Statistical Learning for Nonlinear Dynamical Systems with Applications to Aircraft-UAV Collisions

Xinchao Liu[1], Xiao Liu[1], Tulin Kaman[2], Xiaohua Lu[3], and Guang Lin[4]

[1]Department of Industrial Engineering, University of Arkansas
[2]Department of Mathematical Sciences, University of Arkansas
[3]College of Civil Aviation, Nanjing University of Aeronautics and Astronautics
[4]Departments of Mathematics, Statistics & School of Mechanical Engineering, Purdue University

Online Supplemental Materials

1 Extracting mass matrix and external force from FEA

Constructing the statistical model requires the mass matrix (required by the SVD (13)) and the external force (required by the reduced-order model (15)). Both quantities can be obtained from FEA that solves (6) numerically using the generalized-α time integration method for nonlinear structural dynamics (Farhat et al., 2014).

◇ Equilibrium: $\mathbf{M}\ddot{\mathbf{u}}_{n+1-\alpha_m} + \mathbf{f}_{\text{int}}(\mathbf{u}_{n+1-\alpha_f}, \dot{\mathbf{u}}_n) = \mathbf{f}_{\text{ext}}(t_{n+1-\alpha_f})$

◇ Initial conditions: $\mathbf{u}_0,\ \dot{\mathbf{u}}_0,\ \ddot{\mathbf{u}}_0 = \mathbf{M}^{-1}\left(\mathbf{f}_{\text{ext}}(t_0) - \mathbf{f}_{\text{int}}(\mathbf{u}_0, \dot{\mathbf{u}}_0; t_0)\right)$

◇ Recurrences:

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \Delta t\dot{\mathbf{u}}_n + \Delta t^2((\tfrac{1}{2} - \beta)\ddot{\mathbf{u}}_n + \beta\ddot{\mathbf{u}}_{n+1}) \quad \dot{\mathbf{u}}_{n+1} = \dot{\mathbf{u}}_n + \Delta t((1-\gamma)\ddot{\mathbf{u}}_n + \gamma\ddot{\mathbf{u}}_{n+1})$$

◇ Time average:

$$t_{n+1-\alpha_f} = (1-\alpha_f)t_{n+1} + \alpha_f t_n \quad \mathbf{u}_{n+1-\alpha_f} = (1-\alpha_f)\mathbf{u}_{n+1} + \alpha_f\mathbf{u}_n$$

$$\dot{\mathbf{u}}_{n+1-\alpha_f} = (1-\alpha_f)\dot{\mathbf{u}}_{n+1} + \alpha_f\dot{\mathbf{u}}_n \quad \ddot{\mathbf{u}}_{n+1-\alpha_m} = (1-\alpha_m)\ddot{\mathbf{u}}_{n+1} + \alpha_m\ddot{\mathbf{u}}_n$$

where α_m and β are the parameters that control the stability of the scheme, α_f and γ control the accuracy of the scheme, and Δt denotes the computational time step.

After numerically solving the governing equation (6) by FEA, we can obtain the mass matrix and forces (which are required by the statistical model). Here, we assume that the mass matrix is consistent during the entire collision process, and both the mass matrix and the external forces are determined by all the elements of the discretized structure of interest (i.e., aircraft nose in the numerical example). The procedure of assembling the mass matrix and external force is given as follows:

- ⋄ Input: node matrix $\mathbf{H}$, element matrix $\mathbf{E}$ and nodal forces (extracted from FEA)
- ⋄ MassAssembler3Dshell($\mathbf{H}$, $\mathbf{E}$), $M_{m,n} = \int_\Omega \rho \Phi_n(\mathbf{x}) \Phi_m(\mathbf{x}) d\mathbf{x}_\Omega$, for $m, n = 1, 2, \cdots, N$.
- ⋄ ForceAssembler3Dshell($\mathbf{H}$, $\mathbf{E}$), i.e., $f^{\text{ext}}_{m,i} = \int_\Omega f_i \Phi_m(\mathbf{x}) d\mathbf{x}_\Omega + \int_{\Gamma_s} g_i \Phi_m(\mathbf{x}) d\mathbf{x}_\Gamma$

2 Double Slit experiment (a linear case of ODE)

As discussed in Section 3.2, for a nonlinear governing ODE (13), the nonlinearity in $\mathbf{f}_{\text{int}}$ prevents us from using statistical learning approaches to learn the input-output mapping of $\mathbf{f}_{\text{int}}$. The learning errors are quickly accumulated and magnified due to the chaotic nature of the nonlinear governing equation (see Section 3). If, on the other hand, the governing ODE is linear, Appendix B demonstrates the effectiveness of learning a governing ODE using the statistical hyper-reduction approach.

We consider the classical Thomas Young's Double Slit experiment governed by a partial differential equation (i.e., the wave equation commonly found in science and engineering).

$$\frac{\partial^2 u}{\partial t^2} - \Delta u = f_i \quad \text{in } \Omega \tag{1}$$

with boundary conditions and initial values

$$\begin{aligned} &u = u(t) \text{ on } \Gamma_D, \qquad n \cdot \nabla u = 0 \text{ on } \Gamma_N \\ &u = 0 \text{ in } \Omega \text{ for } t = 0, \qquad \dot{u} = 0 \text{ in } \Omega \text{ for } t = 0 \end{aligned} \tag{2}$$

where u denotes the quantity of displacement. The wave equation describes wave propagation in a median such as a liquid and a gas. As shown in Figure 1, the problem under consideration has a domain composed of a square with two smaller rectangular strips added on one side.

The Dirichlet boundary condition is imposed on the line segments $\Gamma_D = \{x : x_1 = -0.25\}$, and the Neumann boundary condition is imposed on the rest of the boundaries. The source is given by $f_i(x_{\Gamma_D}) = \mu_1 \sin(\mu_2 \pi t)$, where μ_1 and μ_2 are the parameters of interest.

After discretization with a mesh size of 0.025, a number of 5731 nodes and 2968 elements are defined on the domain. Hence, we obtain a 2968×2968 system of ODEs

$$\mathbf{M}\ddot{\boldsymbol{\xi}}(t) + \mathbf{A}\boldsymbol{\xi}(\boldsymbol{t}) = \mathbf{b}(t, \boldsymbol{\mu}), \quad t \in \mathbf{J} \tag{3}$$

where $\boldsymbol{\xi}$ is the time-dependent nodal displacement, $\mathbf{M}$, $\mathbf{A}$ and $\mathbf{b}$ are matrices, $\mathbf{J}$ is a set of discretized time steps, and $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ is a collection of the parameters of interest.

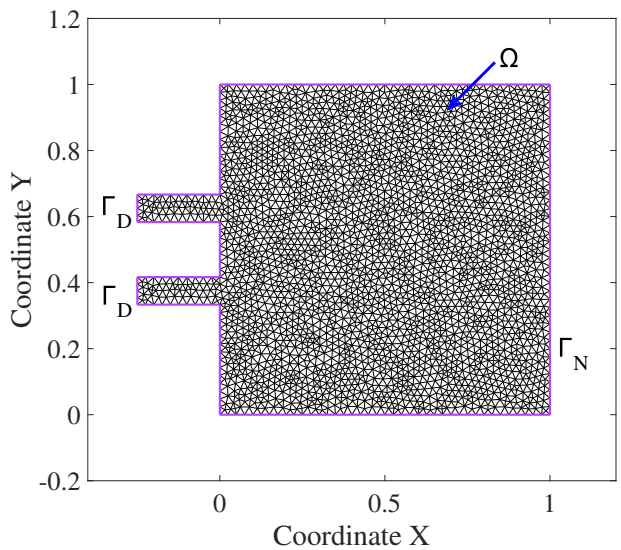

Figure 1: Discretization of the domain $\boldsymbol{\Omega}$ for the Double Slit experiment

Here, our goal is to (i) use the hyper-reduction approach described in Section 3 to learn the parameter-output mapping based on the data generated by the governing ODE; and (ii) predict the solutions at an unknown parameter setting using the statistical model. To prepare the training dataset, we consider values of μ_1 and μ_2 from a mesh grid of $\{80, 84, 88, 92, 96, 100\} \otimes \{3.0, 3.4, 3.8, 4.2, 4.6, 5.0\}$, i.e., 36 combinations of μ_1 and μ_2. Using 10 POD bases, the reduced-order governing physics and compressed snapshot data are prepared using the approach described in Section 3. The Gradient-Boosted Trees are adopted to learn the mapping $\mathbf{f}_{\text{int}}$ in the reduced-order model (Zhang and Jung, 2020).

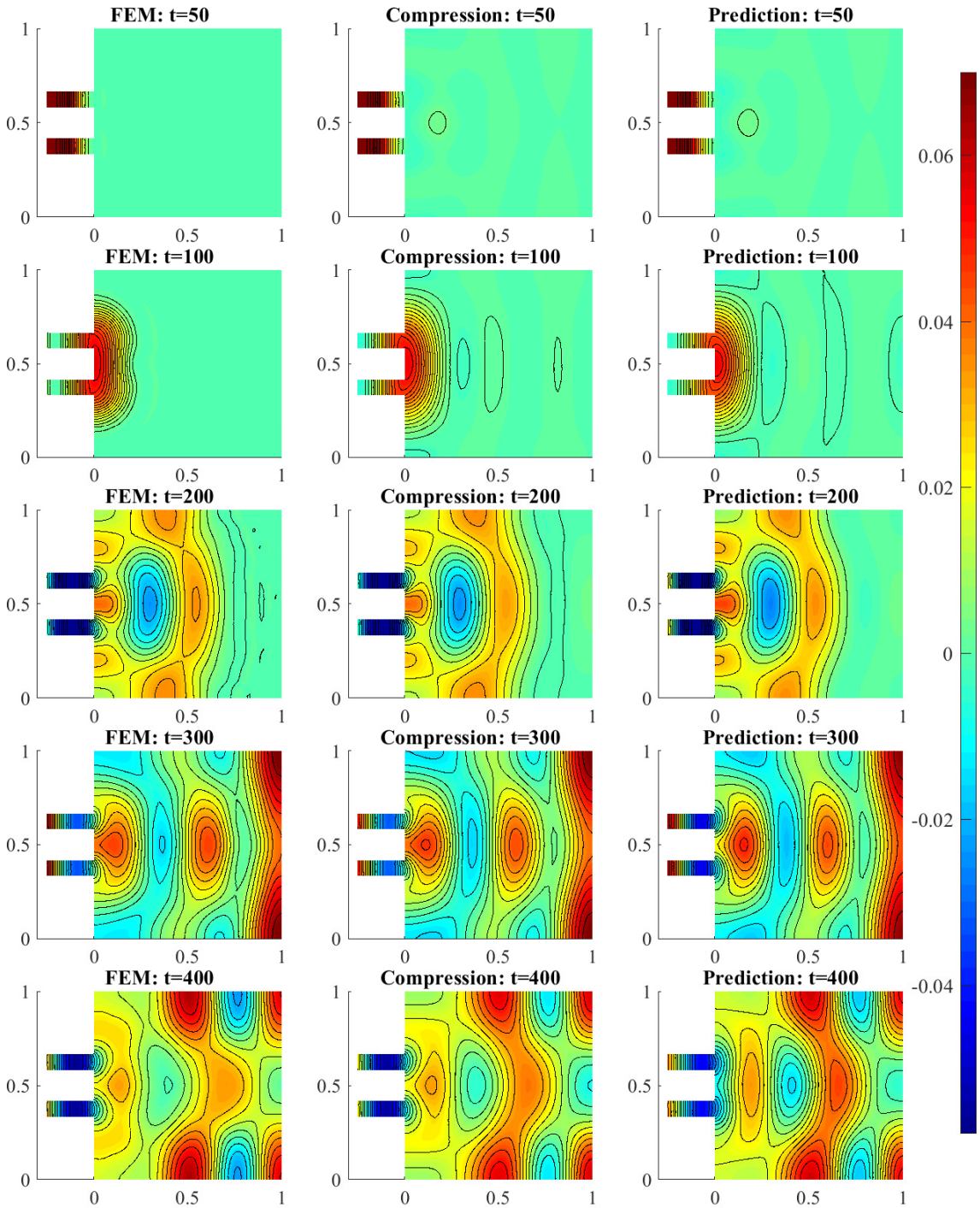

Figure 2: Nodal displacement fields over time for the double-slit experiment. Left column: FEM; Middle column: reconstructed displacement fields by keeping only 10 POD bases; Right column: predicted displacement fields using statistical approaches.

The established statistical model is then used to predict the displacement at the input condition of $\mu_1^* = 85$ and $\mu_2^* = 4.3$. The predicted displacement field is given in the right column of Figure 2. The figure also includes the (actual) displacement generated from FEM (left column), as well as the reconstructed displacement by keeping only 10 POD bases (center column). It is seen that the predicted displacement fields at different time steps well match the outputs generated by FEM, justifying the hyper-reduction approach outlined by (14) for linear problems.

3 Comparison studies and discussions

Appendix C provides additional numerical comparison studies and discussions on the proposed approach, focusing on three important aspects: interpretability, accuracy and computational time.

1). Comparison for interpretability

We first provide a high-level overview to the methods included in the comparison study, and then, discuss their differences in terms of model interpretability.

• Firstly, the proposed approach can be conceptually represented as:

$$\text{Proposed Approach: } \boldsymbol{p} \overset{\text{mGPR}}{\rightarrow} \boldsymbol{\xi^F} \overset{\text{function-to-function regression}}{\rightarrow} \boldsymbol{\xi^q} \overset{\text{mFPCA}}{\rightarrow} \boldsymbol{q} \overset{\text{POD}}{\rightarrow} \boldsymbol{u} \tag{4}$$

For any parameters $\boldsymbol{p}$, the mGPR is used to predict coefficients $\boldsymbol{\xi^F}$ that determine the external force. Then, the function-to-function regression is used to predict the coefficients $\boldsymbol{\xi^q}$ that determine the reduced-order state. Finally, the reduced-order and fully-order state are computed from mFPCA and POD, respectively. We see that, the proposed model is highly interpretable because it explicitly captures the chain relationship: impact conditions→force generated→deformation caused.

• The first Alternative Approach (denoted by AA1) to be compared with is a GP model that directly builds the mapping from parameters $\boldsymbol{p}$ to the full-order state at any given time,

which can be conceptually represented by

$$\text{Alternative Approach 1 (AA1): } \boldsymbol{p} \overset{\text{GP}}{\rightarrow} \boldsymbol{q} \overset{\text{POD}}{\rightarrow} \boldsymbol{u}. \tag{5}$$

Hence, this is a less interpretable and pure data-driven model that directly learns the input-output relationship without explicitly modeling the relationship between force and deformation. In addition, we construct AA1 for individual time steps, i.e., we construct a GP spatial model at each time step without considering the temporal connection.

In particular, let $\{(\mathbf{p}_i, \mathbf{q}(\mathbf{p}_i))\}_{i=1}^{N_\mathbf{P}}$, $\mathbf{p}_i \in \mathbb{R}^p$ and $\mathbf{q}(\mathbf{p}_i) \in \mathbb{R}^K$, AA1 considers a multivariate Gaussian Process Regression (mGPR):

$$\mathbf{g} \sim \mathcal{MGP}(\mathbf{0}, k'_\mathbf{q}, \mathbf{\Lambda}_\mathbf{q}), \qquad \mathbf{q}(\mathbf{p}_i) = \mathbf{g}(\mathbf{p}_i). \tag{6}$$

Then, the system state at any give time step t is given by $\hat{\mathbf{u}}_*(t) = \mathbf{V}(\hat{\mathbf{q}}_*(t) + \bar{\mathbf{q}}(t))$ and $\text{cov}(\hat{\mathbf{u}}(t), \hat{\mathbf{u}}(t')) = \mathbf{V}\text{cov}(\hat{\mathbf{q}}(t), \hat{\mathbf{q}}(t'))\mathbf{V}^T$. As one may see, AA1 is computationally less efficient because it requires a mGPR to be constructed for each time step.

• The second Alternative Approach (denoted by AA2) to be compared with is a GP model that directly builds the mapping from the parameter $\boldsymbol{p}$ to the reduced-order state at a given time, which can conceptually represented by

$$\text{Alternative Approach 2 (AA2): } \boldsymbol{p} \overset{\text{GP regression}}{\rightarrow} \boldsymbol{\xi}^{\boldsymbol{q}} \overset{\text{mFPCA}}{\rightarrow} \boldsymbol{q} \overset{\text{POD}}{\rightarrow} \boldsymbol{u} \tag{7}$$

Hence, this alternative approach used the GPR to directly predict the coefficients $\boldsymbol{\xi}^{\boldsymbol{q}}$ that determine the reduced-order state, and subsequently compute the reduced-order and fully-order states. In other words, AA2 skips the prediction of external force (which is the reason that leads to surface deformation), and is less interpretable.

In our comparison study, let $\{(\mathbf{p}_i, \boldsymbol{\xi}^\mathbf{q}(\mathbf{p}_i))\}_{i=1}^{N_\mathbf{P}}$, $\mathbf{p}_i \in \mathbb{R}^p$ and $\boldsymbol{\xi}^\mathbf{q}(\mathbf{p}_i) \in \mathbb{R}^L$, AA2 considers

a multivariate Gaussian Process Regression (mGPR):

$$\mathbf{g} \sim \mathcal{MGP}(\mathbf{0}, k'_{\boldsymbol{\xi}^{\mathbf{q}}}, \boldsymbol{\Lambda}_{\boldsymbol{\xi}^{\mathbf{q}}}), \qquad \boldsymbol{\xi}^{\mathbf{q}}(\mathbf{p}_i) = \mathbf{g}(\mathbf{p}_i). \tag{8}$$

Once $\hat{\boldsymbol{\xi}}^{\mathbf{q}}_*$ has been obtained, it follows from the mFPCA that $\hat{\mathbf{q}}_*(t) = \bar{\mathbf{q}}(t) + \hat{\mathbf{q}}^Z_*(t)$ and $\text{cov}(\hat{\mathbf{q}}(t), \hat{\mathbf{q}}(t'))$. Finally, based on the POD, the predicted original high-dimensional state vector $\mathbf{u}_*(t)$ is obtained as $\hat{\mathbf{u}}_*(t) = \mathbf{V}\hat{\mathbf{q}}_*(t)$ and $\text{cov}(\hat{\mathbf{u}}(t), \hat{\mathbf{u}}(t')) = \mathbf{V}\text{cov}(\hat{\mathbf{q}}(t), \hat{\mathbf{q}}(t'))\mathbf{V}^T$.

• The third alternative approach (AA3) only slightly differs from our approach. In the proposed approach, we establish the mapping between the cumulative force and displacement, while AA3 establishes the mapping between instantaneous force and displacement. There are two reasons why we chose to model the relationship between cumulative force and displacement:

(i) from the *law of motion*, force corresponds to acceleration while cumulative force is associated with displacement; and

(ii) from the *statistical learning* perspective, cumulative force is smoother and less volatile than instantaneous force. To elaborate, the profile of instantaneous force is shown in the figure on the next page. If we compare the profiles of the instantaneous force to those of the cumulative force (see Figure 9 in the revised paper), we clearly see the cumulative force is much smoother.

Precisely due to this reason, if the instantaneous force is used, more mFPCA bases are needed to retain at least 90% of the total variation. For example, $M = 12$ only retains 49.61% of the total variation of the instantaneous force profile, while $M = 12$ can already capture 97.43% of the total variation of the cumulative force profile (this is what we did in our approach). Hence, in our comparison study, we let $M = 31$ for AA3 which explains 96.13% of the total variation. Obviously, a larger value of M makes the dimension of the problem higher and increases the computational burden (i.e., in the MLE). In summary,

AA3 is slightly less interpretable from the laws of motion, and computationally less efficient.

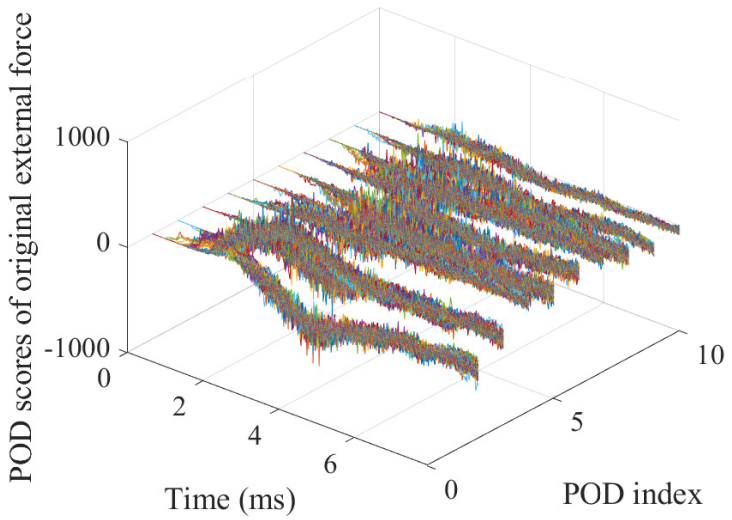

Figure 3: The instantaneous external force profiles

2). Comparison for accuracy

Next, we compare the accuracy of different approaches. As discussed above, the proposed approach explicitly captures the chain relationship (i.e., impact conditions→force→deformation), while AA1 and AA2 use data-driven approaches to directly capture the input-output relationship as accurately as possible (i.e., impact condition→deformation, or, impact condition→reduced-order state→deformation). Hence, one might think that the data-driven driven approaches AA1 and AA2 should be more accurate (we also thought so before the numerical experiment). However, the numerical study shows that the proposed approach has comparable performance in terms of prediction accuracy while greatly improving the model interpretability. This is clearly seen in the figure below, which shows the boxplot of the LOOCV MRE for all 35 collision conditions for the current method, AA1, AA2 and AA3 at 4, 6, and 8 milliseconds after collision.

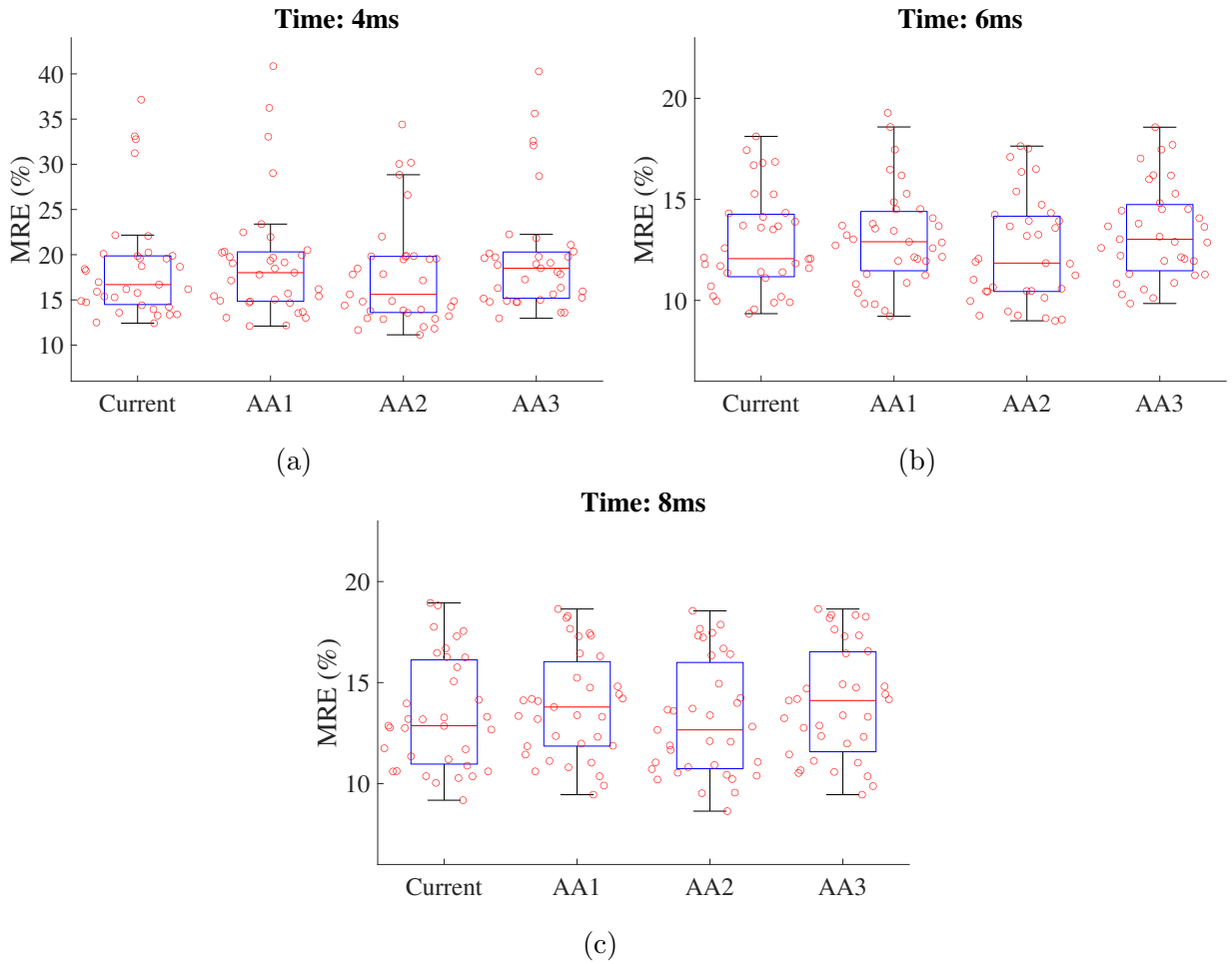

Figure 4: Comparison of LOOCV MRE for the current model, AA1, AA2 and AA3 at different times after collision: (a) 4ms; (b) 6ms; (c) 8ms.

3). Comparison for computational time

Finally, we compare the computational time of the proposed model, AA1, AA2 and AA3. Because all models consume the same amount of time in data pre-processing and the SVD operation, we compare the time required for making statistical inference by different approaches. As shown in the Figure on the next page, the proposed model in this paper is the fastest, while AA1 is the slowest one. This is mainly due to the fact that AA1 requires repeatedly building an mGPR model for each time step. AA3 is also slower because it requires a larger number of bases to capture the instantaneous force profile which makes the dimension of the problem higher.

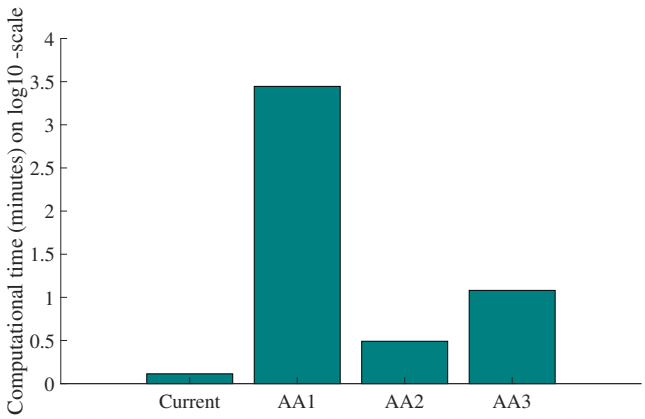

Figure 5: Comparison of computational time (on a base 10 logarithmic scale) for different models.

In summary, *the proposed approach is more interpretable with comparable accuracy to other data-driven approaches, and is computationally faster*, as summarized in the table below.

	Conclusions about the proposed approach
Interpretability	More interpretable than AA1, AA2 and AA3
Accuracy	Comparable performance
Computation	Fastest among all approaches

References

Farhat, C., Avery, P., Chapman, T., and Cortial, J. (2014), "Dimensional reduction of non-linear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency," *International Journal for Numerical Methods in Engineering*, 98, 625–662.

Zhang, Z. and Jung, C. (2020), "GBDT-MO: Gradient-Boosted Decision Trees for Multiple Outputs," *IEEE Transactions on Neural Networks and Learning Systems*.