C, or in a new context C' given small changes ΔM and ΔS, as depicted in Fig. 5. In a situation where one wished to avoid building a prototype for S', there is no direct means of validating a new visualization $V\prime^C = (M'^{hC}, M'^{EC})$, even though this would be ideal for conducting "what-if" games for the new structure. The question of immediate interest is:

Given a virtualization $V^C = (M^{hC}_{\varepsilon_1}, M^{EC}_{\varepsilon_2})$, which is an ε-mirror for S^C; is $V'^C = (M'^{hC}_{\varepsilon'_1}, M'^{EC}_{\varepsilon'_2})$ a mirror for $S\prime^C$ for any values of ε'_1 and ε'_2, and if so, what are the smallest possible values for which this true?

As in the context change scenario, the transfer learning problems would use the information and mapping from the known ε-mirror as the source domain, as shown in Fig. 5. A mapping would then be inferred for the updated model design, again providing an estimated bound on ε'.

[10]Surgery is a mathematical technique for building complicated topological spaces from simpler ones [30]. It may be that the technique can be applied in the context of joining models.

[11]This is similar to the situation in real-time hybrid testing where coordinate sets are defined in each domain, which need to be synchronized in order to form the join. Errors in the synchronization process then give a measure of how imperfect the joint is.

[12]Note that this is rather perverse version of surgery, where undamaged tissue is replaced by damaged.

3.5 An Example Concerning Multifidelity Models: Refinement and Relaxation. This section considers the situation when one has multiple models of the same structure S, in a fixed context C. Suppose that a model M^C is an ε-mirror for S. A modified model $M'^C = \mathrm{Ref}[M^C]$ is a *refinement* of M^C, if it is an ε'-mirror with $\varepsilon' < \varepsilon$. Similarly, A modified model $M'^C = \mathrm{Rel}[M^C]$ is a *relaxation* of M^C, if it is an ε'-mirror with $\varepsilon' > \varepsilon$. For finite element models, these operations can be carried out by refining or coarsening the mesh. In this simplest of situations, one might estimate the values of ε' using analytical error estimates.

This idea is one that can be used in order to answer question (1) in the introduction. In principle one starts with a model M^C which is probably fit-for-purpose and then relaxes the model until one arrives at M'^C with $\varepsilon' = \varepsilon_T$.

Now, it is possible to consider what sort of propositions one might wish to prove in the theory, i.e., consider the hypothesis:

Assume a model $M^A = M_1^{C^1} \underset{M_J}{\oplus} M_2^{C^2}$ is an ε_A-twin for a joined structure $S^1 \underset{S_J}{\oplus} S^2$. Further suppose that $M_1^{C_1}$ is an ε_1-mirror. Now, if $M\prime^A = M\prime_1^{C^1} \underset{M_J}{\oplus} M_2^{C^2}$ is obtained by refining the first submodel, then $M\prime^A$ is an ε'_A-mirror, with $\varepsilon'_A < \varepsilon_A$.

Another strategy for answering question (1) would then be to relax submodels in an assembly until the result is marginally fit-for-purpose.

4 Discussion and Conclusions

This paper proposes some ingredients for a mathematical theory which would allow a general framework for measuring the fidelity of computational models and for understanding the consequences of combining validated models or using them outside their original context. Such a theory would be invaluable in the design and construction of digital twins, because one of the main uses of digital twins will be to make predictions in circumstances where their core models have not been explicitly validated, and it will be critical to obtain estimates of how much models can be trusted when they are used to extrapolate or generalize, i.e., when they are used to make inferences about different structures or in different contexts.

As discussed in the introduction, there are already attempts to define a unifying framework for model calibration and validation. In fact, these papers already go into greater detail on specific technical points than this paper, e.g., they go as far as to propose a Bayesian framework and define appropriate priors, likelihoods etc. [6,7]. The techniques proposed can very much form part of the armory of the more general methodology proposed here. This paper deliberately draws back from some details because the authors believe that important discussions are still to be had. For example, it is not agreed within the broader V&V and uncertainty quantification communities that probability theory is the correct way to approach model bias, or epistemic uncertainty in general. For this reason, some of the definitions given here are independent of whatever uncertainty theory ultimately dominates in a given context. As long as an uncertainty theory singles out some *most highly indicated* model from the population of possible choices, one can base the analysis on the ε-mirror for that single model. For example, in a Bayesian framework, one can apply the idea to the *Maximum a Posteriori* model. Of course, any theorems in the general theory will have to be proved independently for each uncertainty specification.

In some ways, the paper could still—despite the intention of the authors—be considered a *wish list*. In defense of this accusation, the arguments are presented in the real belief that the wishes could come true. The paper presents only the sketchiest arguments as to how the various "theorems" might be proved, or how the relevant estimates could be made; this is because the current authors do not have anything like the complete range of abilities/skills that will be needed in order to assemble the theory. In many ways the paper is intended as a rallying call to the academic community; the skills needed will come from a range of disciplines: pure and applied mathematics, physics, computer science (particularly machine learning), and engineering. The authors here believe that a framework can come together which is more than the sum of its parts and that can be of lasting value in the pursuit of effective computer models and particularly in the construction of digital twins.

Funding Data

- Engineering and Physical Sciences Research Council (EPSRC) (Grant No. EP/R006768/1; Funder ID: 10.13039/501100000266).

References

[1] Tuegel, E., Ingraffea, A., Eason, T., and Spottswood, S., 2011, "Reengineering Aircraft Structural Life Prediction Using a Digital Twin," Int. J. Aerosp. Eng., **2011**, pp. 1–14.

[2] Datta, S., 2017, "Emergence of Digital Twins—Is This the March of Reason?," J. Innovation Manage., **5**(3), pp. 14–33.

[3] Grieves, M., and Vickers, J., 2017, *Digital-Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*, Springer, New York, pp. 85–113.

[4] Wagg, D., Worden, K., Barthorpe, R., and Gardner, P., 2020, "Digital Twins: State-of-the-Art and Future Directions for Modelling and Simulation in Engineering Dynamics Applications," ASCE-ASME J. Risk Uncertainty Eng. Syst., Part B (accepted).

[5] Tao, F., Zhang, M., Liu, Y., and Nee, A., 2018, "Digital Twin Driven Prognostics and Health Management for Complex Equipment," CIRP Ann., **67**(1), pp. 169–172.

[6] Li, C., and Mahadevan, S., 2016, "Role of Calibration, Validation and Relevance in Multi-Level Uncertainty Integration," Reliab. Eng. Syst. Saf., **148**, pp. 32–43.

[7] Nagel, J., and Sudret, B., 2016, "A Unified Framework for Multilevel Uncertainty Quantification in Bayesian Inverse Problems," Probab. Eng. Mech., **43**, pp. 68–84.

[8] Kennedy, M. C., and O'Hagan, A., 2001, "Bayesian Calibration of Computer Models," J. R. Stat. Soc., **63**(3), pp. 425–464.

[9] Bishop, C., 2007, *Pattern Recognition and Machine Learning*, Springer, New York.

[10] Cherkassky, V., and Mulier, F., 1998, *Learning From Data: Concepts, Theory and Methods*, Wiley, Hoboken, NJ.

[11] Worden, K., Barthorpe, R., Cross, E., Dervilis, N., Holmes, G., Manson, G., and Rogers, T., 2018, "On Evolutionary System Identification With Applications to Nonlinear Benchmarks," Mech. Syst. Signal Process., **112**, pp. 194–232.

[12] Worden, K., and Hensman, J., 2012, "Parameter Estimation and Model Selection for a Class of Hysteretic Systems Using Bayesian Inference," Mech. Syst. Signal Process., **32**, pp. 153–169.

[13] Abdessalem, A., Dervilis, N., Wagg, D., and Worden, K., 2018, "Model Selection and Parameter Estimation in Structural Dynamics Using Approximate Bayesian Computation," Mech. Syst. Signal Process., **99**, pp. 306–325.

[14] Rasmussen, C., and Williams, C., 2006, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA.

[15] Bi, S., Prabhu, S., Cogan, S., and Atamturktur, S., 2017, "Uncertainty Quantification Metrics With Varying Statistical Information in Model Calibration and Validation," AIAA J., **55**(10), pp. 3570–3583.

[16] Worden, K., 2002, "Some Thoughts on Model Validation for Nonlinear Systems," Proceedings of Third International Conference on Identification in Engineering Systems, Swansea, UK, Apr. 15–16, pp. 142–154.

[17] Gardner, P., Lord, C., and Barthorpe, R. J., 2019, "A Unifying Framework for Probabilistic Validation Metrics," ASME J. Verif., Validation Uncertainty Quantif., **4**(3), p. 031005.

[18] Turing, A., 1950, "Computing Machinery and Intelligence," Mind, **LIX**(236), pp. 433–460.

[19] Grieves, M., 2005, *Product Lifecycle Management: Driving the Next Generation of Lean Thinking*, McGraw-Hill Professional, New York.

[20] Torrey, L., and Shavlik, J., 2009, "Transfer Learning," *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques*, E. Soria, J. D. Martín-Guerrero, M. Martinez, R. Magdalena, and A. J. Serrano, eds., IGI Global, Hershey, PA, pp. 242–264.

[21] Pan, S. J., and Yang, Q., 2010, "A Survey on Transfer Learning," IEEE Trans. Knowl. Data Eng., **22**(10), pp. 1345–1359.

[22] Weiss, K., Khoshgoftaar, T. M., and Wang, D., 2016, "A Survey of Transfer Learning," J. Big Data, **3**(1), pp. 1–40.

[23] Ladeveze, P., and Leguillon, D., 1983, "Error Estimate Procedure in the Finite Element Method and Applications," SIAM J. Numer. Anal., **20**(3), pp. 485–509.

[24] Ainsworth, M., and Tinsley, J., 1997, "A Posteriori Error Estimation in Finite Element Analysis," Comput. Methods Appl. Mech. Eng., **142**(1–2), pp. 1–88.
[25] Gawthrop, P. J., Wallace, M. I., Neild, S. A., and Wagg, D. J., 2007, "Robust Real-Time Substructuring Techniques for Lightly-Damped Systems," Struct. Control Health Monit., **14**(4), pp. 591–600.
[26] Sun, S., Shi, H., and Wu, Y., 2015, "A Survey of Multi-Source Domain Adaptation," Inf. Fusion, **24**, pp. 84–92.
[27] Farrar, C., and Worden, K., 2012, *Structural Health Monitoring: A Machine Learning Perspective*, Wiley, Chichester, UK.
[28] Barthorpe, R. J., 2010, "On Model- and Data-Based Approaches to Structural Health Monitoring," Ph.D. thesis, University of Sheffield, Sheffield, UK.
[29] Gardner, P., 2019, "On Novel Approaches to Model-Based Structural Health Monitoring," Ph.D. thesis, University of Sheffield, Sheffield, UK.
[30] Cappell, S., Ranicki, A., and Rosenberg, J., 2000, *Surveys on Surgery Theory*, Vol. 1, Princeton University Press, Princeton, NJ.