Chapter 7
A Probabilistic Reasoner Based on Bayes Risk for Damage Detection in Structural Systems

David Najera-Flores, Justin Jacobs, D. Dane Quinn, Michael D. Todd, and Anthony Garland

Abstract Structural health monitoring (SHM) systems are used to inform operation of structural systems subject to loads and environments that may affect their integrity. SHM systems rely on continuous monitoring of the structure to determine its health state. These systems are often coupled with a model of the deployed structure to determine the consequences of changes in the system by forecasting the response to future states. These models, which may be thought of as digital twins, need to be updated to reflect the latest state of the structural system. This work makes use of an uncertainty-aware machine learning model that enforces distance preservation of the original input space to determine deviations from the training data input space distributions. This workflow enables domain shift detection to determine whether damage is present in the structure. The uncertainty metrics generated by this network are then used in a Bayes risk framework to design an optimal damage detector given cost and risk considerations. The approach is demonstrated on a computational example with simulated damage.

Keywords Machine learning · Nonlinear structural dynamics · Structural health monitoring · Damage detection · Data-driven · Structure-preserving

Introduction

A structural health monitoring (SHM) system should enable safe and reliable operation of the monitored system by providing actionable information to stakeholders so they can perform informed decisions about the structure's maintenance. This task involves the mapping from data to knowledge via an inference process. This inference process is also known as *reasoning*. More precisely, a probabilistic reasoner takes into account the various probability distributions from uncertain events to propose a course of action given the cost of decision consequences (risk). This task implicitly involves performing a diagnosis or detection of an underlying change in the system via indirect observable symptoms [1]. The work by the authors in [2] introduced a strategy for domain shift detection based on symptoms related to the dynamic response of a system. That work presented a method based on a distance-preserving, uncertainty-aware machine learning model (ML) that acted as a digital twin. The proposed model was able to ingest data from an external source (such as sensors) and compute the potential energy

David Najera-Flores
ATA Engineering, Inc., San Diego, CA 92128
e-mail: dnajera@ata-e.com

Justin Jacobs
Oak Ridge National Laboratory, Oak Ridge, TN 37830
e-mail: jacobsjw@ornl.gov

D. Dane Quinn
Department of Mechanical Engineering, The University of Akron, Akron, OH 44325
e-mail: quinn@uakron.edu

Michael D. Todd
Department of Structural Engineering, University of California San Diego, La Jolla, CA 92093
e-mail: mdtodd@ucsd.edu

Anthony Garland
Sandia National Laboratories, Albuquerque, NM 87123
e-mail: agarlan@sandia.gov

Roland Platz et al. (eds.), *Model Validation and Uncertainty Quantification, Vol. 3*, Conference Proceedings of the Society for Experimental Mechanics Series, https://doi.org/10.13052/97887-438-0148-1_7

and the damping of the underlying system, as well as the corresponding predictive variance. The work in [2] demonstrated how a Wald-type hypothesis test based on the order statistics of the distribution of the logarithm of the variance over time was useful for robust domain shift detection. The present work builds upon this past work to demonstrate how this workflow can be used to provide actionable information based on the desired sensitivity and specificity of the detector. The approach is demonstrated on a computational example with simulated damage. Section provides technical background, Section presents a computational example with simulated damage to demonstrate the proposed approach, and Section concludes and provides directions for future work.

Background

This section provides a technical description of the ML model used and the detection theory principles employed including the uncertainty-aware ML model and the Bayes risk approach.

Uncertainty-aware ML model

The model introduced in [2] is briefly summarized in this section. The neural network architectures used in this work consist of residual connections with spectral-normalized weights. This choice of architecture ensures distance-preservation as described in [3] by enforcing the lower bound of the Lipschitz constant through its residual connections, and the upper bound by using spectral normalization. This construction represents a bi-Lipschitz constraint that prevents disproportional contraction (also known as feature collapse) or expansion of the input space when it is transformed to the hidden space [3]. There are two separate networks and they both follow the exact same architecture. The input to the network is either $\mathbf{z}_\beta$ or $\dot{\mathbf{z}}_\beta$, which are N_β-dimensional vectors, where N_β is the number of degrees-of-freedom in the boundary of the isolated region. The input layer maps the input to a the hidden space of dimension d with a linear dense layer. The next L hidden layers follow a residual architecture where the input is added to the output after each dense layer. A swish activation function is used at each dense layer. The output layer uses a Gaussian process (GP) with kernel matrix, $\mathbf{K}_{\text{GP}}$. The hidden representation, $\mathbf{h}_i \in \mathbb{R}^{d_{L-1}\times 1}$, of the previous penultimate layer of dimension d_{L-1} serves as input to the GP layer.

As the output Gaussian process is intractable, a low-rank approximation of the kernel matrix is applied via

$$\mathbf{K}_{\text{GP}} = \mathbf{\Phi}\mathbf{\Phi}^T, \tag{1}$$

where

$$\mathbf{\Phi}_i = \sqrt{2\sigma^2/d_L}\cos(-\mathbf{W}_L\mathbf{h}_i + \mathbf{b}_L), \tag{2}$$

are random features that form a random sampling of a square-integrable basis [4] for the output layer. Here $\mathbf{W}_L \in \mathbb{R}^{d_L\times d_L}$ is a fixed weight matrix with entries sampled independent and identically distributed (i.i.d) from a standard normal distribution, $\mathbf{b}_L \in \mathbb{R}^{d_L\times 1}$ is a fixed bias vector with entries sampled i.i.d. from a uniform distribution in the range $(0, 2\pi)$, and σ^2 is the GP kernel variance and is set to 1.0.

The output of the SNGP is then obtained as

$$g(h_i) = \mathbf{\Phi}_i^T\mathbf{u} + c, \tag{3}$$

where $\mathbf{u} \in \mathbb{R}^{d_L\times 1}$ is the vector of learnable weights in the last layer and is initialized by randomly sampling from a zero-mean normal distribution with variance $\tau^2 = 3.0$ (tuned manually for numerical stability) and $c \in \mathbb{R}^{1\times 1}$ is the learnable constant bias. The output $g(h_i) \in \mathbb{R}^{1\times 1}$ is a scalar field corresponding to either the nonlinear potential $V_{\text{nl}}(\mathbf{z}_i)$ or nonlinear damping $\mathcal{I}_{\text{nl}}(\dot{\mathbf{z}}_i)$.

To obtain the prediction variance, we need to construct the model Hessian matrix, $\mathbf{H} \in \mathbb{R}^{d_L\times d_L}$, of the log posterior likelihood at the maximum a posteriori (MAP) estimate. This MAP estimate is induced as the Gaussian process posterior approximated in (1) and (2). Per [3], the model Hessian is defined as

$$\mathbf{H} = \sum_{i=1}^{N_{\text{train}}} \mathbf{\Phi}_i\mathbf{\Phi}_i^T + \mathbf{I}, \tag{4}$$

where N_{train} is the number of training examples. The Hessian matrix corresponds to the precision matrix, or the inverse of the covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{d_L\times d_L}$. Hence, the prediction variance of a new example can be computed as

$$\text{var}(\mathbf{z}_k) = \mathbf{\Phi}_k^T\mathbf{\Sigma}\mathbf{\Phi}_k \tag{5}$$

where $\mathbf{\Sigma}$ is fixed after training but $\mathbf{\Phi}_k$ is a function of the new input $\mathbf{z}_k$. All of the code was implemented using the Python packages Jax [5] and Flax [6].

Damage detector definition

The approach taken in this work is based on statistical detection theory [7]. However, unlike classical detection theory which deals with known signals, the present problem deals with unknown processes, e.g., cases in which the underlying statistical process is not completely known *a priori*. In this work, the probability density function and its corresponding parameters are unknown since they are being generated by a neural network. This statistical process corresponds to the predictive variance associated with the damping and potential energy of the structural system. These signals are used as observable symptoms in lieu of direct observation of damage. To this end, the problem statement in this work consists of the following simple binary hypothesis test:

$$\begin{aligned} H_0 &: \text{damage is not present} \\ H_1 &: \text{damage is present} \end{aligned} \tag{6}$$

Following the hypothesis test framework developed in [2], the two hypotheses rely on the statistical dispersion of the signal being considered as

$$\begin{aligned} H_0 &: \quad \text{IQR}_X - \text{IQR}_Z = 0 \\ H_1 &: \quad \text{IQR}_X - \text{IQR}_Z \neq 0, \end{aligned} \tag{7}$$

where IQR is defined as the interquartile range (IQR) of the random signals X and Z, which correspond to a vector of unordered samples. This vector is defined as the logarithm of the predictive neural network variance within a predefined time window. In this case, the signal X corresponds to the predictive variance associated with a baseline case, which is representative of the healthy state of the structure, while Z represents other time signals to be tested.

An empirical distribution of the random signals, X and Z, is obtained through a bootstrap of K samples drawn within a predefined time window. Consistent with [2], a Wald-type test is carried out explicitly by a Welch's t-test defined as

$$T = \frac{\overline{X} - \overline{Z}}{\sqrt{S_X^2 + S_Z^2}}. \tag{8}$$

The test statistic (8) then follows a t-distribution with degrees of freedom

$$\text{df} = \frac{\left(\frac{S_Z^2}{K} + \frac{S_X^2}{K}\right)^2}{\frac{1}{K-1}\left(\frac{S_Z^2}{K}\right)^2 + \frac{1}{K-1}\left(\frac{S_X^2}{K}\right)^2}. \tag{9}$$

The two-tailed p-value associated with the test statistic in (8) can then be computed as

$$\text{p} = 2(1 - \text{F}(T)) \tag{10}$$

where $\text{F}(T)$ represents the cumulative density function (CDF) of the test statistic. The hypothesis test can therefore be expressed as

$$\begin{aligned} \text{p} \geq \alpha &: \text{fail to reject } H_0 \\ \text{p} < \alpha &: \text{we reject } H_0 \end{aligned} \tag{11}$$

where α represents the significance level of the hypothesis test, also interpreted here as the probability of false alarm (P_{FA}). In other words, the significance level α defines the probability $P_{FA} = P(H_1|H_0)$, or the probability of deciding that the structure is damaged when it is not, also known as a type I error.

Furthermore, this binary hypothesis test also allows us to compute the associated power of the test (denoted as β) which defines the probability that the test correctly rejects H_0 when H_1 is true. This correct inference is known as the probability of detection, $P_D = P(H_1|H_1)$. For most detectors, there exists a trade-off between the probability of false alarms and the probability of detection. This trade-off is the same as the trade-off between the specificity and sensitivity of the test (or the probability of false alarms vs. misses). Therefore, designing a detector is a task that must consider this trade-off in a systematic way. One approach to accomplish this is to rely on the concept of Bayes risk [7], $\mathcal{R}(\theta)$, which combines the weighted likelihood of each possible decision as

$$\mathcal{R}(\theta) = E(C, \theta) = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij} P(H_i|H_j, \theta) P(H_j) \tag{12}$$

where the coefficients C_{ij} represent the cost associated with deciding H_i when H_j is true, θ are the fixed choice parameters associated with the detector or decision function, the likelihood terms $P(H_i|H_j, \theta)$ can be computed from the hypothesis test formulation as described above, and the prior terms $P(H_j)$ can be defined based on prior knowledge (for example, from historical data). The Bayes risk can be interpreted as the smallest integrated risk over all decision functions [8]. These decision functions are parameterized by θ which enables minimization of the risk function by considering all the possible θ.

Analysis

This section will demonstrate the approach described previously by applying the proposed workflow to a structural rod with isolated deviatoric behavior representative of damage.

Problem description

As described previously, a Wald-type hypothesis test was used to detect damage in the rod structure described in [9] and shown in Figure 1. This example is described in detail in the original paper but a general description is provided here. The rod problem was defined using the isolated formulation introduced in [10, 11] where a structure was decomposed into two regions $\mathcal{C}_1$ and $\mathcal{C}_2$, where the equations of motion within $\mathcal{C}_1$ are assumed to be linear and known while region $\mathcal{C}_2$ contains localized nonlinearities such as joints or damaged regions. Based on this formulation and following the work in [10], the equation of motion can be decomposed as

$$\begin{aligned} \mathbf{M}\,\ddot{\mathbf{w}} + \mathbf{C}\,\dot{\mathbf{w}} + \mathbf{K}\,\mathbf{w} &= \mathbf{f}(t) - \mathbf{Q}_\star, \\ \mathbf{M}_2\,\ddot{\mathbf{z}} + \mathbf{C}_2\,\dot{\mathbf{z}} + \mathbf{K}_2\,\mathbf{z} + \mathbf{N}(\mathbf{y}+\mathbf{z}) &= \mathbf{0}, \\ \mathbf{Q}_\star &= \begin{bmatrix} \mathbf{0}_\mathrm{c} & \mathbf{Q} & \mathbf{0}_2 \end{bmatrix}^T, \\ \mathbf{Q} &\equiv \mathbf{M}_{\alpha\beta}\,\ddot{\mathbf{z}}_\beta + \mathbf{C}_{\alpha\beta}\,\dot{\mathbf{z}}_\beta + \mathbf{K}_{\alpha\beta}\,\mathbf{z}_\beta. \end{aligned} \tag{13}$$

where $\mathbf{N} \in \mathbb{R}^{N_2}$ represents the vector for isolated nonlinear forces, $\mathbf{M}, \mathbf{K}, \mathbf{C}$ represent the known mass, stiffness, and damping matrices, respectively while $\mathbf{w} \in \mathbb{R}^{(N_1+N_2)}$ is defined as a mixed displacement vector

$$\mathbf{w} \equiv \begin{bmatrix} \mathbf{x}_1 & \mathbf{y} \end{bmatrix}^T = \begin{bmatrix} \mathbf{x}_\mathrm{c} & \mathbf{x}_\alpha & \mathbf{y}_\beta & \mathbf{y}_\mathrm{n} \end{bmatrix}^T, \tag{14}$$

combining $\mathbf{x}_1 \in \mathbb{R}^{N_1}$, the *exact* response of the system in $\mathcal{C}_1$, with $\mathbf{y} \in \mathbb{R}^{N_2}$, the response of the ideal system, that is, the system in the absence of the nonlinearities, but subject to the force $\mathbf{Q}$ applied at the boundary of the two regions, and referred to as the deviatoric force. This deviatoric force is typically unknown because the exact nature of the nonlinearity in the isolated region is unknown and therefore needs to be learned. Finally, within $\mathcal{C}_2$ the deviatoric response is defined as $\mathbf{z} \equiv \mathbf{x}_2 - \mathbf{y} \in \mathbb{R}^{N_2}$. These coordinates are further decomposed, so that the subscript c represents the DOFs in $\mathcal{C}_1$ that are not coupled to the interface DOFs, while α represents the DOFs in $\mathcal{C}_1$ that are coupled to the interface. Likewise β represents the DOFs in $\mathcal{C}_2$ that are coupled to the interface (i.e., the only DOFs in $\mathcal{C}_2$ that are observable), and n represents the DOFs in $\mathcal{C}_2$ that are not coupled to the interface (i.e., not observable).

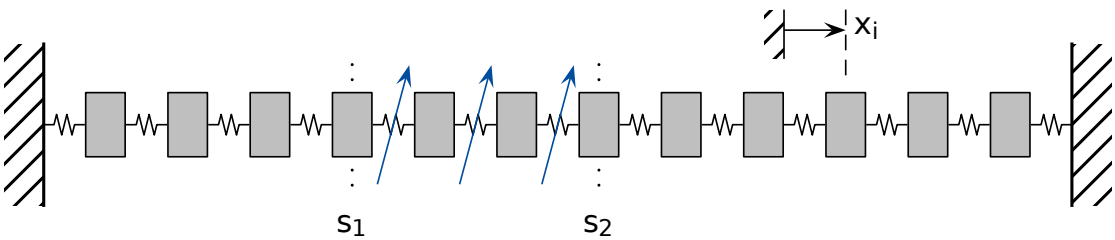

Fig. 1 Exemplar rod discretized into a series of masses, springs, and dampers with an isolated nonlinear region.

For the rod structure, proportional damping was defined as $\mathbf{C} \equiv \xi\,\mathbf{K}$. The nonlinear region $\mathcal{C}_2$ is located in the interval $(s_1, s_2) = (0.25, 0.35)$, so that the isolated region exists between elements $\ell \equiv 16$ and $r \equiv 22$. The deviatoric force across the isolated region is identified as

$$\delta Q \equiv Q_{23} - Q_{15}, \tag{15}$$

The initial conditions of the system are specified in terms of the modal displacements and velocities $(\mathbf{q}(0), \dot{\mathbf{q}}(0))$, where

$$\mathbf{x}(0) = \Phi\, \mathbf{q}(0), \quad \dot{\mathbf{x}}(0) = \Phi \dot{\mathbf{q}}(0), \tag{16}$$

where Φ is the modal transformation matrix.

The nonlinearities in the isolated region are assumed to contain linear mistuning and cubic nonlinearities in terms of both stiffness and damping. Finally, hysteretic nonlinear damping is also present between the elements within $\mathcal{C}_2$, so that the nonlinear force is represented as

$$\begin{aligned} f_{\text{nl},i} = \kappa_i \left(x_{i+1} - x_i\right) + \chi_i \left(\dot{x}_{i+1} - \dot{x}_i\right) \\ +\eta_i \left(x_{i+1} - x_i\right)^3 \\ +\zeta_i \left(\dot{x}_{i+1} - \dot{x}_i\right)^3 + g_{\text{hys},i}, \end{aligned} \tag{17}$$

where the hysteretic force is defined as

$$\begin{aligned} g_{\text{hys},i} &= k_{\text{h}} \left(\left(x_{i+1} - x_i\right) - r_i\right), \\ \dot{r}_i &= \epsilon_i \tan\left(\frac{k_{\text{h}}}{\mu}\left(\left(x_{i+1} x_i\right) - r_i\right)\right), \end{aligned} \tag{18}$$

following the regularized formulation presented in [9].

Unless noted otherwise the parameters of the nonlinear region are chosen as

$$\begin{aligned} \kappa_i = 0, \quad \chi_i = 0, \quad \eta_i = \rho_\eta \cdot 2.621 \times 10^5, \\ \zeta_i = 2.621 \times 10^3, \quad k_{\text{h},i} = \rho_{k_h} \cdot 6.50 \times 10^1, \\ \mu_i = 1.00 \times 10^{-1}, \quad \epsilon_i = 1.00 \times 10^{-2}, \end{aligned} \tag{19}$$

with $\xi = 10^{-4}$, and ρ_η and ρ_{k_h} represent the level of cubic nonlinearity and hysteresis, respectively. For the data generated for training the network, $\rho_\eta = 1.0$, and $\rho_{k_h} = 0.0$. Damage was simulated by adjusting these two coefficients, as summarized in Table 1.

Table 1 Domain shift cases. For the gradual stiffness increase, the max value is provided. The initial condition (IC) coefficients represent the contribution of each mode shape to the initial deformation.

Case name	ρ_η	ρ_{k_h}	IC
Baseline 0 (train)	1	0	(1.0,0.5,2.0)
Baseline 1 (test)	1	0	(0.5,1.0,0.5)
Baseline 2 (test)	1	0	(1.0,1.0,1.0)
Hysteresis Lvl. 1	1	1	(1.0,0.5,2.0)
Hysteresis Lvl. 2	1	10	(1.0,0.5,2.0)
Hysteresis Lvl. 3	1	20	(1.0,0.5,2.0)
Hysteresis Lvl. 4	1	50	(1.0,0.5,2.0)
Hysteresis Lvl. 5	1	100	(1.0,0.5,2.0)
Cubic Stiff. Lvl. 1	2	0	(1.0,0.5,2.0)
Cubic Stiff. Lvl. 2	5	0	(1.0,0.5,2.0)
Cubic Stiff. Lvl. 3	20	0	(1.0,0.5,2.0)
Cubic Stiff. Lvl. 4	100	0	(1.0,0.5,2.0)
Grad. Cubic Stiff. Lvl. 1	5	0	(1.0,0.5,2.0)
Grad. Cubic Stiff. Lvl. 2	20	0	(1.0,0.5,2.0)
Grad. Cubic Stiff. Lvl. 3	100	0	(1.0,0.5,2.0)

In addition, 5% synthetic measurement noise was added to the input data. Previous work [12], demonstrated that these trained neural networks are robust to measurement noise even when added at every iteration during time integration. This noise was added with the intent of approximating a more realistic sensor measurement. The additive noise was modeled as

$$\mathbf{x}_\beta(t) = \overline{\mathbf{x}}_\beta(t) + \delta, \tag{20}$$

where $\overline{\mathbf{x}}_\beta(t)$ is the mean process and δ is the random noise which is assumed to be i.i.d modeled as

$$\delta = \nu \cdot \sqrt{\frac{\Sigma_i^N \overline{x}_\beta(t_i)^2}{N-1}} \cdot \text{Normal}(0,1), \tag{21}$$

where $\nu = 0.05$ and it represents the noise factor which can also be interpreted as the reciprocal of the signal-to-noise ratio (SNR) since it represents the ratio between the noise variance and the signal variance.

The deviatoric response inherits the same additive noise as

$$\mathbf{z}_\beta(t) = \overline{\mathbf{z}}_\beta(t) + \delta, \tag{22}$$

and similarly for the deviatoric velocities, $\dot{\mathbf{z}}_\beta(t)$. These are then used to compute the deviatoric forces as

$$\mathbf{Q}_\star = \frac{\partial V_{\text{nl}}(\mathbf{z}_\beta)}{\partial \mathbf{z}_\beta} + \frac{\partial \mathcal{I}_{\text{nl}}(\dot{\mathbf{z}}_\beta)}{\partial \dot{\mathbf{z}}_\beta}, \tag{23}$$

where $\mathbf{z}_\beta$ and $\dot{\mathbf{z}}_\beta$ are the deviatoric displacements and velocities at the measured boundary (as described in more detail in [9]). The $V_{\text{nl}}(\cdot)$ and $\mathcal{I}_{\text{nl}}(\cdot)$ terms represent the nonlinear potential and damping functions, which are modeled with a structure-preserving, spectral-normalized neural Gaussian process (sp-SNGP) as described in [2]. Consistent with previous work, this work focuses on the study of the damping term variance predicted by sp-SNGP for domain shift detection as an indicator for damage.

Optimizing for probability of detection

Designing an optimal detector depends on the specific objectives of the SHM system. One reasonable objective is to maximize the probability of detection or, in other words, design a highly sensitive detector (i.e., a high recall rate). We can use the receiver operating characteristic (ROC) curve to study this objective. Moreover, the ROC curve also shows the trade-off between recall and precision. In other words, a highly-sensitive test will also result in a larger false alarm rate. To study this trade-off, we consider the case where progressive damage is introduced via a gradual change to the cubic stiffness term. Referring to the nonlinearity described in Eq. 17, the cubic stiffness term is modified as

$$f_{\text{nl},i} = \eta_i(x_{i+1} - x_i)^3 + \eta_{\text{d},i}(t - t_{\text{d}})^2(x_{i+1} - x_i)^3, \tag{24}$$

where

$$\begin{cases} \eta_{\text{d}} = 0.6944, & \text{if } t \geq t_{\text{d}} = 4 \\ 0 & \text{otherwise,} \end{cases} \tag{25}$$

where η is the cubic stiffness coefficient, η_{d} is the damage coefficient, and t_{d} is the time at the onset of damage. As shown, the damage growth is modeled as a quadratic function to represent a self-accelerating degradation process (i.e., the second derivative is non-zero).

Given this progressive damage model, we compute the ROC curve as a function of window size and time. We accomplish this by defining a sliding window across the time history by having consecutive windows of size WS overlap 25 time steps over the range $t = 3$ (3000 time steps) to $t = 4$ (4000 time steps). The following window sizes are considered: $WS = [250, 1000, 3000, 5000]$. Figure X shows the ROC curves for the four WS values considered, with the brightness of each line representing time evolution (darker lines representing later time). As shown, as the lines get darker, the ROC curves show an ideal detector (i.e., the area under the curve (AUC) approaches 1). This result exemplifies how damage is easier to detect as time progresses which is consistent with the self-accelerating damage evolution model. As a result, one may choose the detector that maximizes the probability of detection, which corresponds to the detector that observes data in the most damaged state. However, this behavior may be unacceptable for a realistic SHM system where decisions need to be taken in a promptly manner. To this end, Section will consider the cost associated with each decision as well as the consequence of a delay in decision making.

Given this progressive damage model, we compute the ROC curve as a function of window size and time. We accomplish this by defining a window with a starting point fixed at $t = 4$ and variable window size. Figure 2 shows the ROC curves for 100 WS values corresponding to WS $= \{100, 129, 158...3000\}$, with the color of each line representing time evolution (from blue to red). As shown, as the lines get more red, the ROC curves approach an ideal detector (i.e., the area under the curve (AUC) approaches 1 and the p-value approaches 0). This result exemplifies how damage is easier to detect as time progresses which is consistent with the self-accelerating damage evolution model. This effect is also evident in the IQR

histograms shown in Figure 3 where it can be seen that the distributions are hard to distinguish when small window sizes are used while they become easier to distinguish as more data are collected (also consistent with the small p-values in Figure 2). As a result, one may choose the detector that maximizes the probability of detection, which corresponds to the detector that observes data in the most damaged state. However, this behavior may be unacceptable for a realistic SHM system where decisions need to be taken in a time-critical manner. To this end, Section will consider the cost associated with each decision as well as the consequence of a delayed decision.

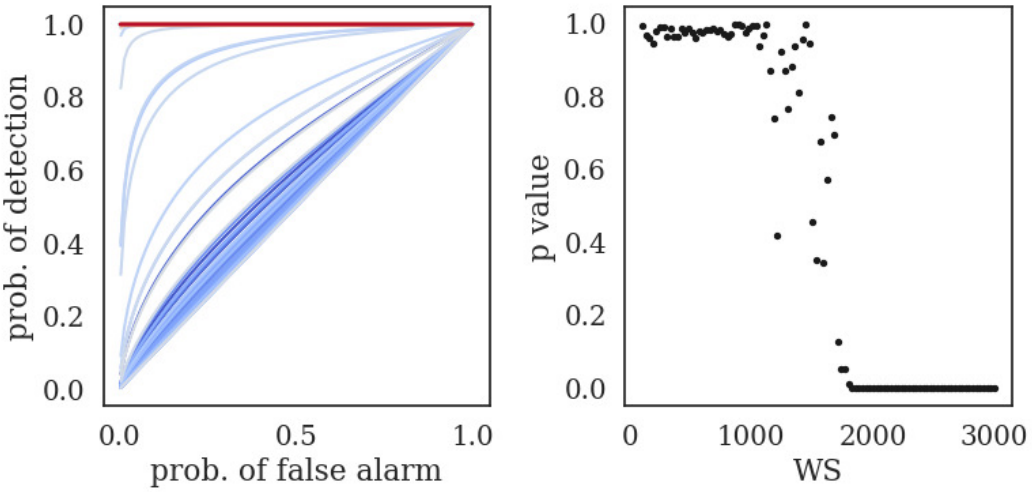

Fig. 2 Left: ROC curves colored by WS value (blue to red: short to large). Right: P-value as a function of WS.

Considering cost of wrong decisions

As described in Section, a detector that only considers probability of detection tends to favor features corresponding to an extended monitoring period. However, waiting for data may have consequences (e.g., risk of structural failure) and/or additional monitoring cost associated with data collection. For this reason, an additional term to model this penalty is added to the Bayes Risk expression as

$$\mathcal{R}(\theta) = E(C,\theta) = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij}P(H_i|H_j,\theta)P(H_j) + \mu\theta^2, \tag{26}$$

where $\mu = 1.0 \times 10^{-7}$ is a normalizing coefficient. A quadratic term was chosen for the penalty term after the quadratic penalty that comes out of assuming a Gaussian prior for θ (see ridge regression in [13]). Furthermore, normalizing the expression by the cost of a correct detection (C_{11}), and setting the cost of normal operations to $C_{00} = 0$, substituting in the probability definitions, and setting $\theta = \text{WS}$, we obtain

$$E(C,\text{WS}) = \lambda(1-P_D)P_{damage} + \gamma P_{FA}(1-P_{damage}) + P_D P_{damage} + \mu\text{WS}^2, \tag{27}$$

where $\gamma = C_{10}/C_{11}$ and $\lambda = C_{01}/C_{11}$.

With this Bayes Risk formulation as our cost model, we can explore the effect of the cost parameters, WS, our threshold for false alarms and misses, and the prior probability. Figure 4 illustrates the effect of the prior probability of damage, the coefficients λ and γ, and WS. The cost has been normalized to be 1.0 at $WS = 0$ for these plots for ease of visualization. The top row of plots considers the variation of γ which represents the ratio of the cost of a false alarm with respect to the cost of a correct detection. So γ essentially models the cost of unnecessary downtime due to a false alarm. For these plots, λ is fixed at 2. As shown in the top row of plots, the cost functions are relatively insensitive to γ given a fixed prior probability of damage. Most of the variation is coming from varying P_{damage}. When $P_{damage} = 0.1$, which means that the prior probability of damage is low (e.g., based on historical trends), the best detector is the one with the smallest WS because, as described above, there is a quadratic cost associated with WS. However, as P_{damage} increases, the minimum cost moves to the right of the x-axis, indicating that more samples are preferable, even if an additional cost is incurred, because the probability of finding damage is higher. The second row of plots shows the sensitivity to λ which represents the cost of

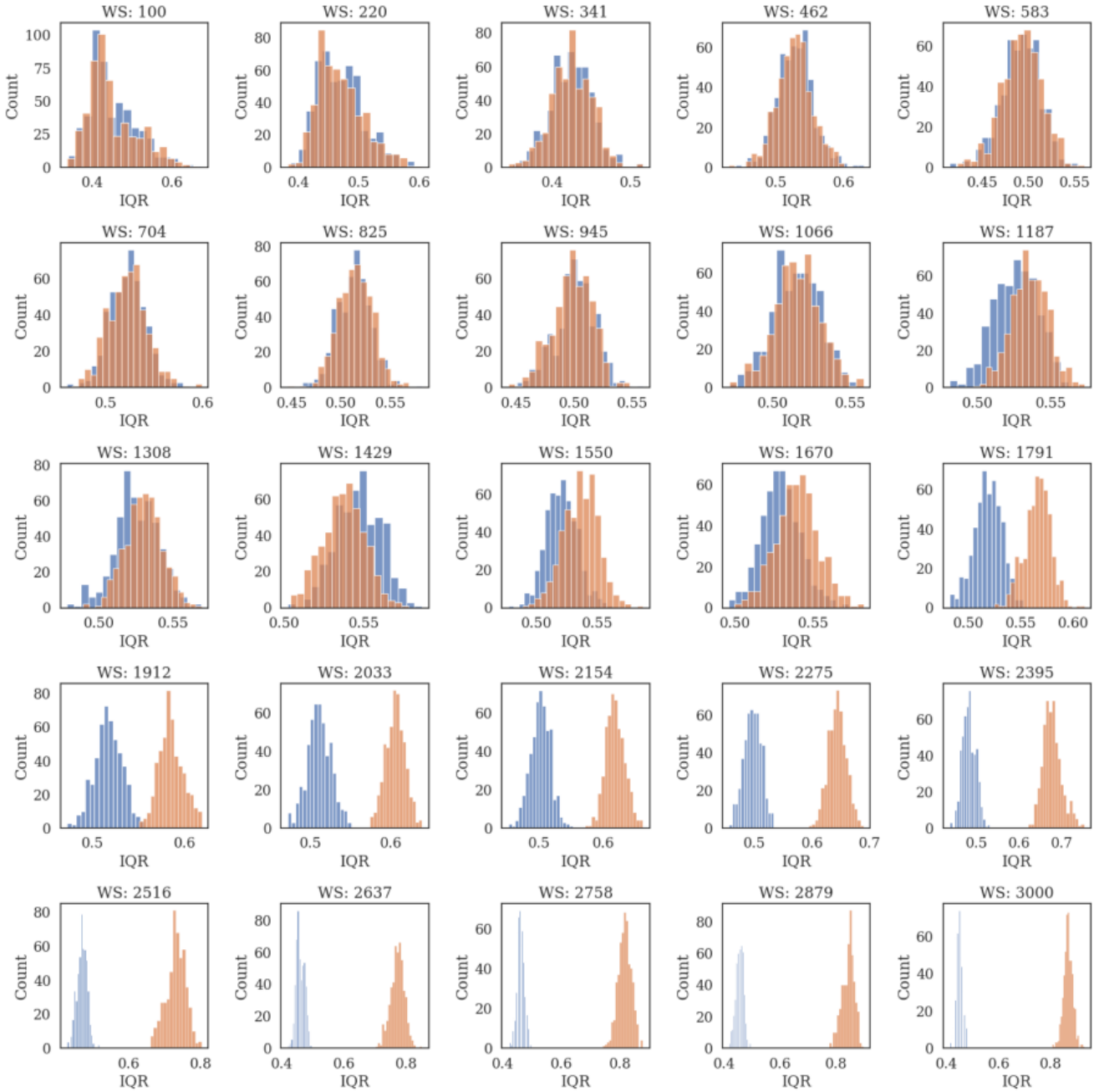

Fig. 3 IQR histograms for varying window sizes.

failure. A large λ is associated with significant consequences of failure, such as it would be the case in a primary structural component. All the plots in the second row favor a non-zero WS except when λ is smallest. There are two valleys in all of the plots and the absolute minimum varies depending on the value of λ. When $\lambda = 0.5$, in other words the consequence of missing to detect damage is half of the cost of correctly detecting it, the optimal detector is that with the smallest WS. This scenario is clearly a very unlikely one since the cost of an error is typically higher than the cost of detecting failure before it happens. As λ increases, the minimum cost moves from WS $= 0$ to the first valley and then to the second valley as P_{damage} also increases. This result is reasonable as the cost associated with the consequences of undetected failure overcome the cost of data collection to a certain extent. As we know from Section , there are diminishing returns beyond a certain WS value (i.e., the p-value plateaus in Figure 2). This result stands in contrast to the conclusion reached from Figure 2 where cost/risk was not considered and the best detector corresponded to the one that collected the most data. To understand why

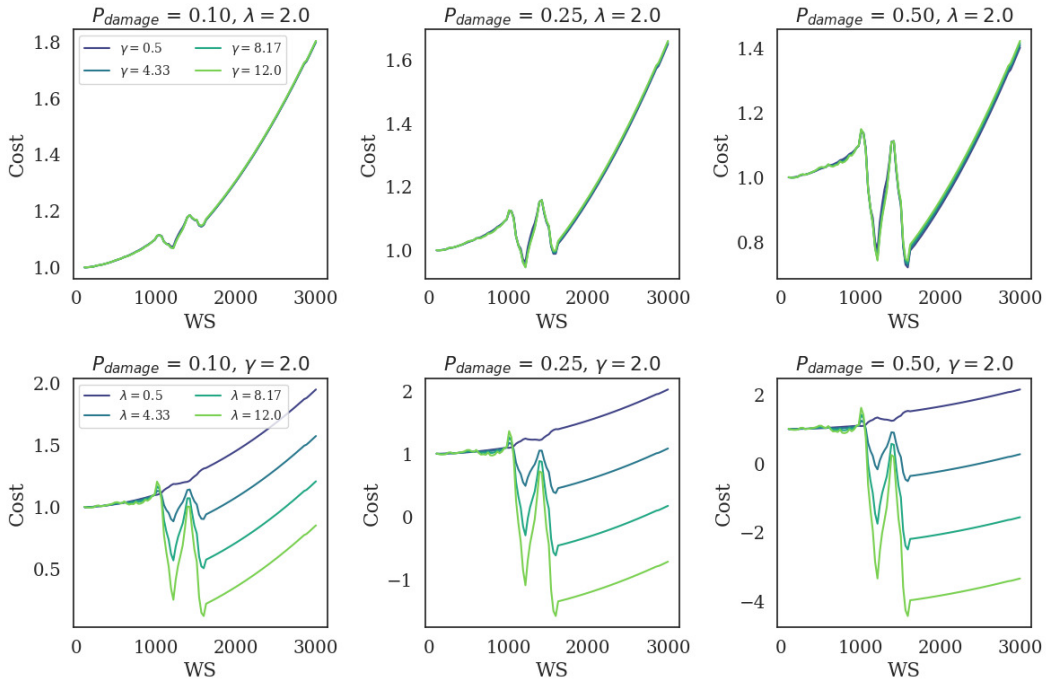

Fig. 4 Cost vs. WS as a function of prior probability of damage and λ.

there is variability between $WS = 1000$ and $WS = 2000$, we can look at the trends of the IQR distributions in Figure 5 (as a reminder, these were obtained through bootstrapping as described in Section). As shown, the IQRs only start to deviate at around the 1000 WS mark and they exhibit an oscillatory pattern because the variances are themselves oscillatory (due to their dependence on the state space response). The first minimum in the cost function corresponds to the first evident deviation of the IQRs around $WS = 1200$. After that, the IQRs tend to be hard to distinguish up until the $WS = 1600$ mark where we find the second minimum of our cost. It is expected that this transition region will vary based on the type of damage and the specific trajectories being considered so a robust detector should consider the second minimum after the oscillatory behavior has settled.

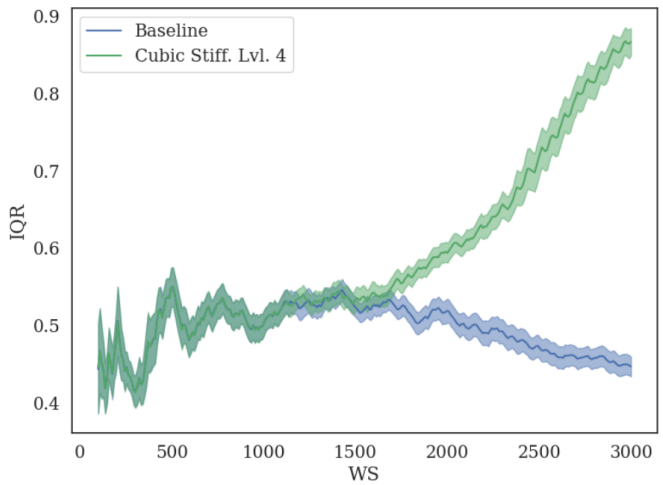

Fig. 5 IQR mean and one standard deviation band around it for the baseline (blue) and the damaged case (green).

The same data are plotted in Figure 6 for a fixed WS $= 1330$ to study the effect of γ, λ, and the probability of false alarm. The value of γ increases as we move from top to bottom, and the value of λ increases as we go from left to right.

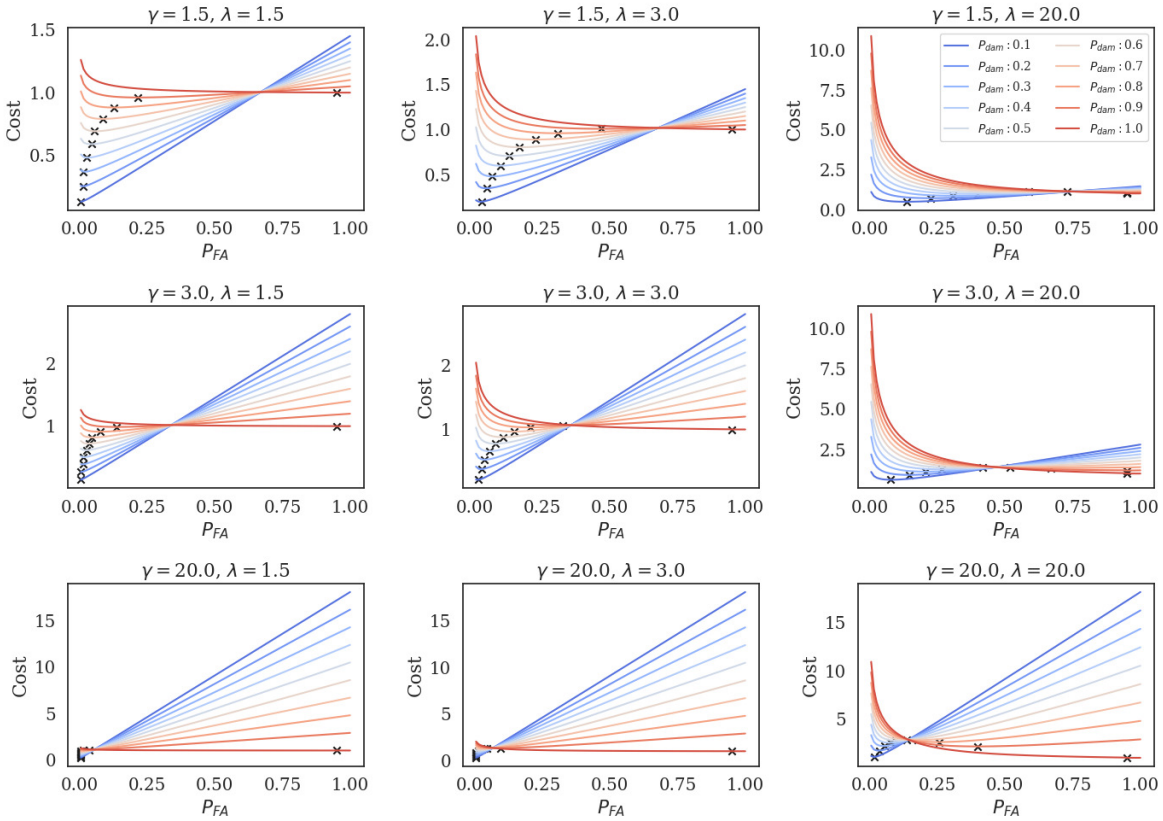

Fig. 6 Cost vs. probability of false alarm as a function of prior probability of damage, λ, and γ.

In all cases, $\gamma > 1.0, \lambda > 1.0$ to study the case where the cost of an error is higher than the cost of a correct detection. In all of these plots the color of the curves represents the prior probability of damage and the minimum cost of each curve is marked with $\times$. If we start by looking at the top plot, we can see that the minimum cost changes as we vary P_{damage}. In other words, the admissible rate of false alarms grows as P_{damage} increases. The minimums are clustered around low false alarm rates when λ is smaller while they quickly move to larger false alarm rates when we increase λ. This trend is due to the cost of an undetected failure growing. In other words, knowing that damage is likely, it is preferable to have a large rate of false alarms since undetected damage is very costly. As we move from top to bottom, the cost of unnecessary downtime increases. One extreme case is the lower left corner where $\gamma = 20.0, \lambda = 1.5$. This case represents an extreme where the cost of unnecessary downtime far exceeds the cost of undetected failure (also knowing that the prior probability of damage is low). In this case, having a very low rate of false alarms is preferable. However, as the probability of damage increases, the plots on the right hand column show that even when the cost of making mistakes is high, a high rate of false alarms is still preferable. When there is a large cost associated with unnecessary downtime, the cost is proportional to the probability of false alarm, except for the case where $P_{damage} = 1$ which is when damage is certain.

Detector performance when considering multiple damage types

Sections and only considered the progressive damage case (gradual cubic stiffness change) as an exemplar. In this section we consider all of the damage cases described in Table 1. We still apply the same Wald-type hypothesis test previously described. However, the power of the test cannot be easily computed across multiple damage cases so we evaluate the detector performance by fixing $P_{FA} = 0.05$ and generating Monte Carlo samples and then counting the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), where TP is equivalent to P_D. The results across all cases are presented in the confusion matrices shown in Figure 7 as a function of WS. This information is further summarized by computing the F1 score as a function of WS, as shown in Figure 8. As illustrated, the F1 score improves with increased WS, leading to the same conclusion drawn in Section that the best detector is the one that collects the most data. However, as shown in Section , this result changes when the cost/risk information is considered.

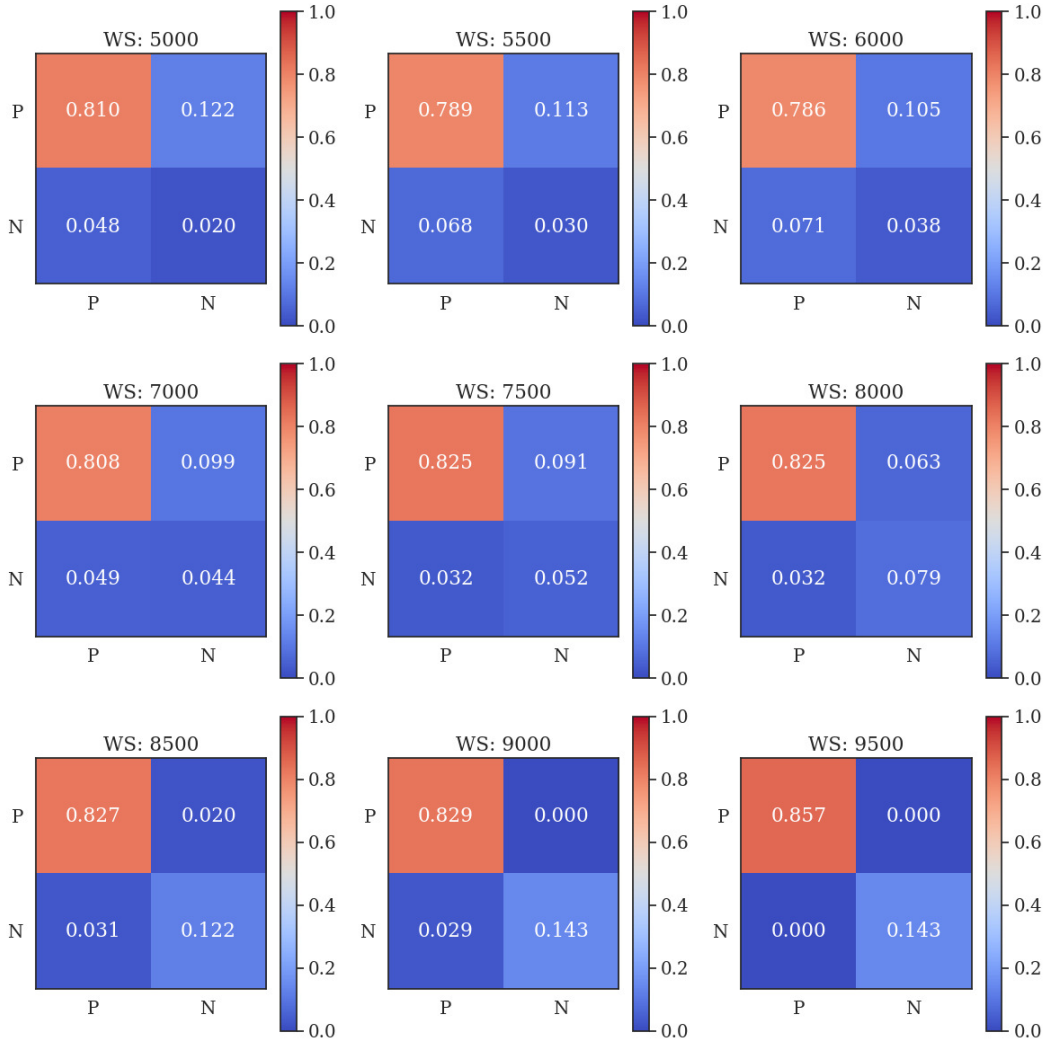

Fig. 7 Confusion matrix across all damage cases as a function of window size. Predicted condition shown across horizontal axis and actual conditions shown across vertical axis.

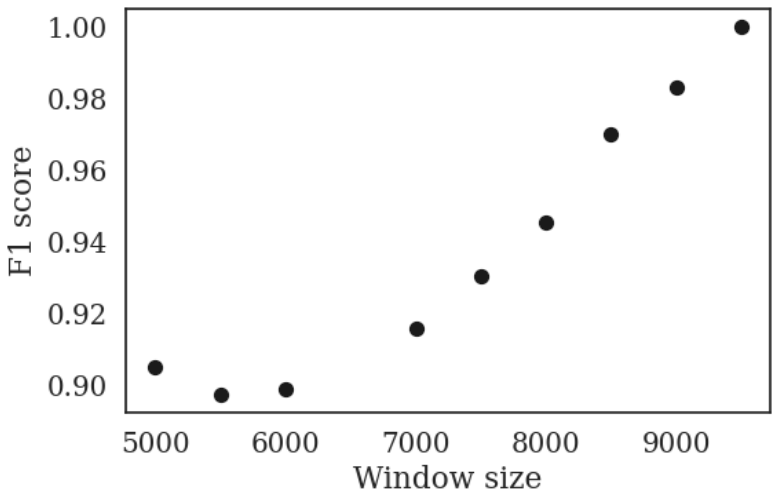

Fig. 8 F1 score as a function of window size.

Model updating

The previous section demonstrated an approach for designing a detector based on desired properties defined by cost or risk of making certain types of mistakes. Once damage or domain shift has been detected, a logical next step is to update the model to reflect the current state of the monitored structure. To demonstrate the strategy for model updating, we will focus on the example with a large change to the cubic stiffness coefficient (Cubic Stiff Lvl. 4 in [2]). This example was selected because it induced the largest error out of all the cases considered. Once damage has been detected using the approach previously presented, the existing sp-SNGP model is re-trained for a few iterations using the new data being observed. For this training, a window size of 3500 time steps from the new data being measured at the boundary of the isolated region was used to update the model. The model updates are performed by the classical backpropagation technique which involves stochastic gradient descent (SGD). The authors in [14] provide an interpretation of SGD with momentum as a posterior sampler in a Bayesian inference context. In this way, the model updating strategy employed in this work can be interpreted as a form of Bayesian inference with adaptive parameters via the ADAM optimizer.

The backpropagation algorithm was run for $10,000$ epochs (it took 15 minutes on a single CPU) to update the network weights. There are two separate networks that model the nonlinear potential energy (V_{nl}) and the nonlinear damping ($\mathcal{I}_{nl}$) respectively. Figure 9 shows a close-up of the nonlinear potential energy and damping time histories and their respective predictive intervals obtained by applying different initial conditions than those used for training. As shown, the model's uncertainty is significantly reduced after the update. Figure 10 illustrates the distributions of log-variances over $10,000$ time steps. As shown, the log-variance is reduced several orders of magnitude in both cases. This result is consistent with the previous findings when running sp-SNGP which are that reduced uncertainty is associated with in-distribution data. This update has effectively shifted the domain of the ML model so that the updated data is now in-distribution.

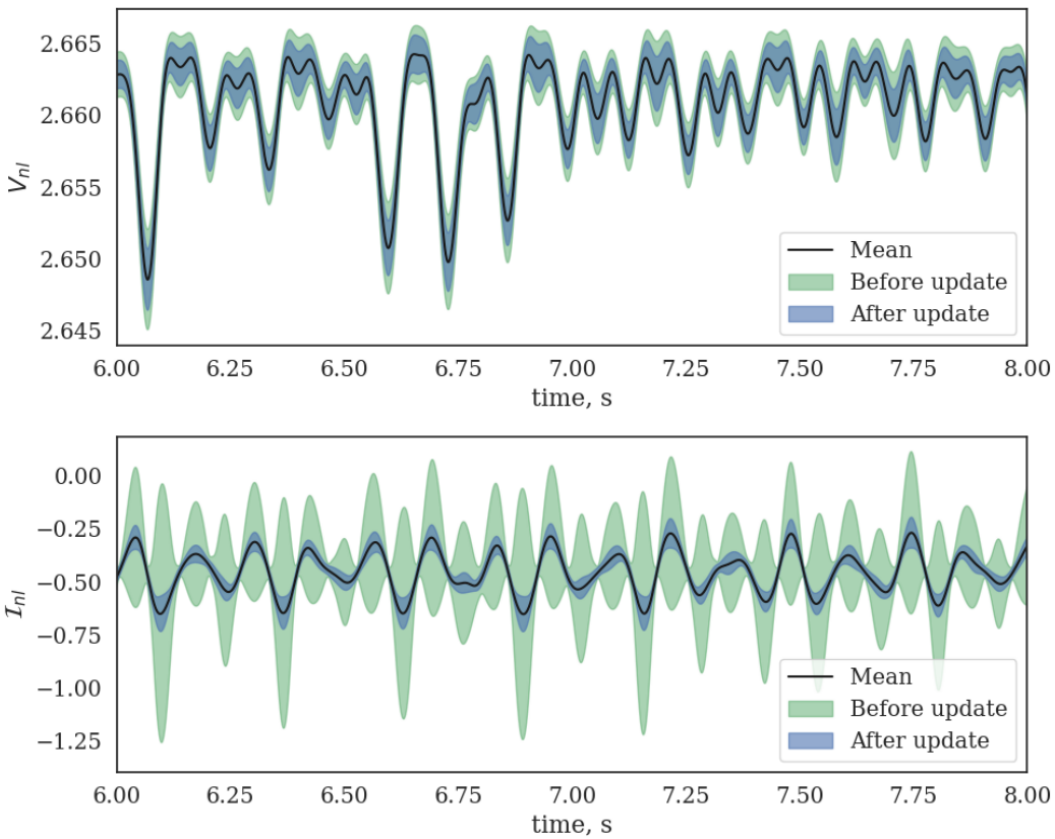

Fig. 9 Nonlinear potential (top) and damping (bottom) time histories. One standard deviation intervals shown.

As mentioned, there are two separate networks that need to be updated. In the absence of prior knowledge about the type of nonlinearity induced by the damage, both networks would need to be updated. The distribution of non-zero log-transformed weights for both networks are presented in Figure 11 before and after the model update. As shown, the distributions before and after are fairly close to each other. To evaluate whether or not they are the same distribution, we employed a Kolmogorov-Smirnov (KS) two-sample test. The p-value for the nonlinear potential weight distributions was 0.030 while for

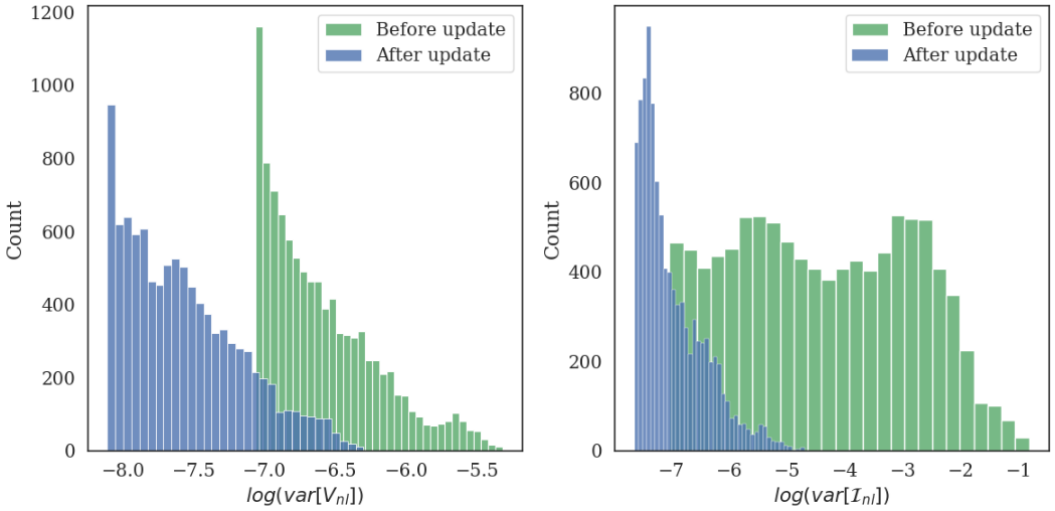

Fig. 10 Histograms of log-variance of nonlinear potential (left) and damping (right) network weights.

the nonlinear damping weight distributions it was 0.626. Considering a threshold of $p = 0.05$, we reject the null hypothesis of the KS test that the nonlinear potential weights were drawn from the same distribution, while we fail to reject the null hypothesis for the nonlinear damping, leading to the conclusion that the two samples were drawn from different distributions. In other words, the difference between the nonlinear potential weight distributions is statistically significant while it is not for the nonlinear damping. This result suggests that the domain shift observed in the Cubic Stiff. Lvl. 4 was caused by a change to the nonlinear potential, which is consistent with the physical change (an update to the cubic stiffness term), which only contributes to the nonlinear potential function. This result can be used to infer the type of change to the system from the data in the absence of any knowledge informing the physical mechanism that is causing the damage symptoms. Future work could extend this work to fully parametric formulation that enables parameter identification.

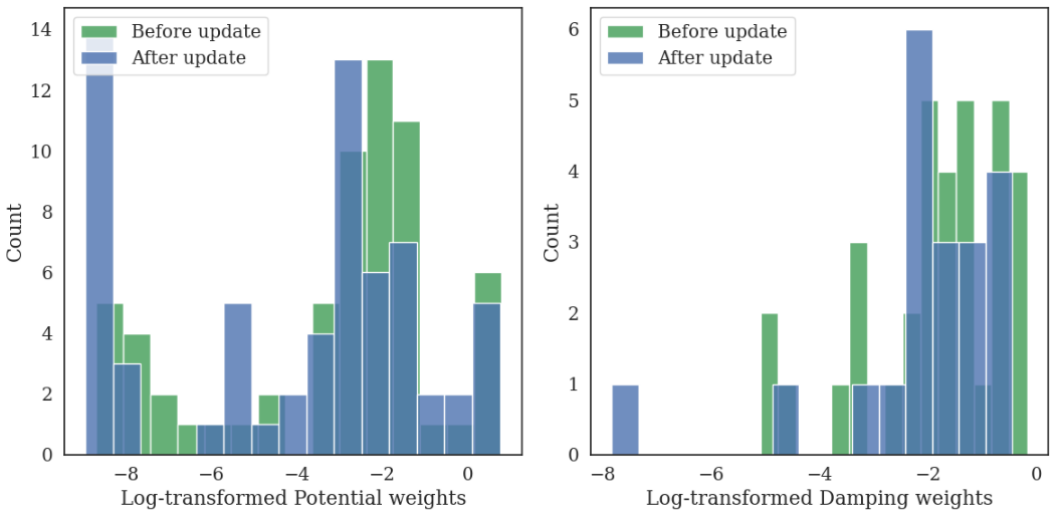

Fig. 11 Histograms of log-variance of nonlinear potential (left) and damping (right).

Conclusion

This work presented a demonstration of a strategy for cost/risk-informed decision-making based on features extracted from an uncertainty-aware ML model. The ML model predictive variance was used as the basis for defining a feature to detect

damage. Based on the work presented in [2] and the discussion in the present paper, it was shown that the order statistics of the log-transformed variance of the damping time histories could be used to detect domain shift which in turn was used as an indicator of damage. While previous work had mostly focused on maximizing probability of detection, the present work demonstrated a more practical case where cost and risk were considered via a Bayes risk formulation, establishing a workflow for probabilistic reasoning. As shown, the optimal number of samples to be collected depends on the cost associated with each decision taken. While this work focused on a single example with multiple damage cases, the workflow developed is extensible and applicable to other structures, and other cost paradigms. Moreover, this work demonstrated how the model can be updated to new data, and how that updates reduce the predictive uncertainty significantly. Furthermore, the statistics of the updated weights were used to infer an understanding of the physical mechanism driving the change. Future work will focus on extensions to understand the physics of the damage mechanism with more detail and extensions to the cost function considered.

Acknowledgments This work was funded by Sandia National Laboratories, Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

References

1. van der Gaag, L.C. and Renooij, S. "On the sensitivity of probabilistic networks to reliability characteristics". In Bouchon-Meunier, B., Coletti, G., and Yager, R.R., editors, *Modern Information Processing*, pages 395–405. Elsevier Science, Amsterdam (2006)
2. Najera-Flores, D.A., Jacobs, J., Dane Quinn, D., Garland, A., and Todd, M.D."Uncertainty-Aware, Structure-Preserving Machine Learning Approach for Domain Shift Detection From Nonlinear Dynamic Responses of Structural Systems".*ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, 11(1):011104 (2024)
3. Liu, J.Z., Lin, Z., Padhy, S., Tran, D., Bedrax-Weiss, T., and Lakshminarayanan, B."Simple and principled uncertainty estimation with deterministic deep learning via distance awareness" (2020)
4. Rahimi, A. and Recht, B."Random features for large-scale kernel machines". *Advances in Neural Information Processing Systems 20*, pages 1177–1184 (2008)
5. Bradbury, J., Frostig, R., Hawkins, P., Johnson, M.J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q."JAX: composable transformations of Python+NumPy programs" (2018)
6. Heek, J., Levskaya, A., Oliver, A., Ritter, M., Rondepierre, B., Steiner, A., and van Zee, M."Flax: A neural network library and ecosystem for JAX" (2023)
7. Kay, S.M.*Fundamentals of statistical signal processing, volume II*.Prentice Hall, Philadelphia, PA (1998)
8. Robert, C.*The Bayesian choice*.Springer Texts in Statistics. Springer, New York, NY, 2 edition (2007)
9. Najera-Flores, D.A., Quinn, D.D., Garland, A., Vlachas, K., Chatzi, E., and Todd, M.D."A structure-preserving machine learning framework for accurate prediction of structural dynamics for systems with isolated nonlinearities".*Mechanical Systems and Signal Processing*, 213:111340 (2024)
10. Quinn, D.D. and Brink, A.R."Global system reduction order modeling for localized feature inclusion".*Journal of Vibration and Acoustics, Transactions of the ASME*, 143:041006 (2021)
11. Vlachas, K., Garland, A., Quinn, D., and Chatzi, E."Parametric reduced-order modeling for component-oriented treatment and localized nonlinear feature inclusion".*Nonlinear Dynamics*, 112(5):3399–3420 (2024)
12. Najera-Flores, D.A., Jacobs, J., Dane Quinn, D., Garland, A., and Todd, M.D."Uncertainty Quantification of a Machine Learning Model for Identification of Isolated Nonlinearities With Conformal Prediction".*Journal of Verification, Validation and Uncertainty Quantification*, 9(2):021005 (2024)
13. Hastie, T., Tibshirani, R., and Friedman, J.H.*The elements of statistical learning*.Springer series in statistics. Springer, New York, NY, 2 edition (2009)
14. M, S., t, Hoffman, M.D., and Blei, D.M."Stochastic gradient descent as approximate bayesian inference".*Journal of Machine Learning Research*, 18(134):1–35 (2017)