Data to decisions: Real-time structural assessment from sparse measurements affected by uncertainty

Citation: Mainini, Laura and Willcox, Laura. "Data to decisions: Real-time structural assessment fromsparse measurements affected by uncertainty." Computers & Structures 182 (April 2017): 296-312 © 2016 Elsevier Ltd.

As Published: http://dx.doi.org/10.1016/j.compstruc.2016.12.007

Publisher: Elsevier BV

Persistent URL: https://hdl.handle.net/1721.1/125256

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Data to decisions: Real-time structural assessment from sparse measurements affected by uncertainty

Laura Mainini and Karen Willcox

Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA

Abstract

This paper proposes a data-to-decisions framework—a methodology and a computational strategy—to assist real-time decisions associated with structural monitoring and informed by incomplete, noisy measurements. The data-to-decision structural assessment problem is described in terms of sensor data measurements (such as strain components) and system capabilities (such as failure indices). A MultiStep Reduced-Order Modeling (MultiStep-ROM) strategy tackles the time-critical problem of estimating capabilities from measured data. The methodology relies on an offline-online decomposition of tasks, and combines reduced-order modeling, surrogate modeling, and clustering techniques. The performance of the approach is studied for the case of uncertain measurements arising from spatially distributed sensors over a wing panel. Both sensor noise and sensor spatial sparsity affect the quality of the information available online. The discussion is supported by three investigations that explore the efficiency of the online procedure for multiple combinations of quantity and quality of sensed data. The method is demonstrated for an unmanned aerial vehicle composite wing panel undergoing local degradation of its structural properties.

Keywords: data-driven reduced-order modeling, data-driven structural assessment, data-to-decisions, sparse and uncertain measurements, real-time capability assessment, self-aware vehicle.

1. Introduction

Advances in sensing technologies offer new opportunities for on-board monitoring of structural health in support of real-time vehicle operational decision making. As noted in Ref. [1], a modern approach to structural health monitoring should combine sensors and automated reasoning techniques, which together can support the data-to-decision flow. One important challenge is the cost (and possibly weight) penalties of sensing technologies, which motivates the question of effective strategies to place and employ a limited number of sensors [2–6], as well as the interest for efficient protocols to handle data acquisition [7, 8]. A second challenge is management of uncertainties in the data-to-decision process, including uncertainty associated with sensor measurements as well as the

uncertainty introduced through inference from sensor data that may be sparse and spatially distributed. A third challenge is the limited computational time and resources associated with data processing and algorithmic execution in support of real-time decision making. These challenges motivate a data-to-decisions mapping that is rapid, adaptive, efficient, effective, and robust.

Existing approaches to structural health assessment include strategies that evaluate changes in structural properties through dynamic analysis and modal characterization [9–15], and strategies that estimate structural parameters via static responses in terms of displacements [16–20] or strain [21, 22]. Both dynamic and static assessment strategies aim to solve a parameter identification problem, with signature analysis, pattern recognition, system identification and model updating being the main approaches used [23–25]. Quantification of the uncertainties that affect the measurements combined with different levels of knowledge about the damage is an important effect that has been studied in several contexts [26–28]. To target rapid online structural assessment, neural networks have been used to cast the damage identification as a pattern recognition problem [12, 20, 29–33], while surrogate modeling has been used to create efficient representations of the behavior of damaged structures [34–40].

This paper addresses real-time structural health monitoring as a system assessment task in support of autonomous operational decisions. The decision-oriented perspective allows reframing the structural assessment task to emphasize the prediction of system capabilities rather than just the identification of system parameters. In particular, we formulate the problem in the form of a SENSE-INFER-PLAN-ACT information flow associated with two distinct sets of quantities of interest: measured quantities of interest and capability quantities of interests (Figure 1). Measurements are physical quantities that can be monitored with sensors (e.g., strain components); they represent a source of information about the state of a system (e.g., damage condition). Capabilities (e.g., failure indices) are quantities that dynamically evolve with the state of a system; they limit the space of possible actions (e.g., vehicle maneuvers). In this SENSE-INFER-PLAN-ACT setting, we then tackle directly the time-critical problem of estimating capabilities from measured data. In doing so, we avoid the costly inverse problems related to the inference step (from measurements to system parameters, Figure 1) and we reduce the computational burden associated with full-order prediction problems (from system parameters to capabilities, Figure 1). Instead, we exploit directly the low-dimensional nature of the measurement-to-capability mapping.

In previous work, we introduced a data-driven adaptive strategy that exploits reduced-order models, surrogate modeling and clustering techniques [41]. Our approach combines parametric proper orthogonal decomposition (POD) [42–45], self-organizing maps (SOMs) [46–49], and local response surfaces, into a MultiStep Reduced-Order Modeling (MultiStep-ROM) strategy. The methodology comprises an offline phase, in which we generate data from high-fidelity physics-based simulations (e.g., a finite element model) and construct a surrogate model, and an online phase, in which the surrogate model is exploited to realize a fast mapping from sensor measurements to vehicle maneuver ca-

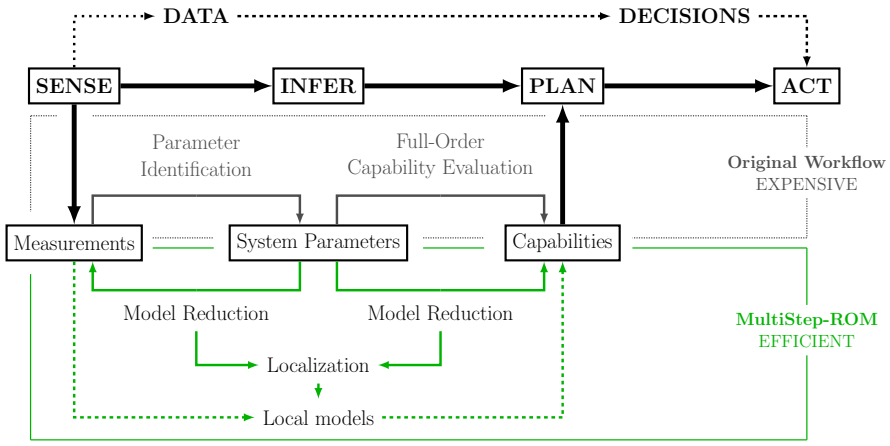

Figure 1: SENSE-INFER-PLAN-ACT framework associated to measurements and capabilities. The dotted box (black) frames the original workflow that includes parameter identification (from measurements to system parameters) and full-order prediction problems (from system parameters to capabilities). The solid line box (green) frames our MultiStep-ROM approach that exploits the low-dimensional nature of the measurement-to-capability mapping: solid arrows indicate the offline phase, dashed arrows denote the online phase.

pabilities. In [41], the mapping strategy was demonstrated in a deterministic structural assessment problem, i.e., assuming that measured values and sensor locations were exactly known.

In this paper, we extend this work to incorporate the effects of uncertainty. In particular, we develop the MultiStep-ROM approach to assist real-time decisions informed by uncertain sparse measurements. We consider uncertainty contributions due to (i) lack of knowledge about the actual location of the sensors and (ii) measurement noise associated with sensor accuracy. To study our computational methodology applied to this class of problems, we define performance metrics to assess accuracy and computational cost of the online procedure. The strategy is demonstrated for a specific data-to-decision problem of autonomous structural assessment onboard an unmanned aerial vehicle (UAV). In particular, we consider a UAV composite wing panel undergoing local degradation of the structural properties (damage). In the context of this representative test problem we investigate the limits and potential (number of sensors needed and robustness to uncertainty) of our approach.

Section 2 of this paper summarizes the MultiStep Reduced-Order Modeling method—originally introduced in Ref. [41]—and expands the method to include further algorithmic choices to improve online performance. Section 3 defines the structural assessment application addressed in this paper. In Section 4, we propose particular metrics to assess accuracy and runtime performance. Section 5 discusses our investigations and the relevant outcomes. Finally, Section 6 summarizes the key findings and concludes the paper.

2. MultiStep-ROM: Methodology and parameters

Our MultiStep Reduced-Order Modeling (MultiStep-ROM) strategy is based on an offline-online structure. During the offline phase we exploit the information provided by high-fidelity complete datasets to obtain an adaptive efficient model. This efficient model is then used during the online phase to rapidly process the information provided by sparse sensor measurements and estimate specific quantities of interest. In the following sections we briefly introduce these two phases together with the related algorithmic parameters. The details of the MultiStep-ROM methodology are described in Ref. [41]. In this paper we extend this technique to estimate system capabilities from sensor measurements that are sparse and affected by uncertainty.

2.1. Offline phase

The offline phase comprises four computational steps dedicated to the construction of an adaptive efficient model. The core idea is to leverage the physics-based knowledge, provided by high-fidelity data sets, by embodying it in a useful form through a combination of reduced-order modeling techniques and localization strategies.

2.1.1. Data collection

High-fidelity data represent our source of information about the physical quantities we aim to process online. These quantities of interest are all functions of the system parameters $\mathbf{x}$ and can be divided into two main groups reflecting the online distinction between quantities measured by the sensors and quantities characterizing system capability, Figure 1. Accordingly, we can distinguish M measured quantities of interest, $\mathbf{q}_m,\ m = 1, ..., M$, and C capability quantities of interest $\mathbf{s}_c,\ c = 1, ..., C$.

The first step collects sets of complete snapshots of measurements and capabilities. In many cases these data are generated by running numerical high-fidelity simulations for n_s different $\mathbf{x}$; however, the source of information can be a diverse set of models of the system, including historical and experimental data. In general, the number of snapshot elements, n_e, can differ for each quantity of interest, while all the quantities need to be evaluated for the same n_s system conditions.

2.1.2. Projection-based model order reduction

The second step exploits the collected data to realize reduced-order models for each measurement $\mathbf{q}_m(\mathbf{x})$ and capability $\mathbf{s}_c(\mathbf{x})$. In particular we employ parametric proper orthogonal decomposition (POD) [50] because it has been shown to be effective in compressing high-dimensional information arising from discretized fields [51–55], such as those in our problem of interest.

Considering the method of snapshots introduced by Sirovich [43] and the connection between POD and singular value decomposition [56, 57], for each quantity of interest we assemble the $n_\mathrm{e} \times n_\mathrm{s}$ matrix of complete snapshots and

compute the POD basis vectors. This orthonormal basis is optimal in the least-square sense [58, 59] and ordered according to the fraction of variance explained by each mode (reflected by the magnitude of the corresponding eigenvalue). The reduction is then achieved by retaining only a small number of the POD modes that capture most of the energy of the system. Thus, each quantity of interest can be approximated in a low-dimensional representation as a combination of its dominant POD modes:

$$\tilde{\mathbf{q}}_m(\mathbf{x}) = \bar{\mathbf{q}}_m + \sum_{j=1}^{n_m} \alpha_j^m(\mathbf{x})\phi_j^m \qquad m = 1, ..., M, \tag{1}$$

$$\tilde{\mathbf{s}}_c(\mathbf{x}) = \bar{\mathbf{s}}_c + \sum_{j=1}^{l_c} \beta_j^c(\mathbf{x})\psi_j^c \qquad c = 1, ..., C. \tag{2}$$

In Equation (1) and (2), $\bar{\mathbf{q}}_m$ and $\bar{\mathbf{s}}_c$ are the average value of $\mathbf{q}_m$, over the snapshots $\{\mathbf{q}_m^i\}_{i=1}^{n_\mathrm{s}}$, and the average value of $\mathbf{s}_c$, over the snapshots $\{\mathbf{s}_c^i\}_{i=1}^{n_\mathrm{s}}$, respectively. In the expansion terms, $\{\phi_j^m\}_{j=1}^{n_m}$ and $\{\psi_j^c\}_{j=1}^{l_c}$ denote the mth measurement POD modes and the cth capability POD modes, respectively; similarly, $\{\alpha_j^m\}_{j=1}^{n_m}$ indicates the mth measurement POD modal coefficient and $\{\beta_j^c\}_{j=1}^{l_c}$ represents the cth capability POD modal coefficient.

The model reduction is realized by truncating the POD expansion such that $n_m \ll n_\mathrm{e}$ and $l_c \ll n_\mathrm{e}$. The criteria to determine such truncation are based on the amount of information recovered with the retained terms. This is commonly assessed by evaluating the amount of cumulative energy $\mathcal{E}$ associated to the preserved modes:

$$\mathcal{E}(n_m) = \frac{\sum_{j=1}^{n_m} \lambda_j^m}{\sum_{j=1}^{n_\mathrm{s}} \lambda_j^m} \quad m = 1, ...M \qquad\qquad \mathcal{E}(l_c) = \frac{\sum_{j=1}^{l_c} \mu_j^c}{\sum_{j=1}^{n_\mathrm{s}} \mu_j^c} \quad c = 1, ..., C, \tag{3}$$

where λ_j^m and μ_j^c are the magnitudes of the POD eigenvalues associated to the jth POD mode of the mth measurement and cth capability, respectively.

This step aims to construct the model so as to avoid the computational cost associated with the direct evaluation of high-fidelity (full-order) models to compute $\mathbf{q}_m(\mathbf{x})$ and $\mathbf{s}_c(\mathbf{x})$. These high-fidelity models are typically too expensive and onerous to run in real-time. With the next two steps we present a method to map rapidly and effectively from measurement POD models to capability POD models. Specifically, we seek a mapping from the measurement modal coefficients $\{\alpha_j^m\}_{j=1}^{n_m},\ m = 1, ..., M$ to capability modal coefficients $\{\beta_j^c\}_{j=1}^{l_c},\ c = 1, ..., C$. A relationship that globally models this mapping with acceptable accuracy may still be complicated and too expensive to meet real-time resource constraints. Therefore, we propose to tackle this problem with a *divide and conquer* strategy comprising the identification of localized subdomains and their characterization with local, simple models.

2.1.3. Localization

The third step of our offline process is dedicated to the empirical identification of subdomains in the space defined by the POD coefficients retained for all measurements and capabilities. The main goal is to find domains in the coefficients space over which we characterize the relationship between measurements and capabilities with simple local models (e.g., low-order response surfaces). Several localization strategies are possible [60–64]; our particular implementation adopts self-organizing maps (SOM), also referred to as Kohonen's method or Kohonen's map [46–49]. SOM is a neural network that allows grouping of sample data according to their intrinsic similarities. It also projects the multidimensional input space onto a lower dimensional topological space where proximity of the clusters reflects their similarity in the higher dimensional input space.

Training data for the SOM is gathered by exploiting the POD models obtained in the previous step. For each quantity of interest, we evaluate the modal coefficients associated to each system state sampled in the first step. For system state $\mathbf{x}$, we collect these modal coefficients in the vector $\boldsymbol{\tau}(\mathbf{x})$:

$$\boldsymbol{\tau}(\mathbf{x}) = \left[\{\alpha_j^m(\mathbf{x})\}_{\substack{j=1,\dots,n_m \\ m=1,\dots,M}}, \{\beta_j^c(\mathbf{x})\}_{\substack{j=1,\dots,l_c \\ c=1,\dots,C}} \right] = \left[\boldsymbol{\alpha}(\mathbf{x}), \boldsymbol{\beta}(\mathbf{x}) \right]. \tag{4}$$

This results in a vector for each one of the n_s sampled states: $\{\boldsymbol{\tau}_i\}_{i=1}^{n_\mathrm{s}}$, where $\boldsymbol{\tau}_i \equiv \boldsymbol{\tau}(\mathbf{x}^i)$ for state $\mathbf{x}^i$. The collection of these vectors constitutes the SOM training dataset. Each vector has dimension $n_\mathrm{pod} = n_\alpha + n_\beta$, where $n_\alpha = \sum_{m=1}^{M} n_m$ and $n_\beta = \sum_{c=1}^{C} l_c$ denote the size of vectors $\boldsymbol{\alpha}(\mathbf{x})$ and $\boldsymbol{\beta}(\mathbf{x})$, respectively.

During the training phase, the SOM computes a set of n_w weight vectors, $\{\mathbf{w}_j\}_{j=1}^{n_\mathrm{w}}$, each of dimension n_pod. These weight vectors are the result of an iterative unsupervised learning [64–67] that groups similar training vectors into n_c clusters. For each cluster $k = 1, ..., n_c$, the training phase also determines the representative weight vector, $\mathbf{w}_k$, as the one for which the distance from the training vectors in k is minimized:

$$k = \underset{j \in 1,\dots,n_\mathrm{w}}{\arg\min} \{\|\boldsymbol{\tau}_i - \mathbf{w}_j\|\}. \tag{5}$$

Specifically, we adopt the distance metric $\|\cdot\|_{\boldsymbol{\Lambda}}$ that we introduced in Ref. [41]:

$$\|\boldsymbol{\tau}_i - \mathbf{w}_k\|_{\boldsymbol{\Lambda}} = \sqrt{(\boldsymbol{\tau}_i - \mathbf{w}_k)^\top \boldsymbol{\Lambda} (\boldsymbol{\tau}_i - \mathbf{w}_k)}. \tag{6}$$

Equation (6) indicates a scaled L_2-norm where the diagonal matrix $\boldsymbol{\Lambda}$ contains the normalized POD eigenvalues, ordered to correspond to the POD coefficients sequence in $\boldsymbol{\tau}$, as defined in (4). At the end of the training phase, the weight vectors represent the cluster prototypes, embodying the average properties of their training elements. By adopting metric (6) as a dissimilarity measure, we aim to achieve a better characterization of the most energetic and informative dimensions of our POD coefficients space. The results discussed in Section 5 of

this paper will reveal the advantages of this particular choice, as it is largely responsible for the efficient behavior of our strategy.

The number of clusters n_c is not imposed *a priori*, but determined by the SOM. However, this output of the localization step is limited by the number of weight vectors that characterize the SOM neural network. Thus, $n_c \leq n_\mathrm{w}$ because each cluster has a representative weight vector, but in principle not all the n_w vectors end up leading a cluster. Moreover, at this stage we must consider another important constraint: each of the n_c clusters must include a minimum number of training vectors, n_τ, to permit the construction of the local models. Thus, since the overall number of training vectors is n_s, the size n_w of the SOM network should be chosen such that $n_\mathrm{w} < n_\mathrm{s}/n_\tau$.

The final SOM clusters empirically represent subdomains in the space of the retained POD coefficients; in the next step, each subdomain is characterized with its own local model. Therefore, the algorithmic parameter n_c plays a key role and represents a trade-off between partitioning the POD coefficients space to have simple local models and avoiding data over-fit.

2.1.4. Local data-fit models

The fourth step finalizes our *divide and conquer* approach to obtain an effective mapping from measurement POD coefficients $\boldsymbol{\alpha}$ to capability POD coefficients $\boldsymbol{\beta}$. We wish to characterize each subdomain k with a dedicated set of simple data-fit models of the capability coefficients as a function of measurement coefficients. In particular, for every single capability coefficient β_j^c we seek a simple response surface (RS) model for $\beta_j^c(\boldsymbol{\alpha})$, valid exclusively for the specific kth subset:

$$\boldsymbol{\beta}^{(k)} \approx \tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha})^{(k)} = \{\tilde{\beta}_j^c(\boldsymbol{\alpha})\}^{(k)}_{\substack{j=1,\dots,l_c \\ c=1,\dots,C}}, \qquad k = 1, ..., n_c, \tag{7}$$

where $\tilde{\beta}_j^c(\boldsymbol{\alpha}) \approx \beta_j^c(\mathbf{x})$ and $\boldsymbol{\alpha} = \boldsymbol{\alpha}(\mathbf{x})$ as defined in (4). Equation (7) indicates the set of functions that constitute the actual online mapping and will eventually be suitable to run in real time.

In order to meet a problem's specific computational constraints and resource limitations, this last offline phase can be properly tuned by acting on a few parameters: (i) the actual number of measurement modal coefficients that constitute the domain of the local RS, n_α^rs, (ii) the degree of the local polynomial RS, p, and (iii) the actual number of retained capability modes, n_β^k. We discuss here some key considerations regarding choices of these parameters.

Response surface approximations tend to be more effective and reliable for smooth functions defined on low-dimensional domains. The final objective here is to obtain reliable approximations while keeping the local models simple. Therefore, we adopt RSs defined over a n_α^rs-dimensional domain ($\boldsymbol{\alpha}^\mathrm{rs} \in \mathbb{R}^{n_\alpha^\mathrm{rs}}$) rather than over the entire n_α-dimensional space of retained measurement POD coefficients ($\boldsymbol{\alpha} \in \mathbb{R}^{n_\alpha}$): if $n_m^\mathrm{rs} \leq n_m$ for each m measured quantity, then $n_\alpha^\mathrm{rs} = \sum_{m=1}^M n_m^\mathrm{rs} \leq \sum_{m=1}^M n_m = n_\alpha$. Our particular strategy takes advantage of the fact that for each measurement m the n_m^rs modal components are

dominant and ordered. In this way, the resulting RS domain still represents a large information content to obtain useful local approximations $\tilde{\beta}_j^c(\boldsymbol{\alpha}^{\mathrm{rs}})$. Then, we act on p and choose simple low-degree polynomials as local RS.

These choices lead to approximations $\tilde{\beta}_j^c$ that are more effective for the first POD coefficients because the first modal terms are smoother than the higher order components. We take into account this fact to realize a further selective reduction of the capabilities terms. We introduce a criterion to cut out all the capability coefficients that are poorly approximated in offline tests. By fixing a maximum allowable offline error for $\tilde{\beta}_j^c$, we aim to limit the large corruption introduced online with badly approximated components (typically the higher order ones). As a consequence $l_c^k \leq l_c$ and $n_\beta^k = \sum_{c=1}^{C} l_c^k \leq \sum_{c=1}^{C} l_c = n_\beta$, that is, the number of capability coefficients to approximate can be reduced and generally differ from cluster to cluster, as the set of local RSs does.

2.2. Online phase

The online phase constitutes the computational procedure to run on-board to execute decisions in real-time. This procedure consists of four steps leading to a final estimate of system capabilities, given sparse sensor measurements. The online process addresses the SENSE-INFER-PLAN-ACT information flow with all the related challenges, including the incomplete measurements and the parameter identification problem (see Figure 1). Our strategy aims to exploit the multi-step adaptive model reduction computed offline to map directly from measurements to capabilities without resolving the expensive inverse problem placed in between.

2.2.1. Measurement coefficients reconstruction: from $\hat{\mathbf{q}}_m$ to $\boldsymbol{\alpha}_{\mathrm{g}}$

The first online step is the most delicate one because it deals with sensor measurements, that is, the on-board source of information about the condition of the system. Since those data feed the entire online process, measurements should be carefully manipulated. A further significant challenge is the fact that online we can usually measure our quantities of interest only in a fraction (f_{p}) of the complete points we have offline. This is due to both the large cost of sensing systems and the limited resources available on-board [68, 69].

We need to extrapolate the most informative content from sparse data with an efficient computational process. In particular, we here wish to translate the sensor measurements into input data suitable for a mapping in the form $\boldsymbol{\beta}(\boldsymbol{\alpha})$, as defined offline with Equation (7). Hence, we have to reconstruct the POD coefficients of the measured quantities $\boldsymbol{\alpha}$ using the information provided by incomplete (sparse) snapshots $\hat{\mathbf{q}}_m$, $m = 1, ..., M$. To do this, for each measurement m we exploit the related POD basis $\{\phi_j^m\}_{j=1}^{n_m}$, computed offline (1), to recover the coefficients using gappy POD (GPOD)[70–72]. This involves solving M linear systems in the form:

$$\mathbf{G}^m \boldsymbol{\alpha}_{\mathrm{g}}^m = \mathbf{f}^m \tag{8}$$

where $\boldsymbol{\alpha}_{\mathrm{g}}^m = \left[\alpha_{\mathrm{g}1}^m, ..., \alpha_{\mathrm{g}n_m}^m\right]^\top$ are the unknown mth POD coefficients to compute, the ijth entry of $\mathbf{G}^m$ is $G_{ij}^m = (\phi_i^m, \phi_j^m)_{\mathrm{g}}$, the ith entry of vector $\mathbf{f}^m$ is $f_i^m = (\bar{\mathbf{q}}_m, \phi_i^m)_{\mathrm{g}}$, and $(\cdot,\cdot)_{\mathrm{g}}$ denotes the gappy inner product which considers only those elements in the vectors that correspond to the available sensed data. Finally, the coefficients reconstructed for the M measurements are all gathered in $\boldsymbol{\alpha}_{\mathrm{g}} = \left[(\boldsymbol{\alpha}_{\mathrm{g}}^1)^\top, ..., (\boldsymbol{\alpha}_{\mathrm{g}}^M)^\top\right]$.

It can be seen that $\boldsymbol{\alpha}_{\mathrm{g}}$ is a n_α-dimensional vector; nevertheless, there may be an advantage in reducing the number of modal components to be reconstructed online via GPOD, such that $n_\alpha^{\mathrm{g}} \leq n_\alpha$ (and $\boldsymbol{\alpha}_{\mathrm{g}}^m = [\alpha_{\mathrm{g}1}^m, ..., \alpha_{\mathrm{g}n_m^{\mathrm{g}}}^m]^\top$ with $n_m^{\mathrm{g}} \leq n_m$). We also require that $n_\alpha^{\mathrm{rs}} \leq n_\alpha^{\mathrm{g}}$, since we wish to recover at least the modal coefficients over which we defined our local RSs; otherwise, we will not have sufficient information to estimate $\boldsymbol{\beta}$ later on (Section 2.2.3). These considerations limit the range of choices for the parameter n_α^{g} and constrain the selection of the modal term $\boldsymbol{\alpha}_{\mathrm{g}}$ to recover .

2.2.2. Classification: from $\boldsymbol{\alpha}_{\mathrm{g}}$ to $\mathbf{w}_k$

The second online step determines the set of models (those corresponding to the k^*th cluster) best suited to represent the system condition revealed by sensor measurements. To do this, we need to classify the sensed information, now condensed and represented in terms of coefficients $\boldsymbol{\alpha}_{\mathrm{g}}$, into one of the n_c clusters determined offline with SOM. The SOM clusters are represented by their leading weight vectors $\{\mathbf{w}_j\}_{j=1}^{n_c}$ (prototypes). Thus, we simply need to find the closest cluster whose weight vector $\mathbf{w}_{k^*}$ minimizes the distance metric (6) defined offline:

$$k^* = \underset{j\in 1,\dots,n_c}{\arg\min}\{\|\boldsymbol{\tau}^* - \mathbf{w}_j\|_{\boldsymbol{\Lambda}_\alpha}\}. \tag{9}$$

In Equation (9), vector $\boldsymbol{\tau}^* = [\boldsymbol{\alpha}_{\mathrm{g}}, \boldsymbol{\beta}^*]$ collects the n_α^{g} coefficients $\boldsymbol{\alpha}_{\mathrm{g}}$, reconstructed via GPOD, and the n_β-dimensional vector of unknowns $\boldsymbol{\beta}^*$, while $\|\cdot\|_{\boldsymbol{\Lambda}_\alpha}$ indicates that the weighted norm $\|\cdot\|_{\boldsymbol{\Lambda}}$ is now computed only over the coefficients $\boldsymbol{\alpha}_{\mathrm{g}}$, that is, over n_α^{g} elements of the vectors. If $n_\alpha^{\mathrm{g}} < n_\alpha$, the tuple $\boldsymbol{\alpha}_{\mathrm{g}}$ includes zero elements corresponding to the non reconstructed modal terms so that $\boldsymbol{\alpha}_{\mathrm{g}} = [(\boldsymbol{\alpha}_{\mathrm{g}}^m)^\top, [\mathbf{0}]^{n_m - n_m^{\mathrm{g}}}]_{m=1}^M$ where $[\mathbf{0}]^{n_m - n_m^{\mathrm{g}}}$ denotes a $(n_m – n_m^{\mathrm{g}})$-dimensional row vector of zeros.

2.2.3. Local approximation: from $\mathbf{w}_k$ to $\boldsymbol{\beta}^{(k)}(\boldsymbol{\alpha}_{\mathrm{g}})$

Once the most representative cluster (the k^*th) is identified, the set of related models $\tilde{\boldsymbol{\beta}}^{(k^*)}$ computed offline is used to approximate each of the $n_\beta^{k^*}$ modal coefficients of the capabilities as a function of the reconstructed $\boldsymbol{\alpha}_{\mathrm{g}}$:

$$\beta_j^c \approx \tilde{\beta}_j^c(\boldsymbol{\alpha}_{\mathrm{g}}^{\mathrm{rs}})^{(k^*)}, \tag{10}$$

where $\boldsymbol{\alpha}_{\mathrm{g}}^{\mathrm{rs}}$ denotes $\boldsymbol{\alpha}_{\mathrm{g}}$ in the n_α^{rs}-dimensional domain of the local RSs. This third online step explains the constraint introduced in Section 2.2.1 on the minimum number of measurement coefficients to reconstruct online via GPOD. According to that, we wish to recover at least the modal components that constitute the

domain of the local RSs ($n_\alpha^{\mathrm{rs}} \leq n_\alpha^{\mathrm{g}}$). Otherwise, in this third stage we would not have enough information to estimate the unknown coefficients β_j^c with our local models. In addition, note that $n_\beta^{k^*} \leq n_\beta$; in practice, we do not recover all the n_β elements of the unknown $\boldsymbol{\beta}^*$, but only the $n_\beta^{k^*}$ components downselected offline (see Section 2.1.4).

2.2.4. Capability estimate: from $\boldsymbol{\beta}^{(k)}(\boldsymbol{\alpha}_{\mathrm{g}})$ to $\mathbf{s}_c(\boldsymbol{\alpha}_{\mathrm{g}})$

The last online step provides the final estimate of the capability quantities of interest, that is the actual information we wish to derive from the quantities measured by the sensors. The expansions (2), computed offline, are here exploited to estimate each capability $\tilde{\mathbf{s}}_c$ as a combination of its POD modes ψ_j^c:

$$\tilde{\mathbf{s}}_c(\boldsymbol{\alpha}_{\mathrm{g}}) = \bar{\mathbf{s}}_c + \sum_{j=1}^{l_c^{k^*}} \tilde{\beta}_j^c(\boldsymbol{\alpha}_{\mathrm{g}})^{(k^*)} \psi_j^c \qquad c = 1, ..., C. \tag{11}$$

Equations (11) differ from the original expansions (2) for the coefficients that are now approximated with the k^*th set of local models (10). Moreover, each capability expansion might not include all the original l_c modal terms retained in (2), but only the subset of $l_c^{k^*}$ components screened with the downselection introduced in Section 2.1.4.

In this paper we employ the criteria introduced in Section 2.1.4 and we set the algorithmic parameters to obtain an efficient implementation of our methodology. We adopt simple local RS models defined on a low-dimensional domain ($n_\alpha^{\mathrm{rs}} \ll n_\alpha$) and retain only the few capability modal terms that satisfy offline constraints ($n_\beta^k \ll n_\beta,\ \forall k$). Such further model reductions permit very efficient implementations and help to speed up the entire online phase.

3. Structural assessment problem

In this paper we consider the application of the strategy to the structural assessment of an unmanned aerial vehicle (UAV). Specifically, we address the test problem of a composite wing panel subjected to static, uniaxal, uniform compression loading. Figure 2 illustrates the layout of the panel. It consists of a 18×18 square-inch component made up of four carbon-fiber layers with symmetric staking sequence $[45°, 0°]_{\mathrm{s}}$. The panel has fastened borders reinforced with two additional plies of orientation $0°$ and $45°$.

The collection of the snapshot datasets is achieved through high-fidelity analyses of panel behavior for different damage conditions. For these purposes, we employ a finite element model: the panel is discretized into $n_{\mathrm{e}} = 3921$ two-dimensional laminate plate elements. The four edges are clamped in fasteners locations, emulating the presence of the bolts along the perimeter. The presence of the damage is simulated by weakening the local stiffness of the elements that belong to the prescribed damage area. This specific modeling choice is one of the simplest options among possible models for structural damage and presents some fidelity limitations. Nevertheless, it still provides a useful basis on which

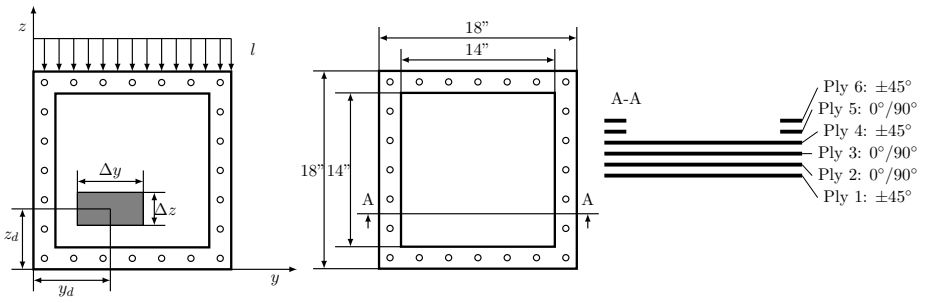

Figure 2: Panel parameters (damage location and damage size) and loading definition. Panel layout and layers sequence.

to investigate the effects of limited and noisy data on our ability to conduct real-time structural assessments.

3.1. Problem setup

Referring to the general structure of the information flow represented in Figure 1, it is possible to describe our structural assessment problem in terms of system parameters $\mathbf{x}$, measured quantities of interest $\mathbf{q}_m$, and capability quantities of interest $\mathbf{s}_c$. The system parameters define a specific damage condition in terms of damage location and extent. Our parameter vector $\mathbf{x}$ includes five components prescribing damage size along y and z directions (Δy and Δz, respectively), damage centroid location on y and z coordinates (y_d and z_d), and damage depth as the number of undamaged plies (d_d):

$$\mathbf{x} = [\Delta y, \Delta z, y_d, z_d, d_d]^\top. \tag{12}$$

The measured quantities of interest are three components of strain, for which values of the specific locations are provided by sensors placed on the external face of ply 4. We consider the normal components along the main orthotropic axes of the ply ($\boldsymbol{\varepsilon}_{\mathrm{n}_1}$ and $\boldsymbol{\varepsilon}_{\mathrm{n}_2}$) and the shear component on the ply plane ($\boldsymbol{\varepsilon}_{\mathrm{s}_{12}}$). This leads to three distinct measurement quantities of interest ($M = 3$) for a given state $\mathbf{x}$:

$$\mathbf{q}_1(\mathbf{x}) = \boldsymbol{\varepsilon}_{\mathrm{n}_1}(\mathbf{x}) \qquad \mathbf{q}_2(\mathbf{x}) = \boldsymbol{\varepsilon}_{\mathrm{n}_2}(\mathbf{x}) \qquad \mathbf{q}_3(\mathbf{x}) = \boldsymbol{\varepsilon}_{\mathrm{s}_{12}}(\mathbf{x}), \tag{13}$$

where each $\mathbf{q}_m(\mathbf{x})$ is an n_{e}-dimensional vector whose components are the element-wise values of the mth strain quantity over the panel. Given fixed loading and boundary conditions, the deformation field is determined by the specific damage affecting the structural integrity of the panel. Therefore, the measured quantities defined in Equation (13) are all functions of the structural parameter $\mathbf{x}$.

We consider a single capability quantity of interest: the failure index, *FI*, defined as the ratio between the experienced stress and the maximum allowable stress (a set of specific properties of the material that characterize its behavior under tension, compression and shear loading). Due to the orthotropic nature of our material, five failure modes characterize a single ply, yielding five different values of *FI* for each ply, in every element of the panel. We condense this data into a single capability vector ($C=1$) of dimension n_e:

$$\mathbf{s}_1(\mathbf{x}) = \mathbf{FI}(\mathbf{x}) \tag{14}$$

whose components are the maximum values of *FI* over all the failure modes and plies, for each element. As for the strain field, given fixed loading and boundary conditions, the stress field is determined by the characteristics of the damage affecting the panel. Hence, the capability quantity defined in Equation (14) is a function of the structural parameter $\mathbf{x}$.

3.2. Algorithmic setup

The study presented in this paper employs two sets of reference data: an *evaluation set*, to compute the surrogate models offline, and a *validation set*, to test the models and simulate the online phase. The evaluation set consists of measurements and capability snapshots computed for $n_\mathrm{s}=3000$ different damage conditions $\mathbf{x}$. The different cases are selected with a Latin hypercube exploration of the parameter space described in Table 1. Similarly, a second different Latin hypercube exploration determines the sample of $n_\mathrm{v} = 500$ damage conditions that constitute the validation set.

Parameters	Min Value	Max Value	Units
Δy damage size	2	8	[in.]
Δz damage size	2	8	[in.]
y_d damage location	4	14	[in.]
z_d damage location	4	14	[in.]
d_d damage depth	1	3	[plies]

Table 1: Damage parameter space: components and related bounds of variation.

We extensively investigated the sensitivity of our methodology with respect to a variety of algorithmic settings. In particular we studied how the tuning parameters affect the accuracy and computational time that characterize the online evaluations. We take into account this experience to choose the algorithmic setup of our strategy. The tuning parameters are: (i) the number of SOM weight vectors, n_w, that is the maximum number of possible SOM clusters partitioning the space of modal coefficients (as $n_c \leq n_\mathrm{w}$), (ii) the number of measurement modal coefficients that constitute the domain of local RS, n_α^rs, (iii) the degree of the local polynomial RS, p, and (iv) the maximum allowable offline error for capability coefficients, m_β, that determines the actual number

of retained capability modes n_β^k. The role of these parameters was already discussed in Sections 2.1.3 and 2.1.4. For all the simulations presented hereafter in this paper we adopt a 5×5 SOM network and linear local RSs $\tilde{\beta}_j^c(\boldsymbol{\alpha}^{\text{rs}})$ whose domain is defined by the first two modal coefficients of each measured quantity ($\boldsymbol{\alpha}^{\text{rs}} = [\alpha_1^1, \alpha_2^1, \alpha_1^2, \alpha_2^2, \alpha_1^3, \alpha_2^3]$). We also allow a 30% relative error in offline reconstruction tests for β_j^c. These choices lead to a specific setup with $n_{\text{w}} = 25$, $n_\alpha^{\text{rs}} = 6$, $p = 1$, and $m_\beta = 30\%$.

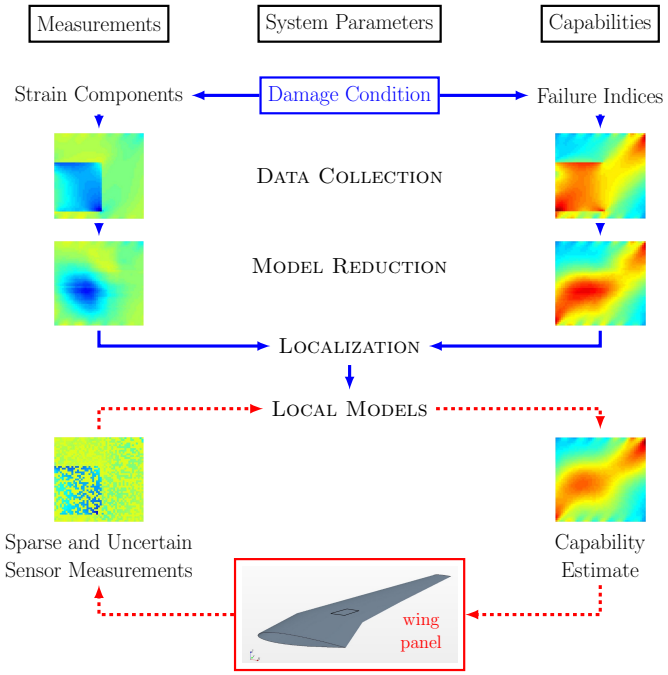

Figure 3: The MultiStep-ROM approach applied to the structural assessment of a wing panel. Solid line arrows (blue flow) illustrate the offline process (data collection, model reduction, localization, local models). Dashed line arrows (red flow) depict the online procedure to estimate capabilities from sparse and uncertain measurements.

Figure 3 illustrates the conceptual stages of our offline-online computational strategy applied to the real-time structural assessment of a wing panel. The solid-line arrows of the diagram depict the multi-step offline procedure (blue flow); the information contained in the evaluation set (3000 damage cases) is progressively condensed (via POD, SOM, and RS) into simple models that embody the knowledge to be used online. The dashed-line arrows visualize the on-board process (red flow) that exploits such knowledge to estimate the Failure Index (capability quantity) from sparse gages of strain components (incomplete measured quantities). The simulation of the real-time process is replicated for the entire validation set, that is, for 500 different damage cases.

4. Assessment metrics

In this paper we assess and discuss the behavior of our computational strategy in the case of sparse measurements affected by uncertainty. In particular we wish to monitor the procedure step by step, throughout the entire online phase. To do that, we propose specific accuracy and runtime metrics for each online stage.

Measurement coefficients reconstruction.. The accuracy of the first step is assessed using the reconstruction error arising from the gappy approximation of the POD coefficients, $\boldsymbol{\alpha}_{\mathrm{g}}$. For each modal term we compute the normalized root mean square error ($\epsilon_{\alpha_j^m}$) over the entire validation set and scale it with its relative energy contribution $\mathcal{E}_j^m = \lambda_j^m / \sum_{i=1}^{n_{\mathrm{s}}} \lambda_i^m$. Then, we average over the reconstructed terms to obtain the accuracy measure E_α^m for each quantity $m = 1, ..., M$:

$$E_\alpha^m = \frac{1}{n_m^{\mathrm{g}}} \sum_{j=1}^{n_m^{\mathrm{g}}} \epsilon_{\alpha_j^m} \mathcal{E}_j^m \quad \text{with} \quad \epsilon_{\alpha_j^m} = \frac{\|\boldsymbol{\alpha}_j^m - \boldsymbol{\alpha}_{\mathrm{g}j}^m\|_2}{\left((\alpha_j^m)_{\max} - (\alpha_j^m)_{\min}\right)\sqrt{n_{\mathrm{v}}}} \times 100\%. \quad (15)$$

In Equation (15), $\boldsymbol{\alpha}_j^m = \{\alpha_j^m(\mathbf{x}^i)\}_{i=1}^{n_{\mathrm{v}}}$ are the jth POD coefficients of the mth quantity for all n_{v} validation cases. Similarly, $\boldsymbol{\alpha}_{\mathrm{g}j}^m = \{(\alpha_{\mathrm{g}j}^m)^i\}_{i=1}^{n_{\mathrm{v}}}$ denotes the corresponding reconstructions obtained from sensor measurements using gappy POD. $(\alpha_j^m)_{\max}$ and $(\alpha_j^m)_{\min}$ are the maximum and minimum values in $\boldsymbol{\alpha}_j^m$, respectively. n_m^{g} indicates the number of reconstructed modes for the mth quantity such that $\sum_{m=1}^{M} n_m^{\mathrm{g}} = n_\alpha^{\mathrm{g}}$. The scaling term $\mathcal{E}_j^m$ mitigates the additive effects that would otherwise lead to larger errors by simply increasing the number of reconstructed modal coefficient; this permits a fairer and more meaningful comparison of the different cases investigated in this paper. Our structural assessment problem has $M = 3$ measured quantities of interest; thus we monitor E_α^1 and E_α^2 for the normal components of strain, and E_α^3 for the shear one.

Classification.. The second online step assigns the sensed condition to the most representative cluster. This classification is done according to the dissimilarity measure we defined in Equation 6 to determine the closest cluster. To assess our online classification step we exploit an internal measure commonly used in cluster analysis [67, 73]: the sum of squared errors, SSE. Specifically, we aim to compare the offline clustering (of $n_{\mathrm{s}} = 3000$ evaluation cases) with the online classification (of $n_{\mathrm{v}} = 500$ validation cases). The SSE is an additive measure that grows with the number of clustered data; as we are comparing two sets with different cardinality, we need to mitigate this effect. Therefore, we rather use the coefficient of determination $R^2 = 1 - \mathrm{SSE}/\mathrm{SST}$, where SST denotes the total sum of squares. Considering the particular dissimilarity measure adopted as clustering and classification criterion, we define the offline and online coefficients of determination accordingly.

Offline, the sum of squared errors and the total sum of squares are computed for the SOM clustering of the training vectors $\{\boldsymbol{\tau}_i\}_{i=1}^{n_s}$ (defined in Section 2.1.3):

$$\mathrm{SSE}_{\mathrm{off}} = \sum_{k=1}^{n_c}\sum_{i=1}^{n_k} \|\boldsymbol{\tau}_i^k - \mathbf{w}_k\|_{\boldsymbol{\Lambda}}^2 \qquad \mathrm{SST}_{\mathrm{off}} = \sum_{k=1}^{n_c}\sum_{i=1}^{n_k} \|\boldsymbol{\tau}_i^k - \mathbf{m}\|_{\boldsymbol{\Lambda}}^2. \tag{16}$$

In Equation (16), $\mathbf{w}_k$ denotes the kth cluster's prototype, that is the SOM weight vector that embodies the average properties of the n_k elements in cluster k, $\mathbf{w}_k \approx \mathbf{m}_k = \sum_{i=1}^{n_k} \boldsymbol{\tau}_i^k / n_k$. The vector $\mathbf{m}$ is the mean vector evaluated over all n_s elements of the evaluation set. Therefore, R^2_{off} indicates how good the SOM clustering is in representing the training data in the evaluation set. Our scaled distance metric prioritizes the characterization of the most energetic components, that is, the first modal terms of each quantity of interest. We can observe this by comparing each cluster's prototype $\mathbf{w}_k$ with the corresponding mean vector $\mathbf{m}_k$ component-wise: they match accurately for the first modes of each quantity, while the approximations progressively worsen for less informative terms (Figure 4).

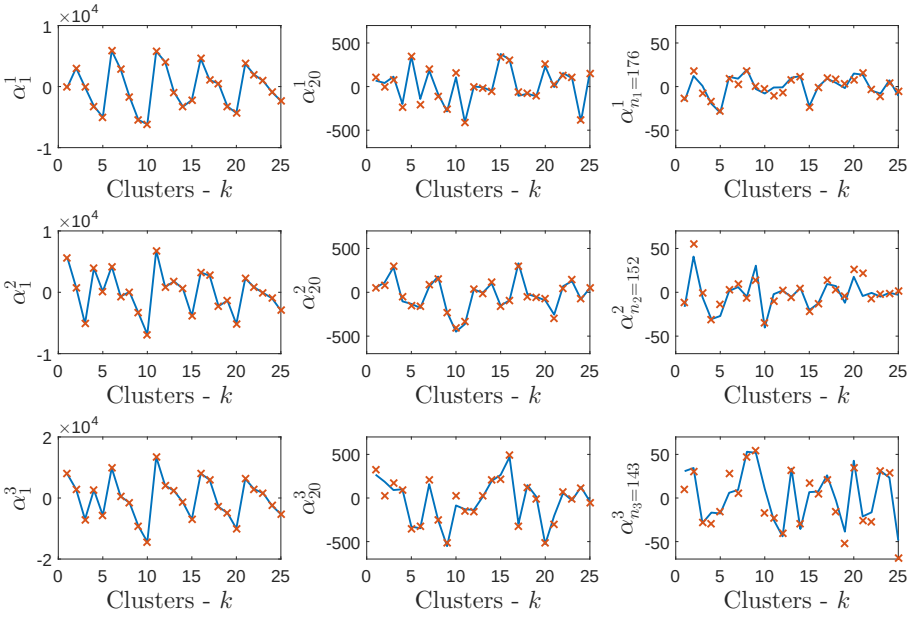

Figure 4: Clusters Prototype $\mathbf{w}_k$ — and Clusters Mean $\mathbf{m}_k$ × compared for different α_j^m. Measure (6) privileges the characterization of the most energetic components, that is, the first modal terms of each quantity of interest. Hence, prototypes and means match very accurately for the first modes (first column), while the approximations progressively (left to right) worsen for less informative terms.

For the online classification task, the sum of squared errors and total sum of squares are evaluated over the validation cases $\{\boldsymbol{\tau}_i\}_{i=1}^{n_v}$ (see $\boldsymbol{\tau}^*$ in Section 2.2.2):

$$\mathrm{SSE}_{\mathrm{on}} = \sum_{k=1}^{n_c}\sum_{i=1}^{n_k} \|\boldsymbol{\tau}_i^k - \mathbf{w}_k\|^2_{\boldsymbol{\Lambda}_\alpha} \qquad \mathrm{SST}_{\mathrm{on}} = \sum_{k=1}^{n_c}\sum_{i=1}^{n_k} \|\boldsymbol{\tau}_i^k - \mathbf{m}\|^2_{\boldsymbol{\Lambda}_\alpha}. \tag{17}$$

In Equation (17), $\mathbf{w}_k$ and $\mathbf{m}$ are again the kth prototype and the mean vector computed offline—as, by construction, both the evaluation set and the validation set are representative samples of the same space and have the same statistical moments. Thus, R^2_{on} indicates how good is the SOM clustering (determined offline) in representing the validation set. We aim to compare different online setups by monitoring their coefficients of determination: the more R^2_{on} approaches the offline reference value, the better the accuracy of the online classification.

Local approximation.. The accuracy metric for the third online step relies on the approximation error introduced by modeling the capability POD coefficients $\boldsymbol{\beta}_j^c$ as function of the reconstructed measurement POD coefficients $\boldsymbol{\alpha}_{\mathrm{g}}$. For each capability quantity of interest we compute the normalized root mean square error $\epsilon_{\beta_j^c}$ of the local approximation $\boldsymbol{\beta}_j^c(\boldsymbol{\alpha}_{\mathrm{g}})$ with respect to the original capability POD coefficients $\boldsymbol{\beta}_j^c$; then we scale it with the related fraction of energy $\mathcal{E}_j^c = \mu_j^c / \sum_{i=1}^{n_{\mathrm{s}}} \mu_i^c$ to mitigate the accumulation effect. For each quantity $c = 1, ..., C$, the accuracy measure E_β^c is computed by averaging $\epsilon_{\beta_j^c}$ over the n_β^k retained terms:

$$E_\beta^c = \frac{1}{n_\beta^k}\sum_{j=1}^{n_\beta^k} \epsilon_{\beta_j^c}\mathcal{E}_j^c \quad \text{with} \quad \epsilon_{\beta_j^c} = \frac{\|\boldsymbol{\beta}_j^c - \boldsymbol{\beta}_j^c(\boldsymbol{\alpha}_{\mathrm{g}})\|_2}{\left((\beta_j^c)_{\max} - (\beta_j^c)_{\min}\right)\sqrt{n_{\mathrm{v}}}} \times 100\%. \tag{18}$$

In Equation (18), $\boldsymbol{\beta}_j^c = \{\beta_j^c(\mathbf{x}^i)\}_{i=1}^{n_{\mathrm{v}}}$ are the jth POD coefficients of the cth capability quantity for all n_{v} validation cases. Similarly, $\boldsymbol{\beta}_j^c = \{\beta_j^c(\boldsymbol{\alpha}_{\mathrm{g}}^i)\}_{i=1}^{n_{\mathrm{v}}}$ denotes the corresponding approximations obtained with the local linear response surfaces. $(\beta_j^c)_{\max}$ and $(\beta_j^c)_{\min}$ are the maximum and minimum values in $\boldsymbol{\beta}_j^c$, respectively. Since for our application $C = 1$, we consider the single metric E_β and drop the superscript.

Capability estimate.. The overall accuracy of our strategy is assessed at the fourth step, the final estimate of capability quantities of interest. For each validation snapshot i, we compute the normalized root mean square error $\epsilon_{\mathbf{s}_c}^i$ in predicting the cth capability snapshots $\tilde{\mathbf{s}}_c(\boldsymbol{\alpha}_{\mathrm{g}}^i)$ with respect to the original FEM values $\mathbf{s}_c(\mathbf{x}^i)$. Then we define our accuracy measure $E_{\mathbf{s}}^c$ averaging over the entire validation set of n_{v} damage cases:

$$E_{\mathbf{s}}^c = \frac{1}{n_{\mathrm{v}}}\sum_{i=1}^{n_{\mathrm{v}}} \epsilon_{\mathbf{s}_c}^i \quad \text{with} \quad \epsilon_{\mathbf{s}_c}^i = \frac{\|\mathbf{s}_c(\mathbf{x}^i) - \tilde{\mathbf{s}}_c(\boldsymbol{\alpha}_{\mathrm{g}}^i)\|_2}{((s_c)_{\max} - (s_c)_{\min})\sqrt{n_{\mathrm{e}}}} \times 100\%. \tag{19}$$

In Eq (19), $\mathbf{s}_c(\mathbf{x}^i)$ denotes the original snapshot of the cth capability quantity for the ith damage case, while $\tilde{\mathbf{s}}_c(\boldsymbol{\alpha}_{\mathrm{g}}^i)$ indicates its final estimate obtained online.

$(s_c)_{\max}$ and $(s_c)_{\min}$ are the maximum and minimum values in snapshot $\mathbf{s}_c(\mathbf{x}^i)$, respectively. For our application $C=1$, so we drop the superscript and refer to the single accuracy metric $E_{\mathbf{s}}$.

Runtime.. Along with accuracy, we are interested in the computational cost associated to the four online steps of our procedure. In particular, we wish to achieve a very effective online processing to obtain informative first-cut capability estimates at minimum computational time. Therefore, in this paper we monitor the runtime, defined as the CPU time in seconds to complete a certain computational step. Specifically, for each validation case $i = 1, ..., n_{\mathrm{v}}$ we record the CPU time t_p^i to compute the pth online step; then we average over the entire validation set to obtain a representative runtime T_p for each computational stage:

$$T_p = \frac{1}{n_{\mathrm{v}}} \sum_{i=1}^{n_{\mathrm{v}}} t_p^i \qquad \forall p. \tag{20}$$

The computational cost of the entire online procedure is denoted by the overall mean runtime $T_{\mathrm{on}} = \sum_{p=1}^{4} T_p$. The procedure is implemented in MATLAB and all test cases are run on an Intel Core i7-2600 at 3.40 GHz.

5. Results and discussion

In this paper we demonstrate the efficiency of the multistep ROM strategy for the case of structural assessment from sparse measurements affected by uncertainties. This specific problem can be described in terms of quantity and quality of sensor information/measurements that feed the online process. In particular, the quantity of sensed data is regulated by two algorithmic parameters, namely the fraction f_{p} of online grid point measurements with respect to the complete data available offline, and the number n_α^{g} of modal coefficients to reconstruct via GPOD. In other words, the quality of sensor measurements depends on uncertainty contributions due to poor knowledge about the exact location of the sensors and/or due to the noise affecting the measurements. This section develops three parametric explorations whose outcomes are discussed following the flow of online computational steps.

5.1. Quantity of sensor information

The first parametric exploration investigates various parameter settings that determine the actual amount of sensed information to process on-board. We evaluate 490 combinations of parameter values resulting from a full factorial exploration that considers 49 levels for the fraction of measured information f_{p}, and 10 levels for the number of modal coefficients recovered with gappy POD n_α^{g} (Table 2).

For the wing panel of Section 3.1, the fraction f_{p} is defined with respect to a complete offline snapshot of $n_{\mathrm{e}} = 3921$ elements; accordingly, $f_{\mathrm{p}} = 0.1\%$ indicates that strain components are measured by sensors randomly placed in only 4 grid points out of 3921, leading to 99.9% of missing data. In Table 2,

Factors	Levels	Values	Units
$f_{\mathrm{p}} = \frac{n_\ell}{n_{\mathrm{e}}} \times 100$	49	0.1:0.1:2.0 2.5:0.5:5.0 6:1:10 15:5:100	[%]
$n_\alpha^{\mathrm{g}} = \sum_{m=1}^{M} n_m^{\mathrm{g}}$	10	6 9 15 30 60 90 150 225 300 471	[-]

Table 2: Parameters that regulate the quantity of sensor information: fraction of measured information (f_{p}, with n_ℓ denoting the number of grid points measured online) and number of modal terms to recover with GPOD (n_α^{g}). The full factorial exploration yields 490 combinations of factor values. Notation $a:c:b$ indicates that we consider values in $[a, b]$ with step c.

the smallest value of $n_\alpha^{\mathrm{g}} = 6$ indicates that we reconstruct only the first two modal coefficients of each measured component of strain (since $n_\alpha^{\mathrm{g}} = \sum_{m=1}^{M} n_m^{\mathrm{g}}$ and $n_m^{\mathrm{g}} = 2\ \forall m$). The largest value of n_α^{g} is determined by the number of modal terms needed to recover the 95% of energy of each measured quantity of interest. For the three quantities of interest, we have $\mathcal{E}(n_m) = 0.95$ when $n_1 = 176$, $n_2 = 152$ and $n_3 = 143$, thus we obtain $n_\alpha^{\mathrm{g}} = n_\alpha = 471$.

Figures 5, 6, 7, 8, and 9 illustrate the assessment metrics defined in Section 4 and computed for all combinations of the parameters in Table 2. The different settings are sorted by n_α^{g}; the 10 sections are clearly visible, each comprising 49 results obtained for different values of f_{p}.

Figure 5 shows the fractions of computational time T_p in ms dedicated to each online step. The area diagram illustrates that the first online step has sensitive scaling to the number of reconstructed modal coefficients. We investigated numerous algorithmic settings corresponding to different values of tuning parameters ($n_{\mathrm{w}}, n_\alpha^{\mathrm{rs}}, p$, and m_β): we observed that the first online step is responsible for the largest runtime fractions. Therefore, in order to contain the overall online processing time, we wish to reduce as much as possible the number of measurement modal coefficients to recover with GPOD. The present study shows the possibility to efficiently act on this parameter without compromising the quality of the final capability predictions.

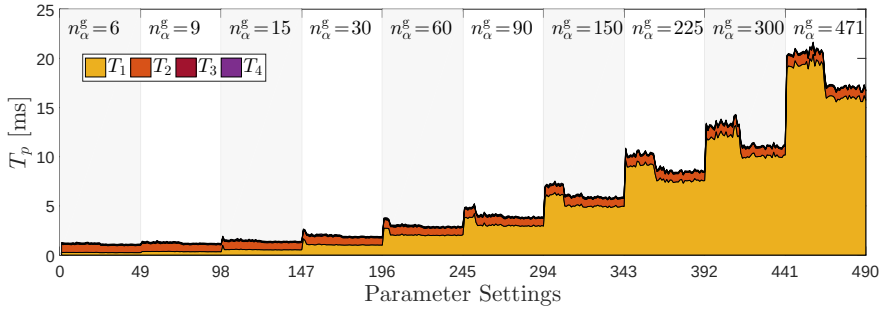

Figure 5: Computational time T_p dedicated to each online step for 490 parameter combinations (Table 2). For each n_α^{g}, settings are sorted by increasing fraction of measurements f_{p}.

Figure 6 depicts accuracy metrics E_α^1, E_α^2, and E_α^3 assessing the first online step. The three measured quantities of interest present similar trends: the

accuracy of the coefficients' reconstruction generally improves by adding sensed data f_{p}, although such enhancement is more effective when only a few modal terms are recovered. In fact, higher order terms are characterized with higher spatial frequencies that need larger fractions of measured information to be captured.

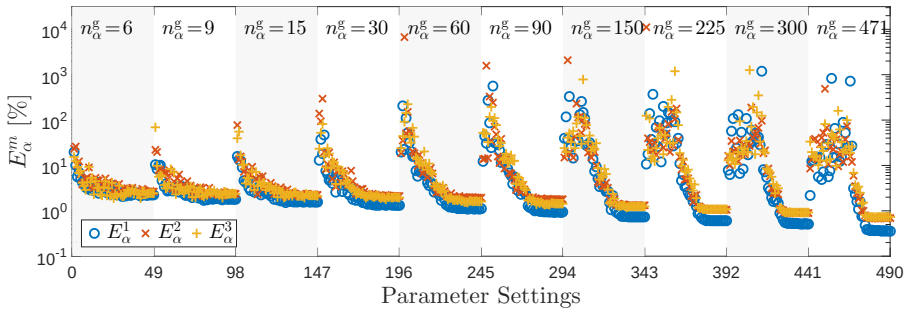

Figure 6: Measurement coefficients reconstruction. Accuracy metrics E_α^m for 490 parameter combinations (Table 2). For each n_α^{g}, settings are sorted by increasing fraction of measurements f_{p}.

In addition, sparse measurements (small f_{p}) and numerous modes to recover (large n_α^{g}) together compromise the conditioning of the gappy matrix, leading to high computational effort and poor accuracy. These phenomena can be observed by comparing the trends for different n_α^{g}. For $n_\alpha^{\mathrm{g}}=6$ we have a faster decay of the error by adding more sensors and we obtain $E_\alpha^m < 5\%\ \forall m$ for $f_{\mathrm{p}} \geq 1.5\%$ in $T_{\mathrm{on}} \simeq 1.1ms$ ($T_1 < 0.3ms$); in contrast, for $n_\alpha^{\mathrm{g}}=471$ the error decay is delayed and we need at least $f_{\mathrm{p}} \geq 7\%$ and $T_{\mathrm{on}} \simeq 18ms$ ($T_1 > 16ms$) to reach similar accuracy values.

We also observe that recovering more terms allows the errors to be lowered below 1% for $f_{\mathrm{p}} \geq 30\%$, but this performance comes at the expense of a 10 to 20 times larger computational effort. Therefore, the challenge is now to exploit at best the information contained within the dominant modal terms. To this end, the scaled L_2 norm introduced in Equation 6 plays an important role. Recall that this scaled norm is used as a dissimilarity measure for both the offline clustering and the online classification steps.

Figure 7 compares the online coefficient of determination R^2_{on} (computed over the validation set) to the reference offline value $R^2_{\mathrm{off}} = 0.84$ (computed over the evaluation set). R^2_{on} generally increases with the fraction of measured data f_{p}, but the actual trend differs with n_α^{g} and nearly reflects the accuracy measures recorded for the first step (Figure 6). However, in contrast with the reconstruction step, here we obtain better performance with small n_α^{g} even for larger fractions f_{p}. Hence, faster settings ($n_\alpha^{\mathrm{g}} = 6, 9, 15$) now outperform expensive settings—not only in terms of computational costs, but also in terms of accuracy.

This result is due to the dissimilarity measure (6) that prioritizes the characterization of the most informative dimensions, which correspond to the dominant

modes of each quantity of interest. This measure allows us to restrict the size of the SOM network by efficiently placing the prototypes in the space of POD coefficients; as a consequence, we obtain a faster classification step by reducing the computational effort to identify the closest cluster. In addition, this dissimilarity measure permits us to efficiently leverage the information content of the most energetic modal terms such that just a few of them are sufficient to achieve good classifications.

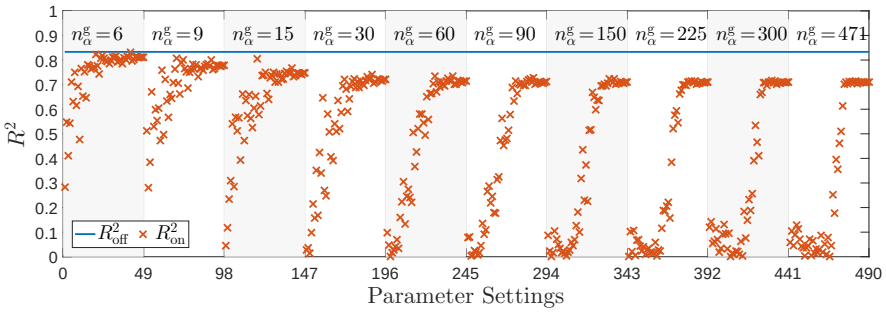

Figure 7: Classification. The online coefficient of determination R^2_{on} is compared to the offline reference value R^2_{off} for 490 parameter combinations (Table 2). For each n_α^g, settings are sorted by increasing fraction of measurements f_p.

Once the right cluster is determined, local approximations are employed to recover the capability POD coefficients. Being in the correct cluster means that the set of local models is the most representative one in the library. Figure 8 depicts accuracy measure E_β assessing the local approximations at the third online step. The trend of E_β is similar for all values of n_α^g: this demonstrates again that the classification step does a good job in exploiting properly the most dominant modes. As a consequence, we can consider the possibility to reconstruct only the POD measurements actually required to compute the local $\tilde{\beta}_j^c(\boldsymbol{\alpha}_g)^{(k^*)}$, i.e., we can set $n_\alpha^g = n_\alpha^{rs}$.

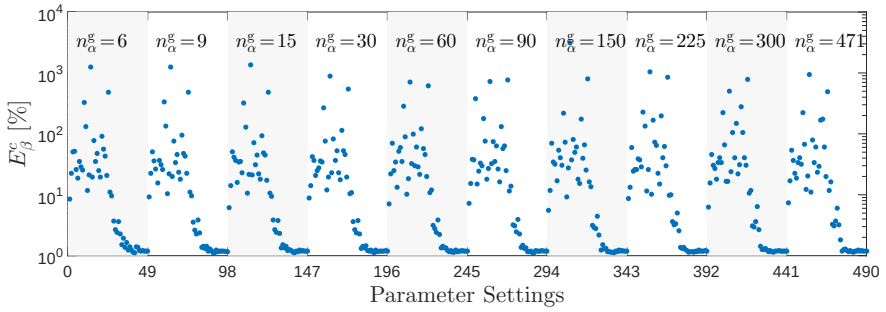

Figure 8: Local approximation. Accuracy metrics E_β for 490 parameter combinations (Table 2). For each n_α^g, settings are sorted by increasing fraction of measurements f_p.

Figure 9 illustrates the accuracy measure $E_{\mathbf{s}}$ assessing the fourth online

step. Similarly to E_β, the final capability estimate presents analogous trends for all levels of n_α^{g}. Much larger variability is instead recorded for different values of f_{p}, as shown in Figure 10. The reason for that behavior is depicted in Figure 11 where we observe: (i) the number of validation damage cases for which the final POD capability expansion includes a certain modal component, and (ii) how this cardinality evolves as we increase the fraction of measured data f_{p}. For $f_{\mathrm{p}} > 10\%$ the cardinality collected for the dominant modes reveals a quantization that resembles the original capability decomposition and the fraction of variance explained by each component. These results confirm that both classification and local approximations work appropriately. For $f_{\mathrm{p}} < 5\%$ the quantization appears much less defined because sensor measurements are insufficient or even misleading due to their sparsity.

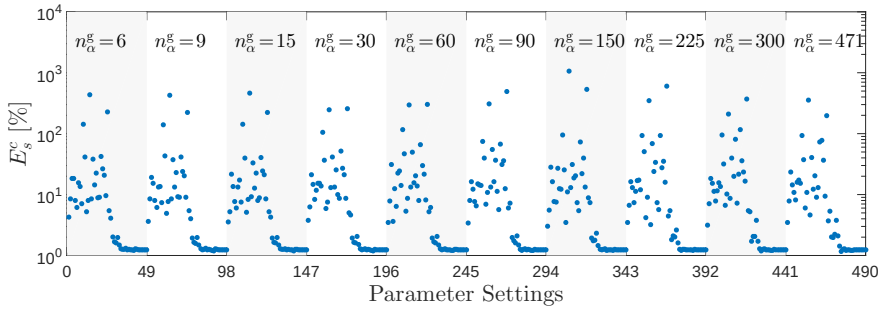

Figure 9: Capability estimate. Accuracy metric $E_{\mathbf{s}}$ for 490 parameter combinations (Table 2). For each n_α^{g}, settings are sorted by increasing fraction of measurements f_{p}.

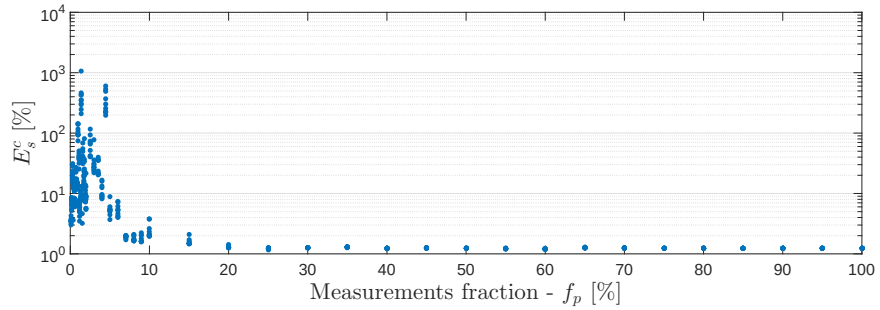

Figure 10: Capability estimate. Accuracy metric $E_{\mathbf{s}}$ for 490 parameter combinations (Table 2) plotted with respect to the fraction f_{p} of grid points measured online.

Taking into account these outcomes, for the algorithmic setup described in Section 3.2 we can adopt $n_\alpha^{\mathrm{g}} = n_\alpha^{\mathrm{rs}}$ to recover only the coefficients $\boldsymbol{\alpha}^{\mathrm{rs}} = [\alpha_1^1, \alpha_2^1, \alpha_1^2, \alpha_2^2, \alpha_1^3, \alpha_2^3]$ in $T_1 < 0.3ms$. This choice (settings 1 to 49 in Figures 5, 6, 7, 8, and 9) allows us to take advantage of the reduced overall computational cost: $T_{\mathrm{on}} \simeq 1.1ms$ is indeed a tiny fraction ($1/10^5$ to $1/10^4$) of the time typically

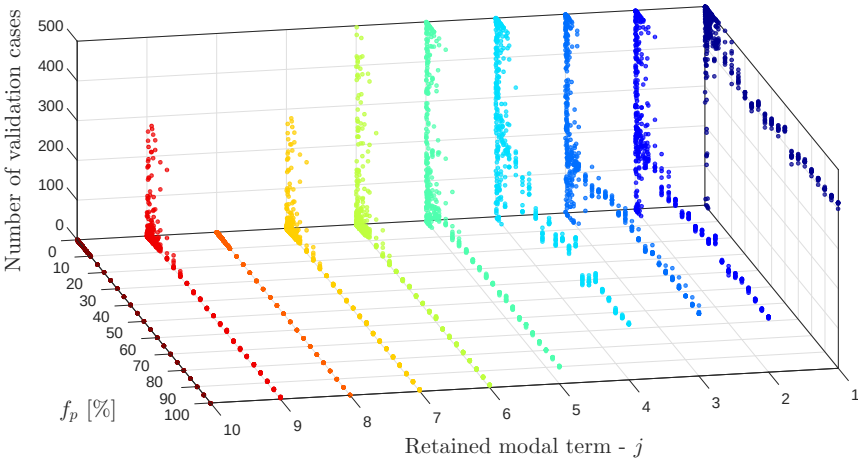

Figure 11: For all given pairs $(f_{\mathrm{p}}, n_{\alpha}^{\mathrm{g}})$, number of damage cases in the validation set that retain the jth modal term in the final POD expansion of $\tilde{\mathbf{s}}_c(\boldsymbol{\alpha}_{\mathrm{g}})$.

required to solve the full-order problem associated to capability evaluation only. Even more, it permits large time savings with respect to the entire original evaluation flow that includes the expensive parameter identification problem (Figure 1). In terms of accuracy, good results are obtained even with sparse measurement data. For example, $f_{\mathrm{p}} = 5\%$ yields $E_{\mathbf{s}} < 6\%$. Adding sensors, we achieve $E_{\mathbf{s}} < 2\%$ with $f_{\mathrm{p}} = 10\%$ (90% of missing data), and $E_{\mathbf{s}} < 1.5\%$ for $f_{\mathrm{p}} \geq 20\%$.

5.2. Quality of sensor information

In this section we conduct two parameter explorations to analyze the robustness of the proposed MultiStep-ROM procedure to uncertainties about sensor locations and measured values. The goal is to understand how the quality of sensor data may affect the accuracy of the online evaluations.

Robustness to sensor locations/positioning.. This study aims to assess the effects of uncertainties about the exact positions of sensors. To do that, we design an exploration that combines full factorial sampling of quantity parameters f_{p} and n_{α}^{g}, and Monte Carlo sampling of sensor locations. For each given pair of parameters $(f_{\mathrm{p}}, n_{\alpha}^{\mathrm{g}})$ we consider multiple random placements of n_{ℓ} sensors, where the number n_{ℓ} is determined by f_{p}. We define 7 levels for factor n_{α}^{g} and 10 levels for factor f_{p} (Table 3). Then, for each measurement fraction f_{p} we consider $n_{\mathrm{p}} = 50$ random placements of the corresponding available sensors. These random positions are selected from among all $n_{\mathrm{e}} = 3921$ grid points, with each grid point having equal probability to be picked as a sensor site. This approach yields larger location uncertainties for larger percentages of missing data, because sensor positions can differ significantly between random placements. Conversely, uncertainty progressively reduces by adding more sensors to increase the fraction of measured grid points f_{p}.

Factors	Levels	Values	Units
$f_\mathrm{p} = \frac{n_\ell}{n_\mathrm{e}} \times 100$	10	0.1 0.2 0.5 1 2 5 10 20 50 100	[%]
$n_\alpha^\mathrm{g} = \sum_{m=1}^{M} n_m^\mathrm{g}$	7	6 9 15 30 60 150 471	[-]

Table 3: Parameters that regulate the quantity of sensor information: fraction of measured information (f_p, with n_ℓ denoting the number of grid points measured online) and number of modal terms to recover with GPOD (n_α^g). The full factorial exploration yields 70 combinations of factor values.

The assessment metrics defined in Section 4 are evaluated for all the resulting 3500 settings. To visualize the outcomes, for each pair $(f_\mathrm{p}, n_\alpha^\mathrm{g})$ we monitor three error indicators computed over the n_p random placements, namely mean, maximum and minimum values of the accuracy measures:

$$\begin{aligned}
(E(f_\mathrm{p}, n_\alpha^\mathrm{g}))_\mathrm{mean} &= \frac{1}{n_\mathrm{p}} \sum_{i=1}^{n_\mathrm{p}} E(f_\mathrm{p}, n_\alpha^\mathrm{g}, S_\mathrm{p}^i(f_\mathrm{p})), \\
(E(f_\mathrm{p}, n_\alpha^\mathrm{g}))_\mathrm{max} &= \max_{i=1,\dots,n_\mathrm{p}} E(f_\mathrm{p}, n_\alpha^\mathrm{g}, S_\mathrm{p}^i(f_\mathrm{p})), \\
(E(f_\mathrm{p}, n_\alpha^\mathrm{g}))_\mathrm{min} &= \min_{i=1,\dots,n_\mathrm{p}} E(f_\mathrm{p}, n_\alpha^\mathrm{g}, S_\mathrm{p}^i(f_\mathrm{p})).
\end{aligned} \tag{21}$$

In Equation (21), E may denote any accuracy metric in Section 4, since these indicators can be computed for every online step; $S_\mathrm{p}^i(f_\mathrm{p})$ is the ith random selection of n_ℓ sensor locations for a given f_p. Figures 12 and 13 show these indicators for E_α^1, E_α^2, E_α^3, and $E_\mathbf{s}$, respectively. The results are plotted for all 70 combinations of parameters $(f_\mathrm{p}, n_\alpha^\mathrm{g})$. The different settings are sorted by n_α^g; the 7 sections are clearly visible, each comprising 10 results obtained for different values of f_p.

Figure 12 confirms that ill-conditioning of the gappy matrix occurs when we try to recover many modal terms with only a small amount of sensed data: it results in high reconstruction errors, as already noted in Section 5.1. All error indicators, from coefficient reconstruction (Figures 12) to capability prediction (Figure 13), present common characteristics of their distributions. The accuracy spread progressively reduces by increasing the number of measurement sites (f_p), until there is no spread for the case with sensors placed everywhere ($f_\mathrm{p} = 100\%$ – settings number 10, 20, 30, 40, 50, 60, and 70). In addition, taking into account that in Figures 12 and 13 the quantities are plotted on a logarithmic scale, we observe that the resulting distributions of error indicators are asymmetric and squeezed towards lower values. Large errors occur rarely, and can be avoided by adding sensors because more measured data contributes to cut the right tail of the distributions (Figure 14). This can also be seen in Figure 15, where the error indicators computed for $E_\mathbf{s}$ progressively collapse onto a single value by increasing the fraction of sensor measurements f_p.

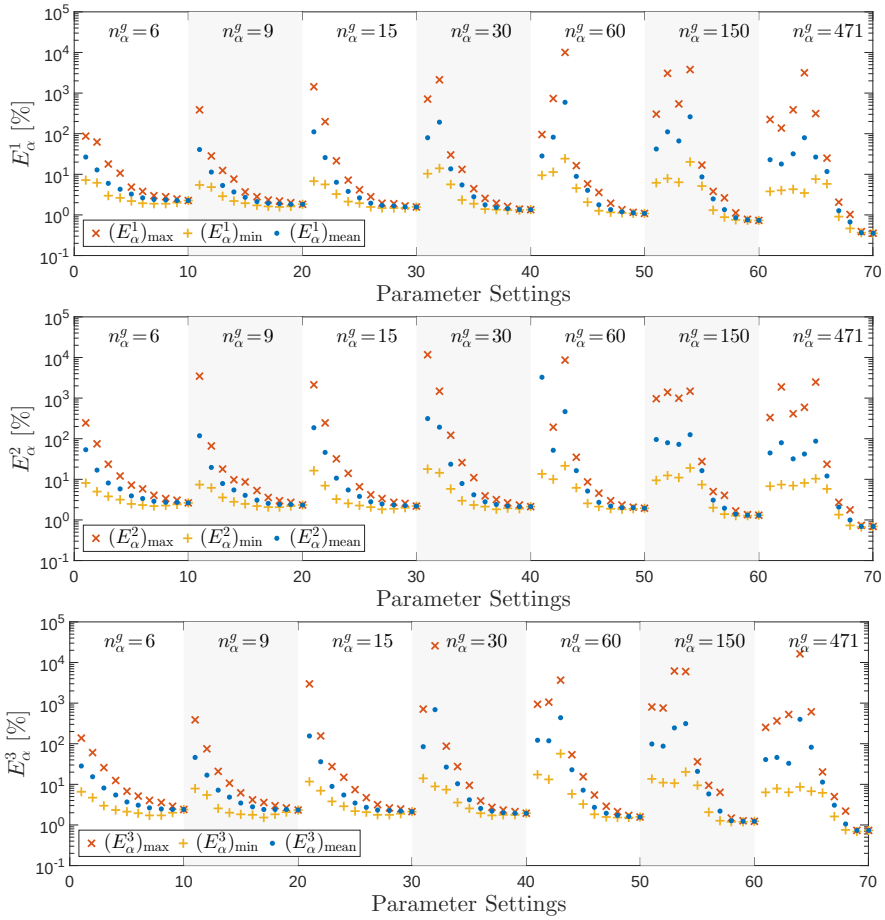

Figure 12: Measurement coefficients reconstruction. Error indicators $(E_\alpha^m)_{\text{mean}}$, $(E_\alpha^m)_{\max}$, and $(E_\alpha^m)_{\min}$ for 70 parameter combinations (Table 3). For each n_α^{g}, settings are sorted by increasing fraction of measurements f_{p}. Each plot depicts a single measured quantity of interest ($M = 3$).

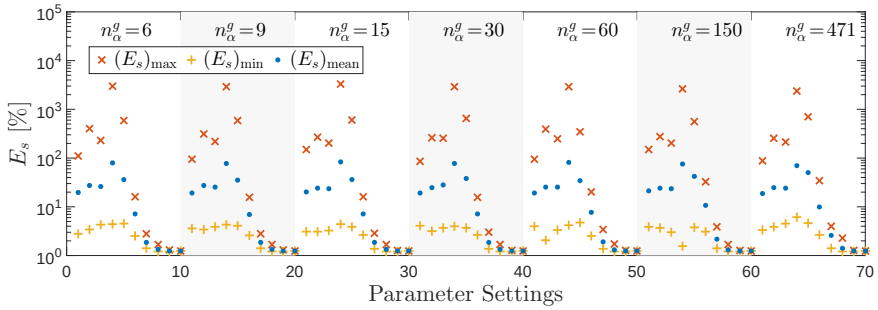

Figure 13: Capability estimate. Error indicators $(E_{\mathbf{s}})_{\text{mean}}$, $(E_{\mathbf{s}})_{\max}$, and $(E_{\mathbf{s}})_{\min}$ for 70 parameter combinations (Table 3). For each n_α^{g}, settings are sorted by increasing fraction of measurements f_{p}.

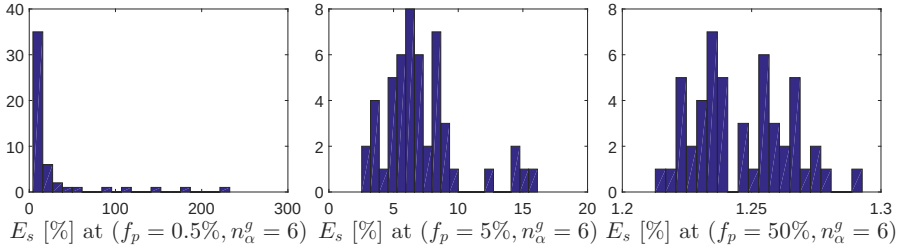

Figure 14: Capability estimate. Distributions of the $n_{\text{p}} = 50$ values of accuracy metric $E_{\mathbf{s}}$ for three specific settings. The diagrams refers to setting number 3, 6 and 9 in Figure 13, respectively.

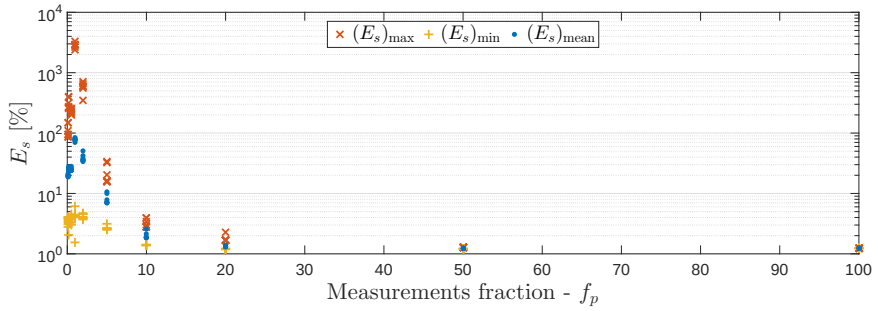

Figure 15: Capability estimate. Error indicators $(E_{\mathbf{s}})_{\text{mean}}$, $(E_{\mathbf{s}})_{\max}$, and $(E_{\mathbf{s}})_{\min}$ for 70 parameter combinations (Table 3) plotted with respect to the fraction f_{p} of grid points measured online.

In summary, the online procedure based on our MultiStep-ROM strategy appears robust to sensor locations for $f_{\text{p}} > 10\%$. In this case, choosing $f_{\text{p}} > 10\%$ yields $(E_{\mathbf{s}})_{\text{mean}} < 1.5\%$, and $(E_{\mathbf{s}})_{\max} < 2.5\%$ $\forall\ n_\alpha^{\text{g}}$. For $f_{\text{p}} < 10\%$ the limited sensed data can be misleading and large errors are possible, although rare in the cases studied.

Robustness to measurements noise.. The third analysis assesses the MultiStep-ROM strategy for the case of measurements affected by uncertainty. In particular we assume that measured quantities of interest are corrupted with i.i.d. additive Gaussian noise such that

$$\check{\mathbf{q}}_m = \hat{\mathbf{q}}_m + \mathcal{N}(0, \boldsymbol{\sigma}^2) \qquad m = 1, ..., M, \tag{22}$$

where $\mathcal{N}(0, \boldsymbol{\sigma}^2)$ is the Gaussian noise with zero mean and variance σ^2. We consider standard deviation σ as an additional parameter in this last investigation. We design an exploration that combines full factorial sampling of parameters f_p, n_α^g and σ, with Monte Carlo sampling of noise contributions. In particular we define 7 levels for factor n_α^g, 10 levels for factor f_p, and 5 levels for factor σ (Table 4); then, for each $\sigma > 0$ we consider $n_\mathrm{n} = 50$ random sampling of noise contribution according to $\mathcal{N}(0, \boldsymbol{\sigma}^2)$.

Factors	Levels	Values	Units
$f_\mathrm{p} = \frac{n_\ell}{n_\mathrm{e}} \times 100$	10	0.1 0.2 0.5 1 2 5 10 20 50 100	[%]
$n_\alpha^\mathrm{g} = \sum_{m=1}^{M} n_m^\mathrm{g}$	7	6 9 15 30 60 150 471	[-]
σ	5	0 10 20 50 100	$[\mu\varepsilon]$

Table 4: Parameters that regulate the quantity and quality of sensor information: fraction of measured information (f_p, with n_ℓ denoting the number of grid points measured online), number of modal terms to recover with GPOD (n_α^g), and standard deviation of the Gaussian model for measurement noise (σ). The full factorial exploration yields 350 combinations of factor values.

For the wing panel described in Section 3, measured quantities of interest assume pointwise values over four orders of magnitude, from $10^{-1}\mu\varepsilon$ to $10^3\mu\varepsilon$. For the damage cases collected in the evaluation and validation sets, average values of strain components range from 10^2 to $10^3\mu\varepsilon$. Therefore, $\sigma = 10\mu\varepsilon$ covers common sensor gage accuracy of $1 - 5\%$, while $\sigma = 100\mu\varepsilon$ corresponds to much less probable corruption ranges of $10 - 100\%$, with noise of the same order of magnitude as the value to measure.

The assessment metrics defined in Section 4 are evaluated for all resulting 14070 conditions. To visualize the outcomes, for each triple $(f_\mathrm{p}, n_\alpha^\mathrm{g}, \sigma)$ we monitor three error indicators computed over the n_n noise contributions, namely mean, maximum and minimum values of the accuracy measures:

$$\begin{aligned}
(E(f_\mathrm{p}, n_\alpha^\mathrm{g}, \sigma))_\mathrm{mean} &= \frac{1}{n_\mathrm{n}} \sum_{i=1}^{n_\mathrm{n}} E(f_\mathrm{p}, n_\alpha^\mathrm{g}, \sigma, S_\mathrm{n}^i(\sigma)), \\
(E(f_\mathrm{p}, n_\alpha^\mathrm{g}, \sigma))_\mathrm{max} &= \max_{i=1,\dots,n_\mathrm{n}} E(f_\mathrm{p}, n_\alpha^\mathrm{g}, \sigma, S_\mathrm{n}^i(\sigma)), \\
(E(f_\mathrm{p}, n_\alpha^\mathrm{g}, \sigma))_\mathrm{min} &= \min_{i=1,\dots,n_\mathrm{n}} E(f_\mathrm{p}, n_\alpha^\mathrm{g}, \sigma, S_\mathrm{n}^i(\sigma)),
\end{aligned} \tag{23}$$

where E may denote any accuracy metric introduced in Section 4. $S_\mathrm{n}^i(\sigma)$ is the ith sample of noise contributions from distribution $\mathcal{N}(0, \boldsymbol{\sigma}^2)$. Figure 16 shows these indicators for E_α^1 (E_α^2 and E_α^3 behave in a similar way), and Figure 17

shows these indicators for $E_{\mathbf{s}}$. The results are plotted for all 350 combinations of parameters $(f_{\mathrm{p}}, n_{\alpha}^{\mathrm{g}}, \sigma)$. The different settings are sorted by σ, whose five sections are clearly visible; for each value of σ we distinguish ten sections associated to f_{p}, each comprising 7 results obtained for different values of n_{α}^{g}. For $\sigma = 0\mu\varepsilon$ (settings 1 to 70) all the indicators defined in Equation 23 coincide and their points are superimposed.

In Figure 16 the error indicators associated to $\sigma = 10\mu\varepsilon$ and $\sigma = 20\mu\varepsilon$ resemble the results obtained without any noise ($\sigma = 0\mu\varepsilon$). Hence, the first reconstruction step is very robust for common values of measurement uncertainty. Only when the noise is of the same order as the quantity to measure do we start to see the errors magnify.

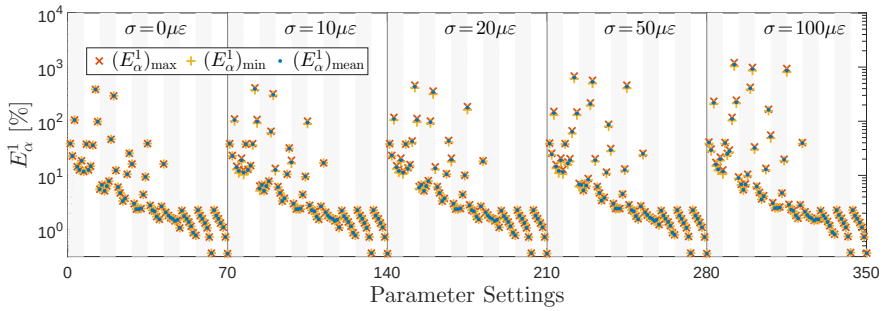

Figure 16: Measurement coefficients reconstruction. Error indicators $(E_{\alpha}^{m})_{\mathrm{mean}}$, $(E_{\alpha}^{m})_{\max}$, and $(E_{\alpha}^{m})_{\min}$ for 350 parameter combinations (Table 4). For each noise distribution (with standard deviation σ), settings are sorted by increasing fraction of measurements f_{p}; for each fraction of measured grid points f_{p}, settings are sorted by increasing number of reconstructed modal terms n_{α}^{g}. In particular, the diagram illustrates E_{α}^{1}.

Figure 17 depicts the error indicators computed for the final capability estimate. It shows that the error progressively increases with higher level of noise. This is mainly due to the error introduced by the local RS approximations of POD coefficients β_{j}^{c}, which are sensitive to the value of their input $\boldsymbol{\alpha}_{\mathrm{g}}^{\mathrm{rs}}$. Hence, this increase in the error is more relevant for cases with $f_{\mathrm{p}} < 10\%$, and, as before, we can achieve a significant mitigation of the uncertainty effects by adding sensors. For $f_{\mathrm{p}} \geq 10\%$ estimate accuracy progressively improves and assumes similar values for all the noise distributions: settings 43-70 ($\sigma = 0\mu\varepsilon$), 113-140 ($\sigma = 10\mu\varepsilon$), 183-210 ($\sigma = 20\mu\varepsilon$), 253-210 ($\sigma = 50\mu\varepsilon$), and 323-350 ($\sigma = 100\mu\varepsilon$) record $E_{\mathbf{s}} < 3\% \;\; \forall \; n_{\alpha}^{\mathrm{g}}$ and $E_{\mathbf{s}} < 2\%$ for $n_{\alpha}^{\mathrm{g}} \leq 60$. Figure 18 illustrates that error indicators for $E_{\mathbf{s}}$ progressively approach the same values as the fraction of sensor measurements f_{p} is increased. Similar results are observed for the error spreads around the mean value. As shown in Figure 19, accuracy dispersion over the n_{n} cases increases with σ (along the column), but larger fractions of measured data f_{p} shrink the error distributions (along the row).

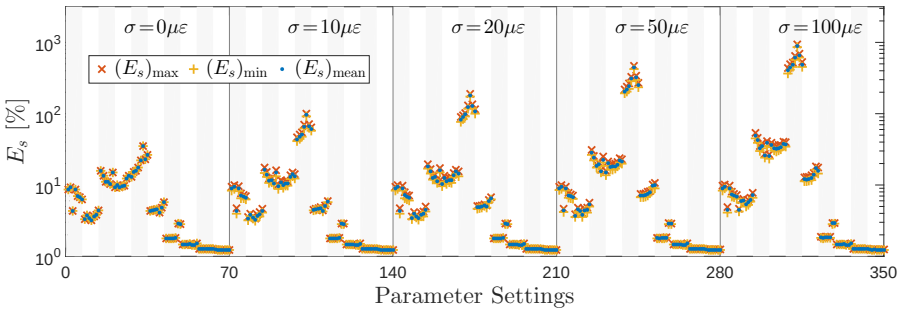

Figure 17: Capability estimate. Error indicators $(E_{\mathbf{s}})_{\mathrm{mean}}$, $(E_{\mathbf{s}})_{\max}$, and $(E_{\mathbf{s}})_{\min}$ for 350 parameter combinations (Table 4). For each noise distribution (with standard deviation σ), settings are sorted by increasing fraction of measurements f_{p}; for each fraction of measured grid points f_{p}, settings are sorted by increasing number of reconstructed modal terms n_{α}^{g}.

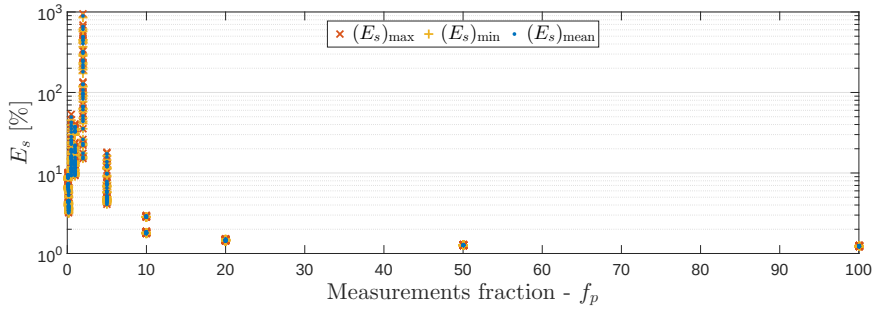

Figure 18: Capability estimate. Error indicators $(E_{\mathbf{s}})_{\mathrm{mean}}$, $(E_{\mathbf{s}})_{\max}$, and $(E_{\mathbf{s}})_{\min}$ for 350 parameter combinations (Table 4) plotted with respect to the fraction f_{p} of grid points measured online.

In conclusion, online evaluations are robust to uncertainties affecting measured values, in particular for $f_{\mathrm{p}} \geq 10\%$. For incomplete measurements presenting more than 90% of missing data, evaluation accuracy becomes more sensitive to noise, because sensed information is insufficient or even misleading to characterize the structural state. In contrast to the important role of f_{p}, the number of modal coefficients to recover, n_{α}^{g}, does not play a significant role in presence of uncertainties. This is mainly due to the underlying algorithmic construction, which prioritizes the informative contributions of the dominant modes. This strategy benefits from the dominant modes being less sensitive to small perturbations of measurement values. Small variations are indeed captured by higher order terms, but can be misleading in presence of uncertainties; therefore, with this implementation of our MultiStep-ROM strategy their contributions are either progressively mitigated or completely discarded.

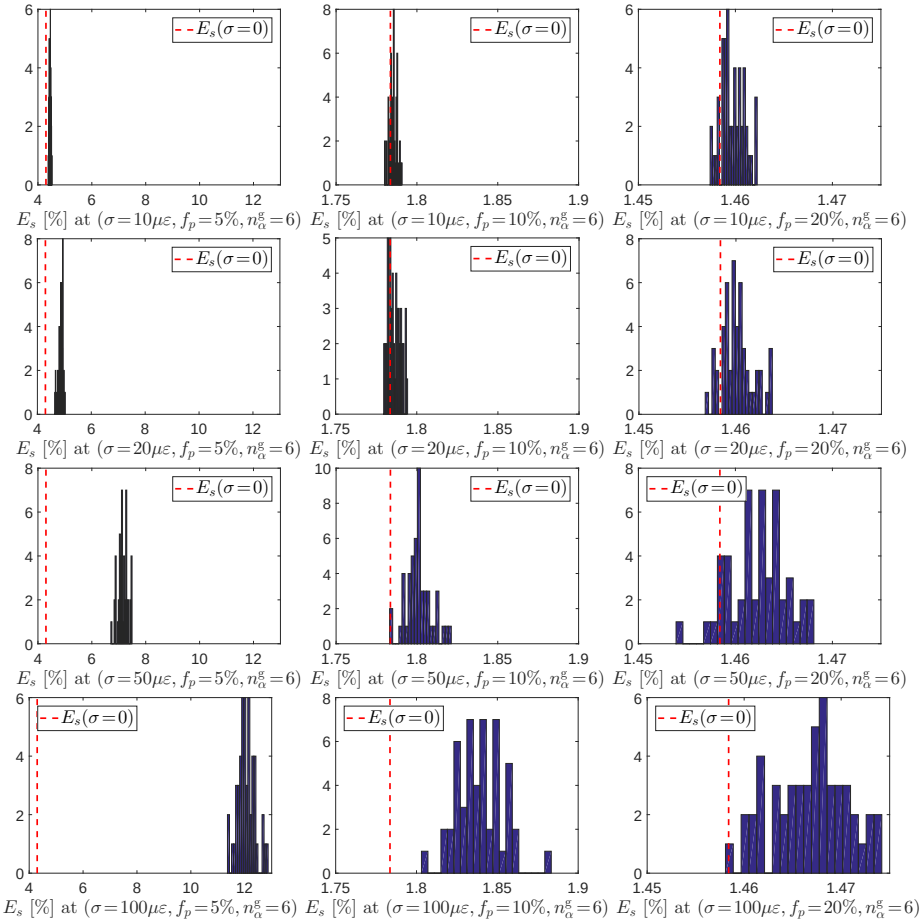

Figure 19: Capability estimate. Distributions of the $n_\mathrm{n}=50$ values of accuracy metric $E_\mathbf{s}$ for twelve specific settings. Each plot includes a dashed reference line that indicates the value of $E_\mathbf{s}$ obtained for the same pair of quantity parameters $(f_\mathrm{p}, n_\alpha^\mathrm{g})$ without measurement noise ($\sigma=0$). Accuracy dispersion over the n_n cases increases with σ (along the column), but larger fractions of measured data f_p shrink the error distributions (along the row).

6. Concluding Remarks

This work discusses the use of our MultiStep-ROM strategy to assist real-time decisions associated with structural assessment tasks and informed by incomplete and noisy measurements. The representative application considered in this paper is a UAV wing panel tested over a variety of local degradations of the structural properties. This testbed lets us explore the issues of quantity and quality of sensor measurements, which are relevant to a broad class of problems. Quantity and quality of measured data are here regulated by a few governing parameters: the fraction f_p of online grid point measurements with

respect to the complete data available offline, the number of modal coefficients n_α^{g} to reconstruct via GPOD, and the standard deviation σ of the Gaussian model of noise. The studies presented in this paper rely on three explorations of these parameters to assess advantages and limitations of the MultiStep-ROM procedure when applied to this class of problems.

The first investigation focuses on the quantity parameters, namely f_{p} and n_α^{g}. The results demonstrate the possibility to reduce the evaluation time through controlling n_α^{g}, the number of modal coefficients to recover via GPOD. In contrast, the accuracy of the final estimate is not affected by n_α^{g}, because the online process benefits from the use of a scaled L_2-norm as a dissimilarity measure for both offline clustering and online classification. For the structural assessment application discussed in this paper, we can evaluate capabilities in $T_{\mathrm{on}} \simeq 1.1ms$ and achieve prediction errors lower than 6% from incomplete on-board measurements with 95% of missing data. In addition, capability estimates improve by adding sensors and, at the same computational cost, it is possible to obtain errors lower than 2% with 90% of missing data.

The second and third explorations focus on the quality of sensor information and consider uncertainty contributions due to (i) lack of knowledge about the actual position of sensors on the measurement grid and (ii) intrinsic accuracy of the sensor gages. The results revealed that, for measurements with less than 90% of missing data, online evaluations are robust to both sensor locations and sensor accuracy. Conversely, if the sensors cover less than the 10% of the reference grid points, measured data may be misleading or insufficient to properly represent the state of the system. This limitation is partially due to the random positioning of the sensors we adopted in this paper; optimal sensor placement strategies are an important area of future work that may address this limitation.

Acknowledgements

This work was supported by the U.S. Air Force Office of Scientific Research grant FA9550-16-1-0108 under the Dynamic Data Driven Application System Program, Program Manager Dr. Frederica Darema, and by the MIT-SUTD Collaboration under the partnership program between Massachusetts Institute of Technology and Singapore University of Technology and Design.

References

[1] D. Huston, Structural sensing, health monitoring, and performance evaluation, CRC Press, 2011.

[2] C. Papadimitriou, Optimal sensor placement methodology for parametric identification of structural systems, Journal of sound and vibration 278 (4) (2004) 923–947.

[3] A. M. Bruckstein, D. L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Review 51 (1) (2009) 34–81.

[4] F. Abazarsa, S. F. Ghahari, F. Nateghi, E. Taciroglu, Responseonly modal identification of structures using limited sensors, Structural Control and Health Monitoring 20 (6) (2013) 987–1006.

[5] L. He, J. Lian, B. Ma, H. Wang, Optimal multiaxial sensor placement for modal identification of large structures, Structural Control and Health Monitoring 21 (1) (2014) 61–79.

[6] Y. Yang, S. Nagarajaiah, Robust data transmission and recovery of images by compressed sensing for structural health diagnosis, Structural Control and Health Monitoring.

[7] U. Yildirim, O. Oguz, N. Bogdanovic, A prediction-error-based method for data transmission and damage detection in wireless sensor networks for structural health monitoring, Journal of Vibration and Control 19 (15) (2013) 2244–2254.

[8] N. Bogdanovic, D. Ampeliotis, K. Berberidis, F. Casciati, J. Plata-Chaves, Spatio-temporal protocol for power-efficient acquisition wireless sensors based shm, Smart Structures and Systems 14 (1) (2014) 1–16.

[9] P. Cawley, R. D. Adams, The location of defects in structures from measurements of natural frequencies, Journal of Strain Analysis for Engineering Design 14 (2) (1979) 49–57.

[10] S. W. Doebling, C. R. Farrar, M. B. Prime, D. W. Shevitz, Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review, Tech. rep., Los Alamos National Lab., NM (United States) (1996).

[11] O. S. Salawu, Detection of structural damage through changes in frequency: a review, Engineering Structures 19 (9) (1997) 718–723.

[12] Y. Yan, L. Cheng, Z. Y. Wu, L. Yam, Development in vibration based structural damage detection technique, Mechanical Systems and Signal Processing 21 (2007) 2198–2211.

[13] J.-H. Park, J.-T. Kim, J.-H. Yi, Output-only modal identification approach for time-unsynchronized signals from decentralized wireless sensor network for linear structural systems, Smart Structures and Systems 7 (1) (2011) 59–82.

[14] C.-H. Loh, K. J. Loh, Y.-S. Yang, W.-Y. Hsiung, Y.-T. Huang, Vibrationbased system identification of wind turbine system, Structural Control and Health Monitoring.

[15] S. Nagarajaiah, Y. Yang, Modeling and harnessing sparse and low-rank data structure: a new paradigm for structural dynamics, identification, damage detection, and health monitoring, Structural Control and Health Monitoring.

[16] M. Sanayei, O. Onipede, Damage assessment of structures using static test data, AIAA Journal 29 (7) (1991) 1174–1179.

[17] M. Sanayei, O. Onipede, S. R. Babu, Selection of noisy measurement locations for error reduction in static parameter identification, AIAA Journal 30 (9) (1992) 2299–2309.

[18] M. R. Banan, M. R. Banan, K. D. Hjelmstad, Parameter estimation of structures from static response. I. Computational aspects, Journal of Structural Engineering 120 (11) (1994) 3243–3258.

[19] M. R. Banan, M. R. Banan, K. D. Hjelmstad, Parameter estimation of structures from static response. II. Numerical simulation studies, Journal of Structural Engineering 120 (11) (1994) 3259–3283.

[20] Z. P. Szewczyk, P. Hajela, Damage detection in structures based on feature-sensitive neural networks, Journal of Computing in Civil Engineering 8 (2) (1994) 163–178.

[21] M. Sanayei, M. J. Saletnik, Parameter estimation of structures from static strain measurements. I: Formulation, Journal of Structural Engineering 122 (5) (1996) 555–562.

[22] M. Sanayei, M. J. Saletnik, Parameter estimation of structures from static strain measurements. II: Error sensitivity analysis, Journal of Structural Engineering 122 (5) (1996) 563–572.

[23] G. W. Housner, L. A. Bergman, T. K. Caughey, A. G. Chassiakos, R. O. Claus, S. F. Masri, R. E. Skelton, T. T. Soong, B. F. Spencer, J. T. P. Yao, Structural control: Past, present and future, Journal of Engineering Mechanics 123 (9) (1997) 897–971.

[24] A. Raghavan, C. E. S. Cesnik, Review of guided-wave structural health monitoring, The Shock and Vibration Digest 39 (2) (2007) 91–114.

[25] Y. Z. Fu, J. Liu, Z. Wei, Z. Lu, A two-step approach for damage identification in plates, Journal of Vibration and Control 22 (13) (2016) 3018–3031.

[26] L. Papadopoulos, E. Garcia, Structural damage identification: A probabilistic approach, AIAA Journal 36 (11) (1998) 2137–2145.

[27] S. Sankararaman, S. Mahadevan, Uncertainty quantification in structural damage diagnosis, Structural Control and Health Monitoring 18 (2011) 807–824.

[28] T. Peng, J. He, Y. Liu, A. Saxena, J. Celaya, K. Goebel, Fatigue damage diagnosis and prognosis using Bayesian updating, in: Proceedings of the 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Boston,MA, 2013.

[29] J. Rhim, S. W. Lee, A neural network approach for damage detection and identification of structures, Computational Mechanics 16 (6) (1995) 437–443.

[30] C. R. Farrar, S. W. Doebling, D. A. Nix, Vibration-based structural damage identification, Philosophical Transactions of The Royal Society Lond. A 359 (1778) (2001) 131–149.

[31] L. H. Yam, Y. J. Yan, J. S. Jiang, Vibration-based damage detection for composite structures using wavelet transform and neural network identification, Composite Structures 60 (2003) 403–412.

[32] J. J. Lee, J. W. Lee, J. H. Yi, C. B. Yun, H. Y. Jung, Neural networks-based damage detection for bridges considering errors in baseline finite element models, Journal of Sound and Vibration 280 (2005) 555–578.

[33] C. R. Farrar, K. Worden, Structural health monitoring: A machine learning perspective, Wiley, 2013.

[34] C. Zang, M. Imregun, Structural damage detection using artificial neural networks and measured frf data reduced via principal component projection, Journal of Sound and Vibration 242 (5) (2001) 813–827.

[35] S. Casciati, Response surface models to detect and localize distributed cracks in a complex continuum, Journal of Engineering Mechanics 136 (9) (2010) 1131–1142.

[36] S. E. Fang, R. Perera, Damage identification by response surface based model updating using D-optimal design, Mechanical Systems and Signal Processing 25 (2011) 717–733.

[37] S.-K. Hong, B. I. Epureanu, M. P. Castainer, D. J. Gorsich, Parametric reduced-order models for predicting the vibration response of complex structures with component damage and uncertainties, Journal of Sound and Vibration 330 (2011) 1091–1110.

[38] H. Y. Gao, X. L. Guo, X. F. Hu, Crack identification based on kriging surrogate model, Structural Engineering and Mechanics 41 (1) (2012) 25–41.

[39] M. J. Pais, F. A. C. Viana, N. H. Kim, Enabling high-order integration of fatigue crack growth with surrogate modeling, International Journal of Fatigue 43 (2012) 150–159.

[40] V. K. Hombal, S. Mahadevan, Surrogate modeling of 3D crack growth, International Journal of Fatigue 47 (2013) 90–99.

[41] L. Mainini, K. Willcox, Surrogate modeling approach to support real-time structural assessment and decision making, AIAA Journal 53 (6) (2015) 1612–1626.

[42] J. L. Lumley, The structure of inhomogeneous turbulent flows, in: Yaglom, Tatarsky (Eds.), Atmospheric Turbulence and Radio Wave Propagation, Macmillian, Moscow and Toulouse, 1967, pp. 166–178.

[43] L. Sirovich, Turbulence and the dynamics of coherent structures. part 1–3, Quarterly of Applied Mathematics 45 (3) (1987) 561–590.

[44] P. J. Holmes, J. L. Lumley, G. Berkooz, J. Mattingly, R. W. Wittenberg, Low-dimensional models of coherent structures in turbulence, Physics Reports 287 (4) (1997) 337–384.

[45] H. V. Ly, H. T. Tran, Modeling and control of physical processes using proper orthogonal decomposition, Journal of Mathematical and Computer Modeling.

[46] T. Kohonen, Self-Organizing Maps, 3rd Edition, Springer-Verlag, New York, 2001.

[47] E. Erwin, K. Obermayer, K. Schulten, Self-organizing maps: ordering, convergence properties and energy functions, Biological Cybernetics 67 (1992) 47–55.

[48] T. Kohonen, Things you haven't heard about the self-organizing map, in: Proc. IEEE Int. Conf on Neural Networks, San Francisco, 1993, pp. 1147–1156.

[49] T. Kohonen, E. Oja, O. Simula, A. Visa, J. Kangas, Engineering applications of the self-organizing map, Proceedings of the IEEE 84 (10) (1996) 1358–1384.

[50] P. Benner, S. Gugercin, K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, SIAM Review 57 (4) (2015) 483–531.

[51] M. Rathinam, L.R.Petzold, A new look at proper orthogonal decomposition, SIAM Journal of Numerical Analysis 41 (5) (2003) 1893–1925.

[52] W. H. Schilders, H. A. van deer Vorst, J. Rommes, Model order reduction: Theory, research aspects and applications, no. 13 in Mathematics in Industry, Springer-Verlag Berlin Heidelberg, 2008.

[53] C. Audouze, F. D. Vuyst, P. B. Nair, Reduced-order modeling of parameterized PDEs using time-space-parameter principal component analysis, International Journal for Numerical Methods in Engineering 80 (8) (2009) 1025–1057.

[54] C. Audouze, F. D. Vuyst, P. B. Nair, Nonintrusive reduced-order modeling of parametrized time-dependent partial differential equations, Numerical Methods for Partial Differential Equations 29 (5) (2013) 1587–1628.

[55] A. Quarteroni, G. Rozza, Reduced order methods for modeling and computational reduction, Vol. 9 of MS&A, Springer International Publishing, 2014.

[56] K. Kunisch, S. Volkwein, Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition, Journal of Optimization Theory and Applications 102 (2) (1999) 245–371.

[57] M. Hinze, S. Volkwein, Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control, in: D. Soresen, P. Benner, V. Mehrmann (Eds.), Dimension Reduction of Large-Scale Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 261–306.

[58] V. R. Algazi, D. J. Sakrison, On the optimality of the Karhunen-Loève expansion, IEEE Transactions on Information Theory 15 (2) (1969) 319–321.

[59] A. Dür, On the optimality of the discrete Karhunen–Loève expansion, SIAM Journal on Control an Optimizaton 36 (6) (1998) 1937–1939.

[60] T. W. Liao, Clustering of time series data – a survey, Pattern Recognition 38 (2005) 1857–1874.

[61] R. Xu, D. C. Wunsch, Survey of clustering algorithms, IEEE Transactions on Neural Networks 16 (3) (2005) 645–678.

[62] M. Filippone, F. Camastra, F. Masulli, S. Rovetta, A survey of kernel and spectral methods for clustering, Pattern Recognition 41 (2008) 176–190.

[63] A. K. Jain, Data clustering: 50 years beyond k-means, Pattern Recognition Letters 31 (2010) 651–666.

[64] B. S. Everitt, S. Landau, M. Leese, D. Stahl, Cluster Analysis, 5th Edition, Probability and Statistics, Wiley, 2011.

[65] J. A. Anderson, Introduction to Neural Networks, MIT Press, Cambridge, MA, 1995.

[66] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[67] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: data mining, inference and prediction, 2nd Edition, Springer Series in Statistics, Springer, 2009.

[68] J. S. Heyman, R. S. Rogowski, Fiber optic sensor technology–an opportunity for smart aerospace structures, in: AIAA/NASA/AFWAL Conference on Sensors and Measurements Techniques for Aeronautical Applications, AIAA-88-4665, Atlanta, Georgia, 1988.

[69] A. Ghoshal, W. Prosser, B. Copeland, A. Chattopadhyayi, Development of piezoelectric acoustic sensors and bio-inspired embedded sensor array architecture for structural health monitoring, in: 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, AIAA 2004-1958, Palm Springs, California, 2004.

[70] R. Everson, L. Sirovich, The Karhunen-Loeve procedure for gappy data, Journal of Optical Society of America 12 (8) (1995) 1657–1664.

[71] T. Bui-Thanh, M. Damodaran, K. Willcox, Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition, AIAA Journal 42 (8) (2004) 1505–1516.

[72] M. Wang, D. Dutta, K. Kim, J. C. Brigham, A computationally efficient approach for inverse material characterization combining Gappy POD with direct inversion, Computer Methods in Applied Mechanics and Engineering 286 (2015) 373–393.

[73] J. Han, M. Kamber, J. Pei, Data Mining, 3rd Edition, The Morgan Kaufmann Series in Data Management Systems, Elsevier, 2012.