Article:
Pitchforth, D.J., Jones, M.R., Gibson, S.J. et al. (2026) Physics-informed kernel mixtures for structural dynamics. Mechanical Systems and Signal Processing, 254. 114330. ISSN: 0888-3270

https://doi.org/10.1016/j.ymssp.2026.114330

Physics-informed kernel mixtures for structural dynamics

D.J. Pitchforth[a], M.R. Jones[a], S.J. Gibson[a], E.J. Cross[a]

[a] *Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom*

Abstract

The effective integration of physical prior knowledge and measured data is critical for developing robust physics-informed machine learning frameworks. Calibrating the balance between structured physical approximations and data-driven flexibility is a key challenge of model development, particularly when physical models are only valid within specific operational regimes. In such cases, a global reliance upon a physical approximation can lead to model misspecification. This paper introduces a physics-informed kernel mixture framework for Gaussian process regression, capable of dynamically varying the reliance upon available physical knowledge based on identifiable switching variables. This ensures that physical understanding is prioritised in valid regimes and relaxed in favour of flexible data-driven components elsewhere. Rather than relying on soft constraints or loss-based penalties, known physics is explicitly embedded within the construction of kernels, enforcing desirable properties within predictions (e.g. quadratic lift force, localised behaviours). This allows model structure to mimic available physical intuition in an interpretable manner. Furthermore, the framework incorporates regime-dependent heteroscedastic noise to accurately capture varying uncertainty across different operational states. The versatility of the physics-informed kernel mixture framework is demonstrated through two distinct engineering case studies: the aerodynamic directional loading of a long-span suspension bridge and the prediction of aircraft wing strain during in-flight manoeuvres. The proposed kernel structures improve predictive accuracy and extrapolation whilst recovering an interpretable representation of regime-dependent system behaviour.

1. Introduction

Engineering knowledge provides a valuable resource for enhancing data-driven models, particularly in applications where measurements are scarce, expensive, or difficult to obtain. Physics-informed machine learning (PIML) [1, 2] seeks to exploit physical insight to improve predictive performance, reduce data requirements, and provide insight into model behaviour. In engineering systems, where physical understanding is often well developed but incomplete, combining physics-based and data-driven approaches offers a promising route to robust and interpretable modelling [3].

A core challenge of PIML is calibrating how strongly a model should rely upon physical knowledge relative to measured data [4, 5]. Physical models are necessarily built on simplifying assumptions and approximations, which may only be valid under particular operating conditions. When these assumptions hold, physical knowledge can provide strong inductive bias [6] and improved extrapolation [7]; when they do not, an over-reliance on physics can be detrimental to predictive accuracy. Conversely, under-utilising available physical knowledge can waste valuable prior information and reduce interpretability.

In many engineering systems, the validity of physical knowledge varies with time, operating regime, or environmental conditions. This motivates a switching perspective, in which the relative contribution of different model components vary across regimes. Engineering structures often exhibit intermittent or condition-dependent physical effects that are poorly captured by data alone, particularly when such behaviour is rare or unevenly sampled [8, 9, 10]. Rather than enforcing a single global description, these systems call for models that adapt their structure across operating conditions. Crucially, this adaptation should be achieved in a controlled and interpretable manner, allowing the influence of physical knowledge to change without sacrificing model transparency.

To capture a continuous, regime-dependent reliance upon physical knowledge, a modelling framework is needed that allows physical assumptions to be strengthened or relaxed as operating conditions change. **Gaussian process (GP)** regression provides exactly this capability, whilst quantifying uncertainty in model components and their relative weighting. GPs are flexible, non-parametric, Bayesian models that have been successfully applied to a range of engineering problems, including crack growth [11, 12], tool wear [13], and wind turbine power curves [14]. The covariance function (kernel) of a GP encodes prior assumptions about system behaviour and can enforce properties such as smoothness, periodicity, or locality [15, 16]. Carefully designed kernels can also embed physically meaningful behaviours [17, 18] or enforce symmetries [19, 20]. Here, structured, physically-derived kernels are combined with more flexible, general-purpose kernels to allow both physical insight and data-driven adaptation within a unified GP framework. The proposed physics-informed kernel mixtures modulate the relative explanatory power of model components as identifiable switching variable(s) change.

1.1. Related work

Modelling systems with regime-dependent behaviour is a foundational challenge in structural health monitoring (SHM) [21], amongst many other fields. This section contextualises the proposed physics-informed kernel mixture by exploring the evolution of general switching models, their application within structural dynamics, and the recent development of physics-informed architectures.

Switching and regime-dependent models

Switching models provide a versatile framework for representing systems whose dynamics change across operating regimes [22]. Within this class, discrete latent-switching approaches, such as Markov-switching models [23, 24], represent system parameters that evolve according to a latent chain, with extensions allowing for state-dependent transitions [25] or variable regime lengths [26]. While these provide robust, interpretable regime representations, their transitions are inherently discrete. Consequently, they are often complemented by models that capture gradual transitions or threshold-crossing behaviour, such as Threshold AutoRegressive (TAR) [27, 28] and Smooth Transition AutoRegressive (STAR) models [29].

Further flexibility is offered by Mixture-of-Experts (MoE) architectures [30, 31], which combine localised sub-models via gating functions. While these approaches are highly effective for point estimation, there is an ongoing interest in integrating them into fully probabilistic frameworks to further enhance the uncertainty quantification required for safety-critical engineering decisions [32, 21]. Gaussian process (GP) mixture-of-experts [33, 34, 35] provide a Bayesian solution, though mixing is typically controlled by latent gating rather than being explicitly tied to physical input locations.

At the kernel level, non-stationarity is addressed through change-point [36, 37] and multi-dimensional change-surface models [38, 39]. Recently, Volodina [40] introduced a framework for non-stationary kernel mixtures using deterministic, input-dependent weights to combine covariance components while maintaining closed-form inference.

Switching models for structural dynamics

In structural health monitoring (SHM), the motivation for switching models often stems from physical phenomena exhibiting distinct regime shifts; for example, the Z24 bridge's natural frequencies exhibit a sudden increase below a 0°C temperature threshold [41]. To capture such non-stationarity, GPs offer a flexible means of adapting covariance structures. Switched latent force models [42] have been successfully applied to mechanical systems with discontinuous forces, such as identifying complex stick-slip friction behavior [43]. Other regional adaptation strategies include Treed GPs [44], which use hierarchical partitioning, and additive GPs [45], which decompose responses into interpretable component functions.

Physics-informed switching models

A small number of physics-informed architectures explicitly model regime-dependent switching between physics and data components. MoE frameworks [31] exemplify this, allowing specialised physics-informed experts to operate in distinct regions of the solution space. For instance, Chalapathi [46] utilised an MoE neural network to enforce hard physical constraints by assigning different experts to specific subproblems. Similarly, Yuan [47] integrated first-principles knowledge with multiple expert networks to predict energy consumption in hybrid manufacturing, using gating mechanisms to adaptively combine outputs across shifting operating regimes.

Adaptive physics-informed modelling focuses on the dynamic balancing of data-driven and physical influences. This is typically achieved in Neural Networks through automated loss weighting [48, 49, 50] or multi-objective optimization [51]. In the GP literature, the enforcement of localised physical constraints, such as monotonicity in heat transfer problems [52], allows models to adapt to local phenomena. While soft constraints are useful to ensure physical plausibility (e.g., rising temperatures), they do not embed knowledge of governing equations into the prediction as we explicitly do here.

1.2. Contributions

While switching models, physics-informed machine learning, and non-stationary Gaussian processes have been widely studied, relatively little attention has been given to modelling the regime-dependent validity of physical knowledge within Gaussian processes. Classical change-point and switching GP models typically focus on statistical regime changes, and many hybrid physics-informed architectures rely on discrete expert selection or gating networks. As a result, the degree to which physical knowledge influences predictions is often implicit, tied to discrete model choice, or difficult to interpret. This motivates a framework in which the contribution of physics and data can vary continuously and is interpretable across operating conditions.

This paper introduces a physics-informed kernel mixture framework that adaptively weights physical and data-driven components according to the validity of the underlying physics. The primary contributions of this work are:

1. **Physics-informed formulation of kernel mixtures:** A novel extension to a kernel mixture framework in which physical intuition of a systems' behaviour is embedded directly within the covariance function structure. The relative contribution of physics-informed and data-driven kernel components vary continuously as a function of independent switching variable(s).
2. **Interpretable and flexible representation of regime transitions.:** Switching behaviour is governed by kernel hyperparameters with clear physical interpretations, providing an intuitive description of when transitions occur and how rapidly they take place. Multiple simpler kernels may be combined as building blocks to represent more complex switching behaviours.

3. **Integrated heteroscedastic uncertainty quantification:** Regime-dependent noise terms allow the model to capture varying levels of uncertainty across different operational states. This is implemented at the covariance-function level, preserving standard GP regression and closed-form inference.

The proposed methodology is demonstrated through two distinct engineering case studies, highlighting its ability to improve predictive performance and capture regime-dependent physical behaviour.

2. Theoretical Foundation: Kernel Mixtures

A natural goal of regime-dependent and switching models is to capture the behaviour of multiple processes and the mechanism through which they combine. Consider M independent processes $f_1(x_1), f_2(x_2), \dots f_M(x_m)$, each representing a different operating regime or physical phenomena. These may be combined within a weighted sum to approximate a systems' overall behaviour:

$$f(x) = \sum_{m=1}^{M} w_m(x_m) f_m(x_m) \tag{1}$$

where $\{x_m\}_{m=1}^{M}$ are the inputs to each process, and the weights $\{w_m(x_m)\}_{m=1}^{M}$ dictate the influence of a particular process $f_m(x_m)$ at any given point. Weighted sums of processes form the basis for many switching model architectures, largely defined by types of weighting function and what kind of model is used to represent $\{f_m(x_m)\}_{m=1}^{M}$.

The framework developed in this paper builds upon the theory of non-stationary kernel mixtures, that allow the covariance structure of a Gaussian process to vary across an input space. Following the general formulation for weighted mixtures [40], a global covariance function $k(x, x')$ can be constructed as a sum of M constituent kernels, each modulated by a deterministic weighting function $w_m(x_m)$:

$$k(x, x') = \sum_{m=1}^{M} w_m(x)\, w_m(x')\, k_m(x, x') \tag{2}$$

Provided that the individual kernels $k_m(x, x')$ are valid (positive semi-definite and non-infinite), the combination of kernels via linear operators remains a valid kernel. This ensures that more complex kernel structures constructed using the framework remain valid. This is a versatile construction and it generalises several established methodologies in the literature:

- **Additive Kernels:** Recovered when weights are globally constant ($w_m(\mathbf{z}) = 1$ or $w_m(\mathbf{z}) = \sigma_f^2$), resulting in a sum of kernels $k = \sum k_m$ across input dimensions [45].
- **Change-point Kernels:** Typically utilised for $M = 2$ using sigmoidal weighting functions $w_m(x) = \frac{1}{1+\exp(-a(z-x_0))}$ to blend two covariance structures across regimes [36].

- **Treed GPs:** Recovered when weights are a discrete indicator function ($w_m \in \{0, 1\}$), determined by the leaves of a decision tree. This partitions the input space, with hard switches between regimes [44].

While these existing approaches effectively capture statistical non-stationarity, they typically assume that switching logic is a function of the primary process inputs x, while also taking a purely data-driven approach to modelling the behaviour in each regime. The following section introduces a physics-informed extension that leverages this mixture logic to include and model the regime-dependent validity of physical prior knowledge using decoupled switching variables.

3. Proposed framework: Physics-informed kernel mixtures

As presented in the previous section, kernel mixtures provide a flexible framework, capable of varying the influence of multiple stochastic model components whilst maintaining closed-form inference. Here, a physics-informed extension is proposed, where kernels are constructed to modulate and interpret the reliance on physical knowledge within Gaussian process regression.

First, a target process $f(x, z)$ is defined as a weighted sum of M independent Gaussian processes:

$$f(x, z) = \sum_{m=1}^{M} w_m(z_m) f_m(x_m) \tag{3}$$

where each component $f_m(x_m) \sim \mathcal{GP}\big(m_m(x_m),\ k_m(x_m, x'_m)\big)$ is modelled using either a physically-derived kernel $k_{Phys}(x_m, x'_m)$ or a flexible, data-driven kernel $k_{Data}(x_m, x'_m)$. This formulation decouples the process inputs x and switching variables z:

- **Process inputs** $x \in \mathcal{X}$: These variables are directly involved in the physical process being modelled, such as time, velocity or structural response measurements.
- **Switching variables** $z \in \mathcal{Z}$: An effective choice of z relates closely to how a physical relationship might change. For example, temperature, humidity, excitation level, or measures of turbulence can all affect the extent to which one might want to rely on a given piece of physical knowledge.

Following the derivation of the kernel mixture in Equation 2, a covariance function may be defined over the joint input space of the process and switch ($\mathcal{X} \times \mathcal{Z}$):

$$k(z, z', x, x') = \sum_{m=1}^{M} w_m(z) w_m(z') k_m(x, x'). \tag{4}$$

This formulation implies that the similarity between two points is not just a function of their process space proximity (x vs x'), but also whether the operating conditions (z and z') place them within the same physical regime. This allows the model to treat identical process inputs differently if they occur under different operating states.

The primary utility of the proposed framework lies in the ability to translate qualitative engineering intuition into a quantitative, probabilistic covariance structure. By deriving kernels from the covariance of a physical process, one can constrain the model to respect physical boundaries. While the framework in this form is general, several key architectures outlined here arise from it that address common challenges in structural dynamics.

Local physics validity: In many engineering scenarios, a physical law is derived under simplifying assumptions (e.g., small-angle approximations or linear elastic behavior) that are only valid near a specific design point z_0. This can be encoded through a localised weighting function:

$$w_{local}(z) = \exp\left(-\frac{(z - z_0)^2}{2\ell^2}\right) \tag{5}$$

where the hyperparameter ℓ determines the width of the localised region and decay of trust in physical knowledge. This weighting function can construct models that rely more heavily on the physical kernel k_{phys} when the operating variable z is near z_0, but smoothly transitions to a more flexible data-driven kernel k_{data} as the system enters other regimes.

Regime dependent transition: Certain physical phenomena in mechanical systems are only triggered when environmental or operational thresholds are met, such as the onset of turbulence at high Reynolds numbers or the amplification of structural non-linearities at high amplitudes. By employing a sigmoid weighting function, physical knowledge can phase in or out across a threshold τ:

$$w_{sig}(z) = \frac{1}{1 + \exp\left(-a(z - z_0)\right)} \tag{6}$$

where the hyperparameters z_0 and a govern the where and how quickly the regime transition occurs. Using a pair of opposing sigmoids allows for a smooth transition between models. This is particularly valuable for avoiding the numerical instabilities and loss of gradient information associated with discrete, non-differentiable change-points.

Multi-physics blending: When a system transitions between multiple distinct physical states, a single switch is insufficient, M components may be handled using a softmax parameterization:

$$w_{sfmax}(z) = \frac{\exp(g_m(z))}{\sum_{j=1}^{M} \exp(g_j(z))} \tag{7}$$

where $g_m(z)$ are auxiliary functions, often linear transforms, that define the dominance of each regime. This provides a mechanism to blend multiple states or phenomena present within a system, with independent specification of when each occurs. This formulation of softmax ensures that all weights consistently sum to 1.

Composite physical validity: In many mechanical systems, a physical law may be valid only within a specific operating window that is defined by multiple constraints. For example, a physical

approximation might only be trusted during high-speed operation (sigmoid transition) but only for a specific temperature range (local validity). This can be encoded by defining a composite weight:

$$w_{comp}(z_1, z_2) = w_{sig}(z_1) \times w_{local}(z_2) \tag{8}$$

This multiplicative structure acts as a logical AND gate, ensuring the physical kernel k_{phys} only influences the posterior when both operational conditions are satisfied. Similar composite weights may be constructed however best reflects physical intuition for a system. For example, later in this paper, two sigmoids are composed to identify high speed northerly and southerly wind loading of a bridge deck.

Regime-dependent heteroscedastic noise: Beyond the signal mean and covariance, physical knowledge may extend to the expected noise present in a system. For instance, signals from a machine tool may be significantly noisier during high-speed drilling. In such cases, the mixture can include a heteroscedastic noise term $\epsilon(z)$, where the observation noise is itself a function of the operating variable z:

$$y = f(x, z) + \epsilon(z), \quad \epsilon(z) \sim \mathcal{N}\left(0, \sum_{m=1}^{M} w_m(z)\sigma_{n,m}^2\right) \tag{9}$$

By assigning different noise levels to different regimes, a more representative representation of uncertainty present within the system may be attained. This noise formulation is inherently less flexible than fully non-parametric heteroscedastic models [53], as it requires a user to map their noise assumptions to weighting functions. However, in circumstances where it is desirable for noise to vary in a structured manner (e.g. local change, linear/quadratic growth), it allows for a higher degree of control over model behaviour. Exact inference is also maintained for the resulting mixture.

Application to structural dynamics

The proposed framework addresses the inherent non-stationarity of structural systems by encoding regime-dependent constraints directly into the covariance structure. By mapping the switching variable z to specific weighting functions, the model transitions between physically-derived and data-driven kernels without the need for discrete boundaries. For smaller dataset sizes, the approach ensures the resulting process remains a valid Gaussian process with an exact marginal likelihood. Consequently, the model offers a tractable solution for Structural Health Monitoring (SHM) tasks where the reliability of physical laws is often regime dependent. The applicability of models is demonstrated in the following case studies involving a long-span suspension bridge and an aircraft wing.

4. Case study aims and benchmark models

The case studies in this paper perform two main roles; firstly, they aim to outline the kernel building process, showing how physical understanding of a system can be embedded within a model using the

proposed framework. Although there are several decisions to be made (e.g. choice of weighting function, switching variable), the authors argue that these may be made fairly intuitively, and using physical insight to guide these decisions can be a strength of the approach. The second role of the case studies is to compare the performance of the physics-informed kernel mixtures against a selection of benchmark model architectures. These included data-driven, physics-informed, global and switching approaches. They were selected to isolate and assess both the value of physical knowledge incorporation and the benefit of a switching architecture. The benchmark models, against which we compare our approach, are defined as follows:

- **SE kernel:** A global, stationary squared exponential kernel $k_{SE}(x, x')$. This acts as a standard non-parametric benchmark that assumes no regime-specific behavior or physical structure.
- **Spectral mixture kernel:** A highly flexible kernel that models the spectral density of a process as a mixture of Gaussians [54]. Using Q additive kernel components, the kernel is able to discover localised, periodic structure in data, without physical insight.
- **Global physics-informed kernel:** A global application of the additive kernel $k(x, x') = k_{Phys}(x, x') + k_{Data}(x, x')$. This determines if a physical approximation may be utilised appropriately by the model without an explicit switching mechanism.
- **Black-box switching kernel:** A non-stationary mixture of two data-driven kernels ($k_{SE,1}$ and $k_{SE,2}$). To ensure a fair comparison, this benchmark utilises the same switching variables z and sigmoidal weights $w(z)$ as the proposed model. This isolates the contribution of the physical kernel.

To measure the performance of models, two error metrics were used; the Normalised Mean Squared Error (NMSE) and the mean standardised log loss (MSLL). The NMSE quantified the deviation of the predictive mean from the true target and the MSLL provided a probabilistic measure of performance. The MSLL is defined:

$$MSLL = \frac{1}{N} \sum \{ -\log p(\mathbf{y}_\star | X, \mathbf{y}, X_\star) + \log p(\mathbf{y}_\star; \mathbb{E}(\mathbf{y}), \mathbb{V}(\mathbf{y})) \}, \tag{10}$$

where the training set is denoted $\{X, \mathbf{y}\}$, test inputs $X_\star$ and test set observations $\mathbf{y}_\star$. As this metric is a measure of the negative log probability of the model, a lower value is indicative of a better fit. Models were optimised in line with standard GP practice [15], using the Negative Log Marginal Likelihood (NLML) as a cost function. Details of model training and computation time are provided in Appendix C.

5. Case study A: Directional wind loading of the Tamar bridge

The Tamar Suspension Bridge is located in the south-west of the UK and was the feature of a monitoring campaign led by the Vibration Engineering Section (VES) at the University of Sheffield. The bridge is 643m long, with two towers, each 73m in height; it also forms part of a major connection to the city of Plymouth. For additional details on both the bridge and monitoring campaign, the reader is directed towards [55] and [56].

Available data from the bridge spans a three-year period from 2007 to 2011, with access to measurements from accelerometers, strain gauges, anemometers, temperatures, humidity and traffic levels. Here, a section of this dataset (2500 points) is used from the summer of 2008, focussing on measurements of wind speed, wind direction and vertical deck acceleration. The aim of this case study is to investigate the presence of changing physical behaviours from a full scale engineering structure. The example chosen here is the dependency of response to wind load on wind direction. Orientated in the east-west direction, the bridge is subject to higher deck accelerations when winds blow perpendicular to this at high speeds [55]. This is shown in plots of wind speed vs deck acceleration, separated by wind direction in Figure 1.

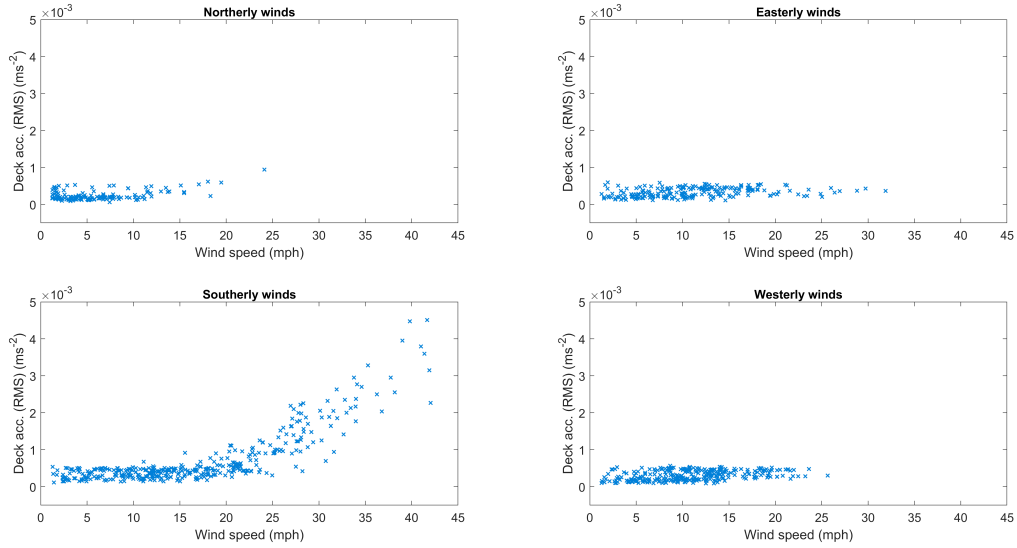

Figure 1: Wind speed vs vertical deck acceleration for a section of the Tamar bridge dataset. Winds are separated by direction, into northerly winds ($0 \pm 20°$), easterly winds ($90 \pm 20°$), southerly winds ($180 \pm 20°$) and westerly winds ($270 \pm 20°$). The northerly and southerly winds blow sideways across the bridge and induce much higher deck accelerations, particularly at high wind speeds.

5.1. Kernel design

When utilising the framework, the construction of the kernel should mimic available physical intuition of how a system behaves. This will influence the choice of physical approximations, derived physical kernels, weighting functions and switching variables. For the Tamar bridge, the directional relationship between wind speed and deck response $y(u, \theta)$ is formulated:

$$y(u, \theta) = f_1(u) + f_2(u, \theta) + \epsilon \tag{11}$$

where u is wind speed, θ is wind angle, $f_1(u)$ is the deck response to low-speed winds, $f_2(u, \theta)$ is the deck response to high-speed lateral winds and ϵ is noise. The cause of this directional relationship may be explained as lift force, incited by a pressure difference as wind flows over the bridge deck [57]:

$$L = \frac{1}{2}\rho C_L A u^2 \tag{12}$$

where ρ is fluid density, C_L is lift coefficient, A is projected area and u is wind speed. The quadratic term in this expression aligns with the observations in Figure 1, particularly for southerly winds. To incorporate this knowledge within our model, a second order polynomial kernel is used to represent lift force:

$$k_{Lift}(u, u') = \sigma_L^2 (u u^T + c)^2 \tag{13}$$

where σ_L^2 scales the variance and c allows the introduction of lower order terms (linears and constants). Draws from this kernel will be quadratic, enforcing desirable structure within the predictions of the model. To modulate the reliance upon the physical kernel, a composite weighting function (Equation 8), with two sigmoids is used:

$$w_{comp}(u, \theta) = w_{sig}(u) \times w_{sig}(\cos(2\theta)) \tag{14}$$

where w_{sig} are sigmoidal weighting functions (Equation 6), u is wind speed and θ is wind angle. The term $\cos(2\theta)$ is used as measure of how well a wind direction aligns with north or south, enforcing symmetry within the kernel. This is used to construct a physics-informed kernel mixture of the form:

$$\begin{aligned} k(\theta, \theta', u, u') = & \underbrace{w_{comp}(u, u', \theta, \theta')\, w_{comp}(u, u', \theta, \theta')\, k_{Lift}(u, u')}_{\text{Introduce lift force at high N/S winds...}} \\ & + \underbrace{w_{comp}(-u, -u', -\theta, -\theta')\, w_{comp}(-u, -u', -\theta, -\theta')\, k_{SE}(u, u')}_{\text{... otherwise, use a Squared Exponential kernel.}} \end{aligned} \tag{15}$$

The aim of this structure is to allow the knowledge of lift force to be utilised only for higher speed northerly and southerly winds. The construction of the kernel required a physical intuition of this,

however it was not necessary to know the specific speed or angle when this should occur. By fitting this kernel to data, it is possible to retrieve and interpret this information.

Additionally, to capture the growing variance present within the data when lift is induced, it is possible to introduce regime-dependent heteroscedastic noise. Here, in line with Equation 9, a noise scaling linearly with wind speed is used:

$$\epsilon(u) \sim \mathcal{N}\left(0, w(u)\,\sigma_n^2\right) \tag{16}$$

where $w(u) = u$ for linearly scaling noise. A prediction incorporating the heteroscedastic noise is shown in Figure 3. Its implementation is also compared with a more flexible approach to modelling heteroscedasticity, where an additional Gaussian process prior is placed over the noise (see Appendix B for details). We term this the non-parametric heteroscedastic noise model.

5.2. Results

The performance of the physics-informed kernel mixtures is evaluated against the benchmark models described in Section 4. Predictions of the best performing benchmark model and the proposed approach are shown in Figures 2 and 3. For the Tamar bridge, the capability of models to predict well at high deck responses and wind speeds was of particular interest. This is when the environmental loading of the bridge is highest and when prediction accuracy is most critical, but it is also however where we have the fewest data points. To test and demonstrate extrapolation capability, models were trained on a random scatter of 300 datapoints up a maximum of 30mph wind speeds. This training scenario is designed to showcase how physics-informed kernels can work in the common situation of incomplete data across operational windows. For a full training data regime, one typically wouldn't expect to see any advantage in performance of a physics-informed kernel. In such a case, purely data-based models would naturally be just as capable of capturing the trends.

A summary of kernel structures and a comparison of model performance is shown in Table 1. For conciseness, only deck response plots for selected models are shown in this section, with all other model plots available in Appendix A.1.

The best performing benchmark model was the SE kernel mixture with an NMSE of 31.16% NMSE and MSLL of -0.844. This model contained equivalent switching capabilities to the proposed approach, a composite sigmoidal weighting function acting on wind speed and direction, but did not contain a physically derived kernel. A prediction of deck response for this model is shown in Figure 2. Due to its switching architecture, it is able to identify the changing regimes fairly well. However, due to relying upon stationary SE kernels, the model is not able to extrapolate beyond the 30mph wind speeds observed during training. Instead, the prediction trends back toward the posterior mean and the predictive variance expands to reflect the model's reduced confidence in its prediction.

Table 1: Model performance and computation times for Case Study A: Tamar Bridge wind loading. Additional details of optimisation, hardware and computation cost are provided in Appendix C.

Model	**Kernel structure**	**NMSE (%)**	**MSLL**
SE kernel	k_{SE}	64.42	-0.778
SM kernel	$\sum_q^{Q=3} k_{SE}^q \, k_{Cos}^q$	75.70	-0.369
PI global	$k_{Lift} + k_{SE}$	32.51	-0.825
SE mixture	$w_{comp} k_{SE,1} + (1 - w_{comp}) k_{SE,2}$	31.16	-0.844
PI mixture	$w_{comp} k_{Lift} + (1 - w_{comp}) k_{SE}$	26.10	-0.847
Het. PI mixture	$k_{PI} + \epsilon(u)$	**25.38**	**-1.058**
Het N-P. PI mixture	See Appendix B	29.99	-1.037

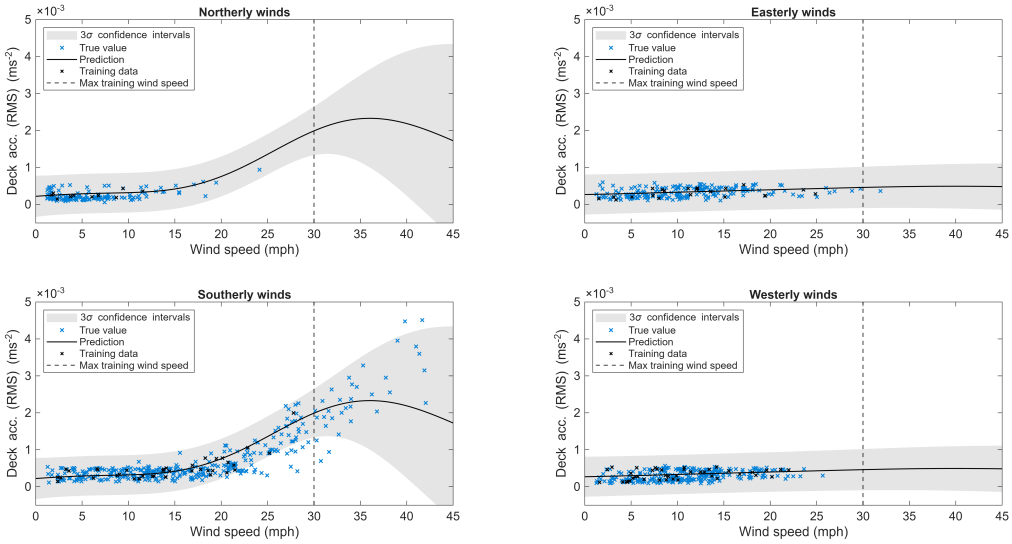

Figure 2: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using an SE kernel mixture. The model is able to identify the regime-dependent behaviour, but is unable to extrapolate effectively beyond the 30mph maximum wind speeds observed during training.

The best performing model was the physics-informed kernel mixture with linearly scaling heteroscedastic noise, achieving an NMSE of 25.38% and a MSLL of -1.058. The influence of the physically derived lift kernel can be seen in Figure 3. Unlike the SE kernel mixture, the model is able to extrapolate beyond the maximum wind speeds seen during training. In this region, without access to measured data, the model relies upon the quadratic approximation of lift force. Importantly, due to the switching architecture, it is not introduced in to the easterly or westerly winds. For the global physics informed model however, this was not the case, resulting in increased model errors. For example, the prediction

of deck acceleration for easterly and westerly winds was pulled upwards by the presence of high speed southerly training data. Although these points do occupy different places in the input space $[u, \theta]$, the global physical kernel could not effectively separate these relationships. A suspected reason for this is an over estimation for the lengthscale acting on wind angle, preventing predictions varying quickly enough w.r.t this variable.

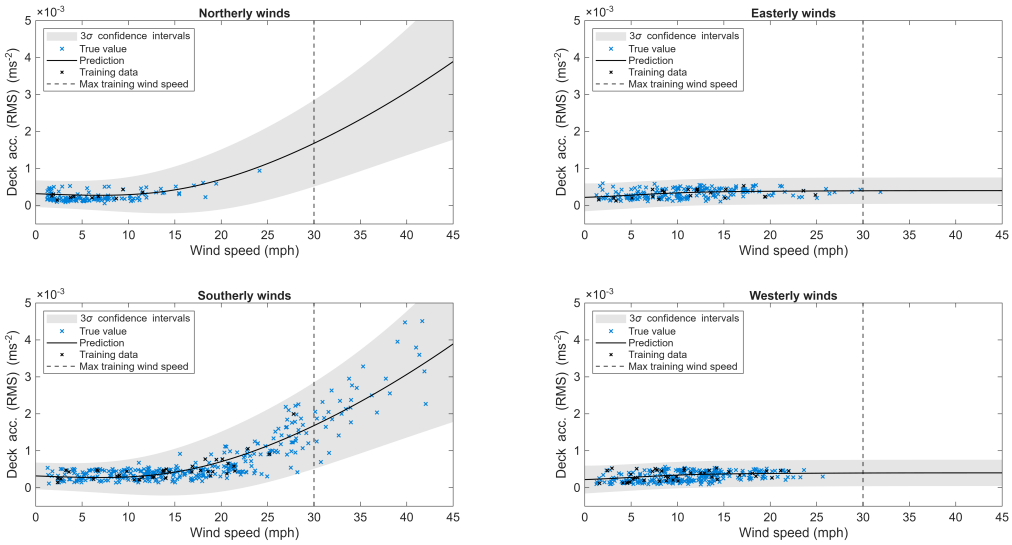

Figure 3: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using a physics informed change-point kernel (Equation 15) with linearly scaling heteroscedastic noise (Equation 16). The model is able to identify regime-dependent behaviour and extrapolate beyond the 30mph maximum wind speeds observed during training. The regime-dependent noise also provides a more realistic capture of uncertainty, decreasing model confidence when lift is induced.

Along with a change in mean signal behaviour, introduced via a quadratic lift force, there is a growing noise level when lift is induced. Here, as wind speed increases, so too does the magnitude of aerodynamic fluctuations and turbulence-induced vibrations. Models utilising a global noise variance are unable to capture this effectively, either over estimating the noise level at lower wind speeds, underestimating the noise during lift, or a compromise of both. The two heteroscedastic models are better able to reflect the underlying confidence in model predictions, including our trust within the physical lift approximation. A quadratic approximation of lift (Equation 12) is based on quasi-steady pressure differences, neglecting unsteady flow and local pressure fluctuations [58]. These differences are amplified at higher wind speeds. Based on this physical intuition of how model noise should behave, embedding this within the kernel structure enforced desirable behaviour without the additional flexibility of a fully non-parametric heteroscedastic approach. When available, an awareness of where physical approximations remain valid can be a valuable resource to encode within a model structure.

The prediction of the non-parametric heteroscedastic noise model is shown in Figure 4. This model was able to discover the growing noise variance with wind speed during lift, without additional physical insight encoded in the noise process. However, overall, the model performs worse than the linearly scaling noise of the kernel mixture. At lower data quantities, such as the test scenario in the case study, the flexibility of the non-parametric heteroscedastic model makes it challenging to fit both the changepoint and a flexible noise model, which results in inflated uncertainty bounds compared to the linear noise model. For example, at high easterly and westerly winds, where no lift was generated the non-parametric heteroscedastic noise expands greatly. Given the training data, and the detection of growing noise with wind speed, this is an expected outcome. Without higher speed wind data (>30mph), or an enforced regime switching noise (the linear noise model) there is little to model behaviour in these regions. An expansion of uncertainty in low data areas is generally a desirable property of a model to avoid overfitting, however physical knowledge can provide a valuable resource to overcome high uncertainty in a meaningful way, even in a reduced training data regime.

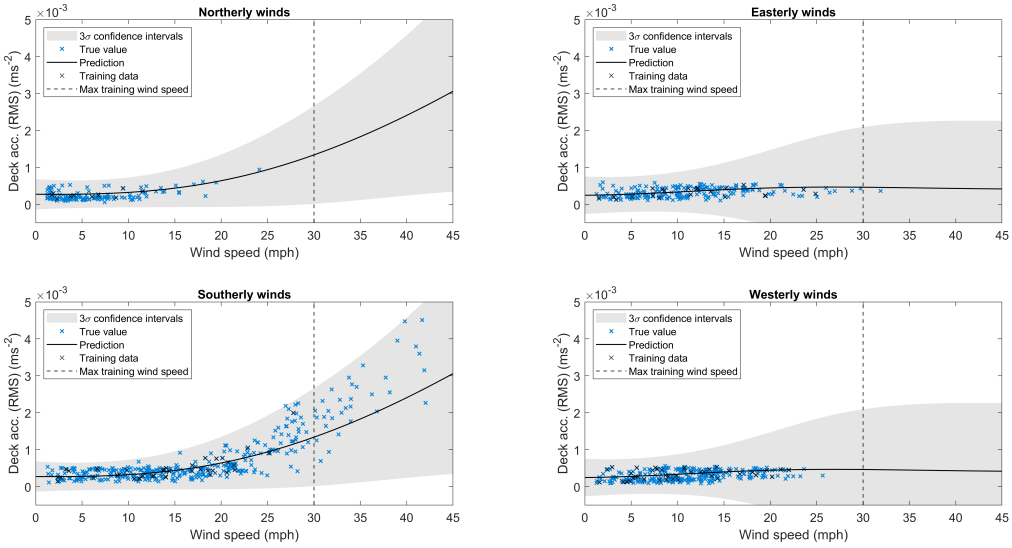

Figure 4: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using a physics informed change-point kernel with fully non-parametric heteroscedastic noise. The model is able to identify regime-dependent behaviour and extrapolate beyond the 30mph maximum wind speeds observed during training. The regime-dependent noise is more flexible than the enforced linear scaling of the proposed approach in Figure 3.

Missing data for specific regimes is a common phenomena within many engineering datasets due to high variability of environments, rarity of events and other factors. For the section of the dataset used for this work, there was a lack of high speed northerly wind data, making predictions in this region challenging. In the UK, the prevailing wind direction is typically south-westerly, causing an increased likelihood to encounter higher wind speeds in this direction [59]. This is further exaggerated near the

south-west coast, where the Tamar bridge is located. The rarity of higher speed northerly winds does not make a models' performance during their occurrence any less important and an effective model should be able to cope with reduced data in this region. The choice of the switching architecture was important here; opting to use a sigmoidal weight (Equation 6) on $\cos(2\theta)$ rather than a local weight (Equation 5) directly on θ. This performed two roles: to enforce symmetry within the kernel, utilising high speed southerly wind data to inform northerly predictions and to avoid complications around the $0° - 360°$ threshold.

A point of discussion relevant to kernel design is the number of hyperparameters within the model. Typically, the construction of more complex kernels comes at the expense of the introduction of additional hyperparameters, therefore increasing computational demands for optimisation. Here, the number of hyperparameters grows from three for the SE kernel, to eleven for the physics-informed change-point kernel leading to an increase in computation time. To alleviate this, several of the introduced hyperparameters were bounded during the optimisation. For example, the switching location of the wind speed sigmoid x_{w0} was bounded between 5 and 30 mph. The use of physical knowledge to reduce optimiser search space is one of the advantages of hyperparameters with physical meaning.

Another common motivation for PIML methods is an ability to provide additional interpretability within results. This is achieved here through the physically informative hyperparameters of the sigmoid kernels. Plots displaying the sigmoids learned from the data, constructed using these hyperparameters, are shown in Figure 5. These plots are able to intuitively display, in terms of where and how quickly, how the relationships modelled within the data change. Here, the model was able to learn from the data that changing direction of wind rapidly introduced lift force as wind approached northerly and southerly directions ($\cos(2\theta) \to 1$). The dependency of the relationship on wind speed was more gradual, seen within the shallower gradient of the learned wind speed sigmoid. The yellow region of the 2d sigmoid surface highlights a region of high speed northerly and southerly winds in which the lift force kernel has the highest explanatory power within the model.

6. Case study B: Changing physical behaviour of an aircraft wing

The second case study focuses on the prediction of aircraft wing strain during aerial manoeuvres and how a changing physical relationship is induced due to the inputs of the pilot. Reed [60] previously achieved success modelling wing strain with MLP networks, resulting in an ability to estimate fatigue damage for the wing within $\pm 3\%$. Later work of Fuentes [61] and Gibson [62] on the dataset used GP regression to enable access to a full predictive distribution over the strains. Gibson explored how draws from the predictive distribution may be propagated to achieve distributions over the fatigue damage.

In this work we will illustrate the proposed method using a dataset from an aircraft in flight which

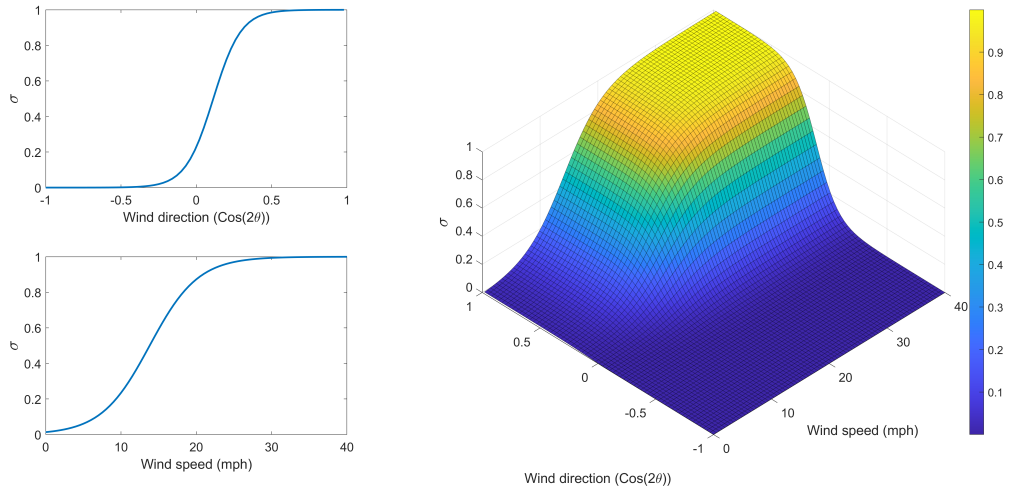

Figure 5: Learned sigmoid functions for the wind direction (top left) and wind speed (bottom left) from the physics-informed change-point kernel. The 2D sigmoid surface (right) across both wind direction and speed highlights the region in yellow in which the physics-based kernel becomes active. This represents a region of northerly/southerly winds at high speeds.

includes time histories of strain measurements and pilot controls (see [60, 61, 62]). Models predict wing strain from a subsection of the dataset, where the pilot used an input to the rudder to cause the aircraft to turn. A time series of the wing strain and rudder angle is shown in Figure 6. At high rudder angles, an aircraft will turn about its Z axis, and the introduction of some additional dynamic behaviour can be seen. A comparison of the induced dynamic behaviours across the three aerial manoeuvres is shown in Figure 7. A similar frequency content, with peaks at $11Hz$ and $32.5Hz$ is shared between the three turns.

6.1. Kernel design

The wing strain is considered to consist of two main parts: a quasi-static component, capturing the more slowly varying relationship between flight parameters and strain; and a dynamic component, induced by oscillations occurring during manoeuvres. In this case, the manoeuvre of interest is a sharp turn due to a high rudder angle. To represent this, the problem is structured as:

$$y(X) = \underbrace{f_1(X)}_{\text{Quasi-static component}} + \underbrace{f_2(X)}_{\text{Dynamic, manoeuvre-induced behaviour}} + \underbrace{\epsilon}_{\text{Noise}} \tag{17}$$

To capture the relationship between measured flight parameters and the wing strain, in line with previous kernel selections within the literature [61, 62], we employ the Squared Exponential (SE) kernel:

$$k_{SE}(X_{Fl}, X'_{Fl}) = \sigma_f^2 \exp\left(-\frac{1}{2}(X_{Fl})^T \Lambda^{-1}(X'_{Fl})\right) + \sigma_n^2 \delta_{ij} \tag{18}$$

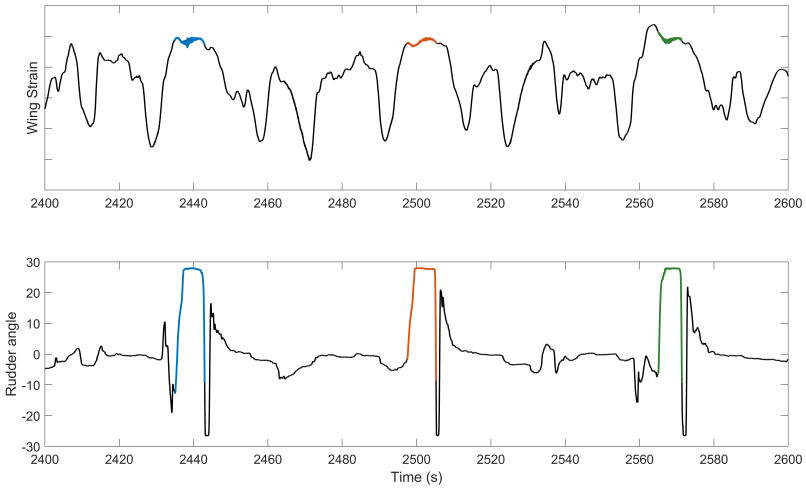

Figure 6: Time series plots of wing strain and rudder angle. Three consecutive, aerial turns are highlighted in blue, orange and green.

where σ_f^2 is the signal variance, σ_n^2 is the noise variance, Λ is the matrix of length scales such that $diag(\Lambda) = [l_1^2, l_2^2, ..., l_D^2]$ for a D dimensional input and X_{Fl} is a matrix of measured flight parameters.

To capture the oscillatory behaviour, highlighted in Figure 7, a covariance function is employed that can model vibrating systems and has been directly derived from an equation of motion of a stochastic vibrating system (SDOF kernel [17]):

$$k_{SDOF}(\tau) = \frac{\sigma^2}{4m^2\zeta\omega_n^3} e^{-\zeta\omega_n|\tau|}\left(cos(\omega_d\tau) + \frac{\zeta\omega_n}{\omega_d} sin(\omega_d|\tau|)\right) \tag{19}$$

where the hyperparameters of the kernel now relate to physical properties of a SDOF oscillator: m is the mass, $\zeta = c/2\sqrt{km}$ is the damping ratio, $\omega_n = \sqrt{k/m}$ is the natural frequency and $\omega_d = \omega_n\sqrt{1-\zeta^2}$ is the damped natural frequency. Draws from this kernel are constrained to obey the behaviour of an SDOF oscillator, a useful property to encode, particulary where we are interested in fatigue performance. Here the SDOF kernel aims to capture the oscillatory behaviour of the wing, similarly to how previous implementations of the kernel have been used to represent modes of an oscillating cantilever beam [63]. Due to the presence of two peak frequencies, believed to be caused by two modes of the wing being excited, it is necessary to use a sum of SDOF kernels:

$$k_{MDOF} = \sum_{i=1}^{N} k_{SDOF}^i(t, t') \tag{20}$$

where, here, $N = 2$. A feature of the dataset, and one of the primary reasons it has been selected for this case study, is that the dynamic behaviour is intermittent. To allow the SDOF kernel to be phased

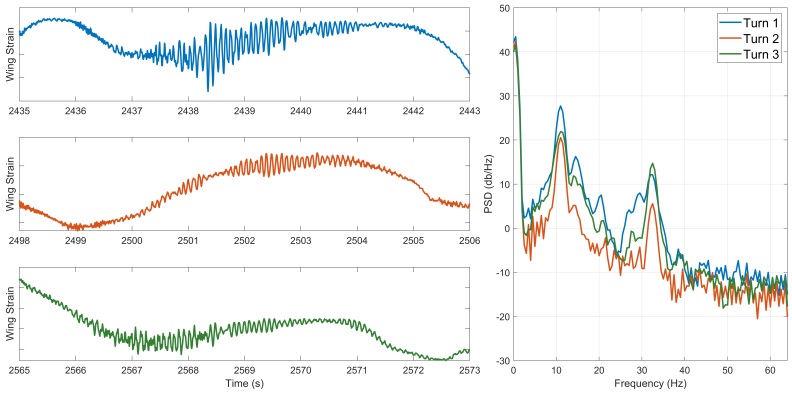

Figure 7: A comparison of the time series and frequency content of the three sharp aerial turns from Figure 6. The same colours: blue, orange and green are used for each turn. Two peak frequencies of $11Hz$ and $32.5Hz$ are observed across the three turns.

in and out appropriately, it is integrated in a change-point structure of the form:

$$k(X, X') = \underbrace{k_{SE}(X_{Fl}, X'_{Fl})}_{\text{Quasi-static signal component captured with an SE kernel...}} + \underbrace{k_{\sigma}(\text{RUDD}, \text{RUDD}') \left(\sum_{i=1}^{N} k^{i}_{SDOF}(t, t') \right)}_{\text{... whilst introducing the SDOF kernel(s) for rudder-based manoeuvres.}} \tag{21}$$

where X_{Fl} is a matrix of flight parameters, RUDD is the rudder angle and t is the time vector. Note an important difference here compared with the Tamar Bridge kernel construction in Equation 15; rather than switching to and from a pair of kernels (k_{Lift} and k_{SE} for the Tamar case study), the SE kernel k_{SE} remains active and the physics-informed kernel k_{SDOF} is being phased in and out. The most appropriate way to utilise sigmoid kernels, and therefore define how the kernel may switch, will vary by use case.

6.2. Results

A focus of the earlier Tamar Bridge case study was the utilisation of physical knowledge to assist with extrapolation and heteroscedastic noise, here, for the second case study, an emphasis is instead placed on interpolation and dynamic behaviour. A key strength of the SDOF kernel is an ability to help with the upsampling of response data [17]. To test this capability, models were trained on data at 64Hz and tasked with upsampling to 128Hz. The three manoeuvres, highlighted in Figure 6, took place over a time period of 200 seconds, resulting in a test set size of 25,600 datapoints and a training set size of 12,800 datapoints. For purposes of model comparison, without additional variation induced

by approximate inference schemes, exact inference is used for the models in the case study, however it is acknowledged that a sparse implementation would be more practical for datasets of this size. The performance of models is compared in Table 2 (NMSE) and 3 (MSLL). Details of hardware and computation times are provided in Appendix C.

Table 2: Predictive accuracy (NMSE (%)) comparison for Case Study B: Aircraft wing strain during in-flight manoeuvres.

Model	**Kernel structure**	**Full series**	**Turn 1**	**Turn 2**	**Turn 3**
SE kernel	k_{SE}	0.01380	22.851	1.876	5.413
SM kernel	$\sum_q^{Q=3} k_{\text{SE}}^q \, k_{\text{Cos}}^q$	0.00881	4.613	0.354	1.579
PI global	$k_{\text{SE}} + \sum_i^{N=2} k_{\text{SDOF}}^i$	0.01340	22.860	1.859	5.404
SE mixture	$k_{\text{SE}} + w_{\text{sig}} \, k_{\text{SE}}$	0.01160	5.859	0.445	2.022
PI mixture	$k_{\text{SE}} + w_{\text{sig}} \sum_i^{N=2} k_{\text{SDOF}}^i$	**0.00749**	**2.944**	**0.301**	**0.872**

Table 3: Probabilistic performance (MSLL) comparison for Case Study B: Aircraft wing strain during in-flight manoeuvres.

Model	**Kernel structure**	**Full series**	**Turn 1**	**Turn 2**	**Turn 3**
SE kernel	k_{SE}	-4.240	0.255	-1.968	-1.484
SM kernel	$\sum_q^{Q=3} k_{\text{SE}}^q \, k_{\text{Cos}}^q$	-4.453	-1.525	-2.101	-1.998
PI global	$k_{\text{SE}} + \sum_i^{N=2} k_{\text{SDOF}}^i$	-4.290	0.257	-1.971	-1.485
SE mixture	$k_{\text{SE}} + w_{\text{sig}} \, k_{\text{SE}}$	-4.461	-1.499	-1.947	-1.881
PI mixture	$k_{\text{SE}} + w_{\text{sig}} \sum_i^{N=2} k_{\text{SDOF}}^i$	**-4.615**	**-1.759**	**-2.107**	**-2.095**

Across the full duration of the time series, all candidate models demonstrated a high level of predictive fidelity, reflected in the low NMSEs ($< 0.014\%$) and MSLLs (< -4.240). This indicates that even the baseline architectures capture the primary quasi-static trends effectively. As a task of interpolation, one would expect a GP to perform well here, even without any inclusion of prior knowledge. The primary areas of interest, and where model performance deviates most, is when dynamic behaviour is induced due to the turning of the aircraft. In terms of overall error, the performance during these manoeuvres does not greatly impact the overall time series, however for fatigue analysis, the estimation of peak loads (e.g. during turns) and identification of loading cycles is of high importance.

The prediction of the physics-informed kernel mixture across the full timeseries had an NMSE of 0.00749%, beating the 0.00881% NMSE of the best performing benchmark model (spectral mixture kernel). A large contributor to this was an improved ability to capture the dynamic content through the phasing in of SDOF kernels during turning manoeuvres. This is highlighted by the prediction plots

in Figures 8 and 9. The NMSE for turns 1, 2 and 3 dropped from 4.613%, 0.354% and 1.579% for the SM kernel to 2.944%, 0.301% and 0.872%, with the physically-informed change-point kernel better able to capture the structure present within the signal. Both kernel structures are inherently well suited to modelling periodic structure present within signals, however they differ in their methodology. The SM kernel is designed to perform automated 'pattern discovery' without requiring prior knowledge of a system's physical properties [54]. Although it may enforce locality of behaviour to some extent, via multiplicative k_{SE} kernel components, it does not involve an explicit switching mechanism. Further work pursuing the integration of switching capabilities in this context could be an interesting avenue.

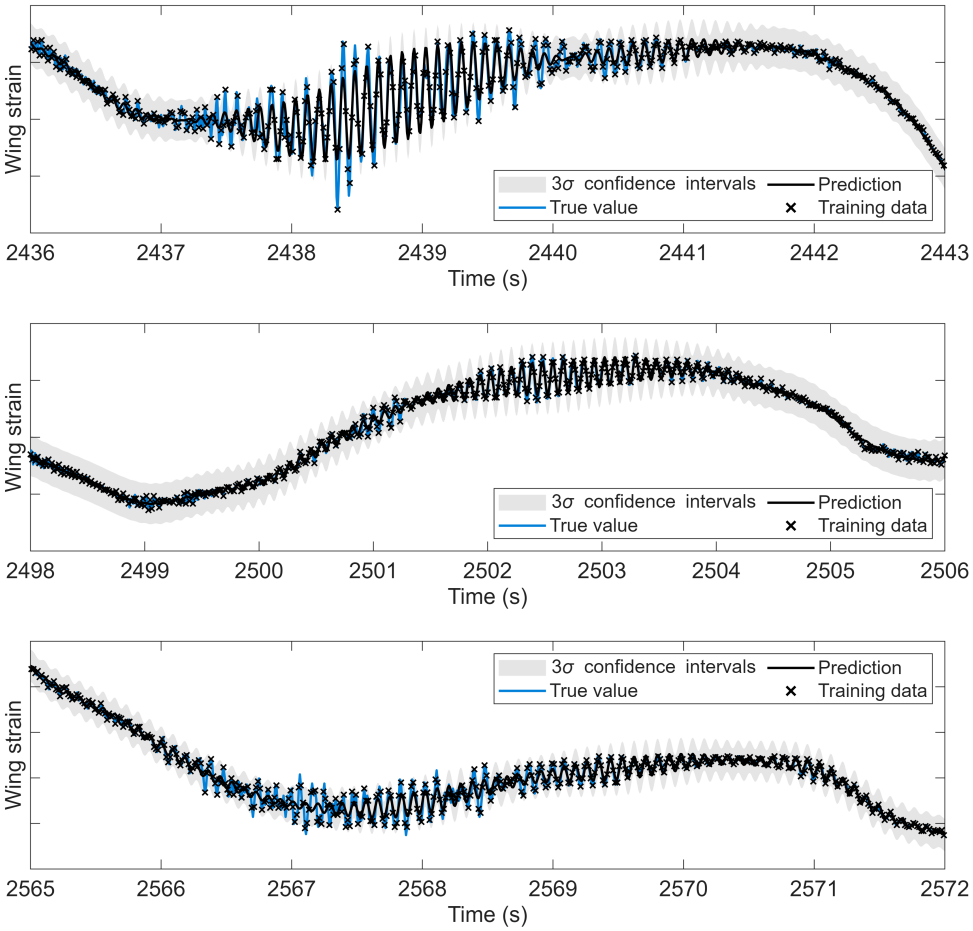

Figure 8: Predictions of inner port side wing strain during the three in-flight turns using a spectral mixture kernel with 3 components.

The benefits of both the switching architecture and the physically derived SDOF kernel can be isolated through a comparison of the proposed approach with the global physics-informed model and the SE

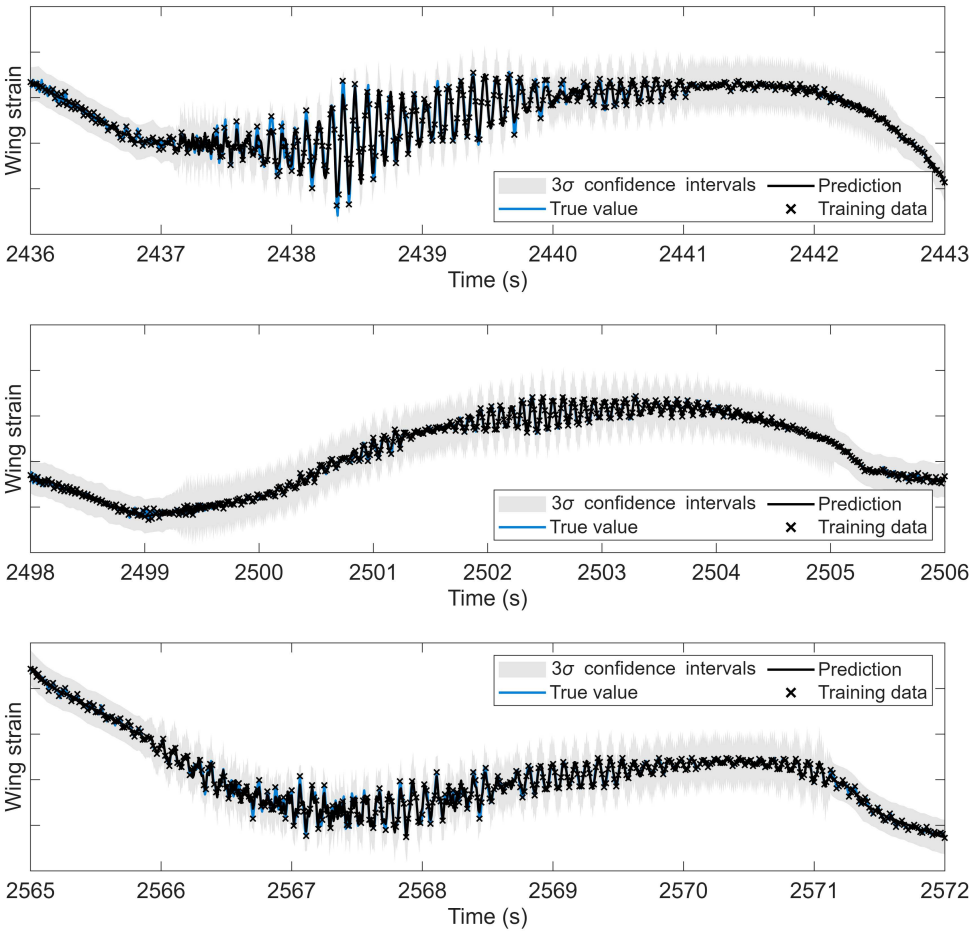

Figure 9: Predictions of inner port side wing strain during the three in-flight turns using a physics-informed change-point kernel.

kernel mixture. These models contain either SDOF kernels or a switching mechanism, but not both. In the case of the SE kernel mixture, the model is able to switch to a lower length scale during high rudder angles and is better able to capture the higher frequency behaviour present in the signal. As expected, this performed better than a single stationary SE kernel, but was worse than either the spectral mixture or the proposed approach. SE kernels, even with short lengthscales, are not derived specifically for modelling periodic signals. The performance of the global physics-informed kernel was much closer to that of a standard SE kernel than any of the switching approaches, with the SDOF kernel failing to capture the high frequency content during manoeuvres. Without a switching mechanism, the fit of the SDOF kernel to the whole timeseries is calculated during training. Since the manouvres represent only a small percentage of the time series, low amplitude, low frequency fits of the kernel produce

low log loss minima. To alleviate these problems, it is possible to bound or constrain the physical hyperparameters of the kernel e.g. analyse the frequency content of a signal prior to modelling and fix the ω_n. However, since high frequency oscillations were not desirable during the majority of the signal, this could worsen overall performance for the global approach.

6.3. Model interpretability

As also discussed in the Tamar Bridge case study, another notable benefit from PIML models alongside increased performance is an additional degree of insight in to model results. For the change-point kernel, this came from two main sources: the hyperparameters of the weighting functions, providing information about how the model was switching between kernel components, and the hyperparameters of the SDOF kernel, relating to physical properties of an SDOF oscillator. Here, we look to interpret the model's switching behaviour, with plots constructed using the learned sigmoid kernel hyperparameters shown in Figure 10. The degree of interpretability is inherently tied to the choice of weighting functions and kernels. In this example, the location and gradient terms of the sigmoid function determine the specific rudder angle and speed at which dynamic behaviour is introduced.

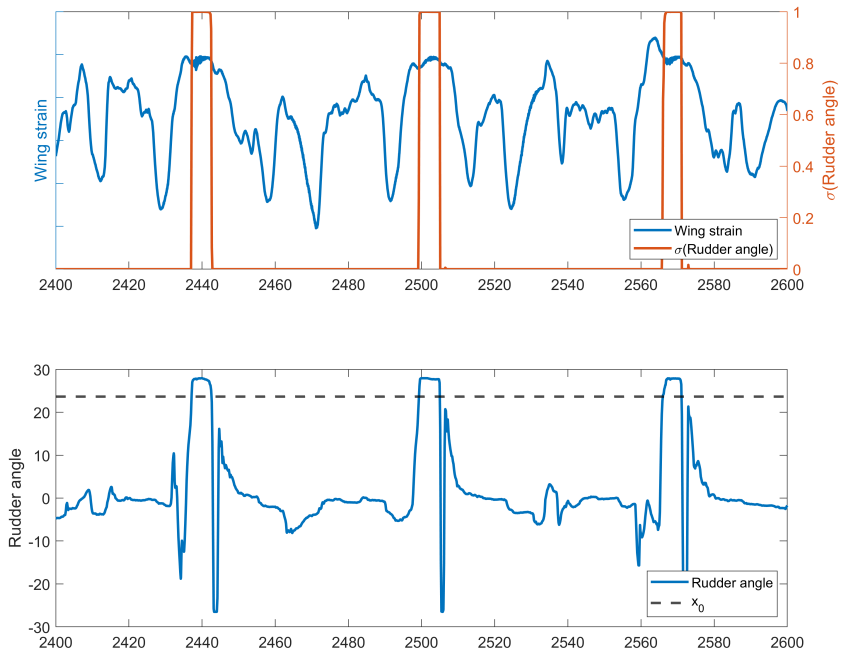

Figure 10: Interpretation of model switching using the learned sigmoid kernel hyperparameters. (Top) A time series plot of the rudder angle passed through the sigmoid function, σ(Rudder angle). The 'active' region is observed to coincide with the oscillatory behaviour seen in the wing strain. (Bottom) A time series plot of the rudder angle, showing where it exceeds the sigmoid switching location x_0.

The top subplot of Figure 10 shows the reconstructed sigmoid function from the learned hyperparameters $x_0 = 22.59°$ and $a = 3.099$. Within the change-point structure, this produces a switching of the SDOF kernel from fully inactive, to fully active within in a window of approximately $21 - 24°$ rudder angle. The relatively steep gradient of the sigmoid, combined with a rapid change of rudder angle during the in-flight manoeuvre, produces a sudden switch within the model. An almost step function like time series plot of the rudder angle passed through the sigmoid function $\sigma(\text{RUDD})$ can be seen within the bottom subplot of Figure 10. For the modelling of in-flight manoeuvres, one might expect a fairly sudden switching of behaviour, however for any application, similar plots to this can provide a useful way to interpret when and how different parts of the model are being phased into or out of predictions.

7. Summary and conclusions

This paper presented a physics-informed kernel mixture framework for regime-dependent reliance upon physical knowledge in Gaussian process regression. Building upon existing non-stationary and change-point kernel architectures, that primarily capture statistical differences in signals, the proposed approach uses switching mechanisms to model and intepret the reliance upon physically-derived kernels based on environmental switching variables. This provides a promising tool to cope with both regime-switching and varied presence of known or partially known physical behaviours. Two engineering case studies were used to highlight how the model structure can be used to improve performance and provide insight in to a models' results.

A dataset of directional wind loading of the Tamar bridge was used to highlight a case where physical knowledge of a phenomena may only apply in particular regimes. The lift force induced by high wind speed, represented here with a polynomial kernel, only acted when wind hit the bridge side-on. The change-point kernel was able to learn this and present results of the changing regimes in an interpretable manner. The model was able to extrapolate for regions of high speed northerly winds, for which a scarcity of measured data was available. Additionally, it was shown how the change-point kernel may be used in a heteroscedastic setting, where observation noise is a function of the inputs; in this case, wind speed, providing better fitted confidence bounds.

Data from an aircraft was used to investigate oscillatory behaviour within the wings, induced during in-flight manoeuvring. The behaviour was found to occur during sharp turns of the aircraft (high rudder angles), and was represented using a SDOF kernel within the change point structure. The model was able to effectively upsample from 64Hz to 128Hz and provide a breakdown of static and dynamic signal components. Compared with a purely data-based approach, model performance during three aerial manoeuvres was improved from 22.851%, 1.876% and 5.413% to 2.944%, 0.301% and 0.872% respectively, with performance across the complete time series improving by almost a factor of two.

The work and contribution here exploits the ease of making combinations in the Gaussian process framework. Here we are explicitly advocating for embedding known physics into kernel functions, where the physics will then play an integral role in model predictions (as opposed to, for example, a soft constraint). The framework shown here then provides a suitable means in which the reliance on this physics can change based on how relevant it is to operational conditions. The paper has shown that this works very well where we indeed have access to or intuition of some of the governing processes at work in the situation we are modelling - of course, this would not be true without that physical insight. The approach here is not intended to be generic, rather to exploit what is known in the right places. In our approach we are reliant, therefore, on having knowledge of some of the governing physics and being able to derive appropriate kernels to represent it. In cases where such insight is not available, future work could explore the integration of automated kernel discovery methods [64] or the use of pattern discovery kernels [54] within the switching framework to handle less-defined physical regimes.

One of the benefits of the implementation here is closed form inference for posterior distributions, which was of benefit in particular in the first test case where we embedded a physics-informed heteroscedastic noise. Staying closed form, however, for the majority of GP models, requires a restriction to small datasets. To extend the applicability of this framework to large-scale SHM datasets, the adoption of sparse GP approximations and variational inference techniques remains a necessary and promising avenue for future research.

Acknowledgements

The authors gratefully acknowledge the support of The University of Sheffield and Innovate UK through the OLLGA project grant 10040817. We would like to thank the Vibration Engineering Section (VES) for their work collecting the Tamar bridge dataset.

References

[1] G. Karniadakis, Y. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nature Reviews Physics (2021) 1–19.

[2] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, M. Walczak, J. Pfrommer, A. Pick, R. Ramamurthy, J. Garcke, C. Bauckhage, J. Schuecker, Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems, IEEE Transactions on Knowledge and Data Engineering (2021).

[3] E. J. Cross, S. J. Gibson, M. R. Jones, D. J. Pitchforth, S. Zhang, T. J. Rogers, Physics-informed

machine learning for structural health monitoring, in: Structural Health Monitoring Based on Data Science Techniques, Springer, 2022, pp. 347–367.

[4] E. J. Cross, T. J. Rogers, D. J. Pitchforth, S. J. Gibson, S. Zhang, M. R. Jones, A spectrum of physics-informed Gaussian processes for regression in engineering, Data-Centric Engineering 5 (2024) e8.

[5] M. Haywood-Alexander, W. Liu, K. Bacsa, Z. Lai, E. Chatzi, Discussing the spectrum of physics-enhanced machine learning: a survey on structural mechanics applications, Data-Centric Engineering 5 (2024) e31.

[6] A. Cicirello, Physics-enhanced machine learning: a position paper for dynamical systems investigations, Journal of Physics: Conference Series 2909 (2024) 012034.

[7] D. J. Pitchforth, T. J. Rogers, U. T. Tygesen, E. J. Cross, Grey-box models for wave loading prediction, Mechanical Systems and Signal Processing 159 (2021) 107741.

[8] W. Wan, B. Zhang, The intermittent leakage phenomenon of incipient cracks under transient conditions in pipeline systems†, International Journal of Pressure Vessels and Piping 186 (2020) 104138.

[9] V. Belenky, K. M. Weems, C. C. Bassler, M. J. Dipper, B. L. Campbell, K. J. Spyrou, Approaches to rare events in stochastic dynamics of ships, Probabilistic Engineering Mechanics 28 (2012) 30–38.

[10] S. Xue, G. Xu, W. Xie, L. Xu, Z. Jiang, Characteristics of freak wave and its interaction with marine structures: A review, Ocean Engineering 287 (2023) 115764.

[11] S. Pfingstl, M. Zimmermann, On integrating prior knowledge into Gaussian processes for prognostic health monitoring, Mechanical Systems and Signal Processing 171 (2022) 108917.

[12] D. An, N. H. Kim, J.-H. Choi, Practical options for selecting data-driven or physics-based prognostics algorithms with reviews, Reliability Engineering & System Safety 133 (2015) 223–236.

[13] D. Kong, Y. Chen, N. Li, Gaussian process regression for tool wear prediction, Mechanical Systems and Signal Processing 104 (2018) 556–574.

[14] T. Rogers, P. Gardner, N. Dervilis, K. Worden, A. Maguire, E. Papatheou, E. Cross, Probabilistic modelling of wind turbine power curves with application of heteroscedastic Gaussian process regression, Renewable Energy 148 (2020) 1124–1136.

[15] C. E. Rasmussen, C. K. I. Williams, Gaussian processes for Machine Learning, The MIT Press, 2005.

[16] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions, Journal of Mathematical Psychology 85 (2018) 1–16.

[17] E. J. Cross, T. J. Rogers, Physics-derived covariance functions for machine learning in structural dynamics, IFAC-PapersOnLine 54 (2021) 168–173, 19th IFAC Symposium on System Identification SYSID 2021.

[18] M. Kok, A. Solin, Scalable magnetic field slam in 3d using Gaussian process maps, in: 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 1353–1360.

[19] M. M. Noack, J. A. Sethian, Advanced stationary and nonstationary kernel designs for domain-aware Gaussian processes, Communications in Applied Mathematics and Computational Science 17 (2022).

[20] S. Klus, P. Gelß, F. Nüske, F. Noé, Symmetric and antisymmetric kernels for machine learning problems in quantum physics and chemistry, Machine Learning: Science and Technology 2 (2021).

[21] C. R. Farrar, K. Worden, Structural Health Monitoring: A Machine Learning Perspective, John Wiley & Sons, 2012.

[22] Z. Tan, Y. Wu, On regime switching models, Mathematics 13 (2025).

[23] J. D. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57 (1989) 357–384.

[24] J. D. Hamilton, Time series analysis, Princeton university press, 1994.

[25] Z. Psaradakis, M. Sola, Markov-switching models with state-dependent time-varying transition probabilities, Econometrics and Statistics 29 (2024) 49–63.

[26] S. Chiappa, Explicit-duration markov switching models, Foundations and Trends in Machine Learning 7 (2014) 803–886.

[27] H. Tong, Threshold Models in Non-linear Time Series Analysis, Springer, 1983.

[28] H. Tong, Threshold models in time series analysis-30 years on, Statistics and Its Interface 4 (2011) 107–118.

[29] T. Teräsvirta, Specification, estimation, and evaluation of smooth transition autoregressive models, Journal of the American Statistical Association (1994).

[30] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, G. E. Hinton, Adaptive mixtures of local experts, Neural Computation 3 (1991) 79–87.

[31] S. E. Yuksel, J. N. Wilson, P. D. Gader, Twenty years of mixture of experts, IEEE Transactions on Neural Networks and Learning Systems 23 (2012) 1177–1193.

[32] A. H.-S. Ang, W. H. Tang, Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering, John Wiley & Sons, 2006.

[33] V. Tresp, Mixtures of gaussian processes, in: Advances in Neural Information Processing Systems (NIPS), 2000, pp. 654–660.

[34] C. E. Rasmussen, Z. Ghahramani, Infinite mixtures of gaussian process experts, in: Advances in Neural Information Processing Systems (NIPS 2001), Max Planck Institute for Biological Cybernetics, 2002.

[35] T. Härkönen, S. Wade, K. Law, L. Roininen, Mixtures of gaussian process experts with smc2, Journal of Machine Learning Research 26 (2025) 1–38.

[36] A. G. Wilson, R. P. Adams, Gaussian process change point models, in: International Conference on Machine Learning (ICML), 2013.

[37] D. Garreau, S. Arlot, Consistent change-point detection with kernels, Electronic Journal of Statistics 12 (2018) 4440–4486.

[38] W. Herlands, A. Wilson, H. Nickisch, S. Flaxman, D. Neill, W. Van Panhuis, E. Xing, Scalable gaussian processes for characterizing multidimensional change surfaces, in: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, Vol. 51 of Proceedings of Machine Learning Research, PMLR, 2016, pp. 1013–1021.

[39] S. Arlot, A. Celisse, Z. Harchaoui, A kernel multiple change-point algorithm via model selection, Journal of Machine Learning Research 20 (2019) 1–56.

[40] V. Volodina, D. Williamson, Diagnostics-driven nonstationary emulators using kernel mixtures, SIAM/ASA Journal on Uncertainty Quantification 8 (2020) 1–26.

[41] B. Peeters, G. De Roeck, One-year monitoring of the z24-bridge: environmental effects versus damage events, Earthquake Engineering & Structural Dynamics 30 (2001) 149–171.

[42] M. A. Álvarez, J. Peters, B. Schölkopf, N. D. Lawrence, Switched latent force models for movement segmentation, in: Advances in Neural Information Processing Systems, Vol. 23, 2010, pp. 55–63.

[43] L. Marino, A. Cicirello, A switching gaussian process latent force model for the identification of mechanical systems with a discontinuous nonlinearity, Data-Centric Engineering 4 (2023) e18.

[44] K. Worden, E. Cross, On switching response surface models, with applications to the structural health monitoring of bridges, Mechanical Systems and Signal Processing 98 (2018) 139–156.

[45] D. Duvenaud, H. Nickisch, C. E. Rasmussen, Additive gaussian processes, in: Proceedings of the 25th International Conference on Neural Information Processing Systems, 2011, pp. 226–234.

[46] N. Chalapathi, Y. Du, A. Krishnapriyan, Scaling physics-informed hard constraints with mixture-of-experts, in: B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, Y. Sun (Eds.), International Conference on Learning Representations, 2024, pp. 36015–36033.

[47] M. Yuan, J. Liu, Z. Chen, Q. Guo, M. Yuan, J. Li, G. Yu, Predicting energy consumption for hybrid energy systems toward sustainable manufacturing: A physics-informed approach using pi-mmoe, Sustainability 16 (2024).

[48] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics 378 (2019) 686–707.

[49] Z. Xiang, W. Peng, X. Zheng, X. Zhao, W. Yao, Self-adaptive loss balanced physics-informed neural networks for the incompressible navier-stokes equations, arXiv preprint arXiv:2104.06217Adaptive self-balancing PINN loss weights (2021).

[50] C. Zhou, J. Chen, Z. Yang, C. E. Png, Dual-balancing for physics-informed neural networks, arXiv preprint arXiv:2505.11117Adaptive dynamic loss balancing for PINNs (2025).

[51] S. Perez, S. Maddu, I. F. Sbalzarini, P. Poncet, Adaptive weighting of bayesian physics informed neural networks for multitask and multiscale forward and inverse problems, Journal of Computational Physics 491 (2023) 112342, adaptive Bayesian PINN weighting strategies.

[52] D. Kochan, X. Yang, Blending physics with data using an efficient gaussian process regression with soft inequality and monotonicity constraints, Frontiers in Mechanical Engineering 10 (2025) 1410190, adaptive GP regression with physical constraints.

[53] M. Lázaro-Gredilla, M. K. Titsias, Variational heteroscedastic Gaussian process regression., in: ICML, 2011, pp. 841–848.

[54] A. G. Wilson, R. Adams, Gaussian process kernels for pattern discovery and extrapolation, in: Proceedings of the 30th International Conference on Machine Learning, Vol. 28, 2013, pp. 1067–1075.

[55] E. J. Cross, On structural health monitoring in changing environmental and operational conditions, Ph.D. thesis, University of Sheffield (2012).

[56] K. Y. Koo, J. M. W. Brownjohn, D. I. List, R. Cole, Structural health monitoring of the Tamar suspension bridge, Structural Control and Health Monitoring 20 (2013) 609–625.

[57] K. Kerenyi, T. Sofu, J. Guo, Hydrodynamic forces on inundated bridge decks, United States Department of Transportation (2009).

[58] J. D. Holmes, Wind Loading of Structures, 3rd Edition, CRC Press, 2015.

[59] A. Lapworth, J. McGregor, Seasonal variation of the prevailing wind direction in britain, Weather 63 (2008) 365–368.

[60] S. C. Reed, Development of a parametric-based indirect aircraft structural usage monitoring system using artificial neural networks, The Aeronautical Journal 111 (2007) 209–230.

[61] R. Fuentes, E. Cross, A. Halfpenny, K. Worden, R. Barthorpe, Aircraft parametric structural load monitoring using gaussian process regression, in: Proceedings of the European Workshop on Structural Health Monitoring, 2014.

[62] S. J. Gibson, T. J. Rogers, E. J. Cross, Distributions of fatigue damage from data-driven strain prediction using gaussian process regression, Structural Health Monitoring 22 (2023) 3065–3076.

[63] D. J. Pitchforth, T. J. Rogers, U. T. Tygesen, E. J. Cross, Incorporation of partial physical knowledge within gaussian processes, in: Proceedings of the 30th International Conference on Noise and Vibration Engineering (ISMA 2022), 2022.

[64] D. Duvenaud, J. Lloyd, R. Grosse, J. Tenenbaum, G. Zoubin, Structure discovery in nonparametric regression through compositional kernel search, in: Proceedings of the 30th International Conference on Machine Learning, Vol. 28, 2013, pp. 1166–1174.

[65] M. Titsias, Variational learning of inducing variables in sparse gaussian processes, in: Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, Vol. 5, 2009, pp. 567–574.

Appendix A. Prediction plots for all models

Appendix A.1. Case study A: Directional wind loading of the Tamar bridge

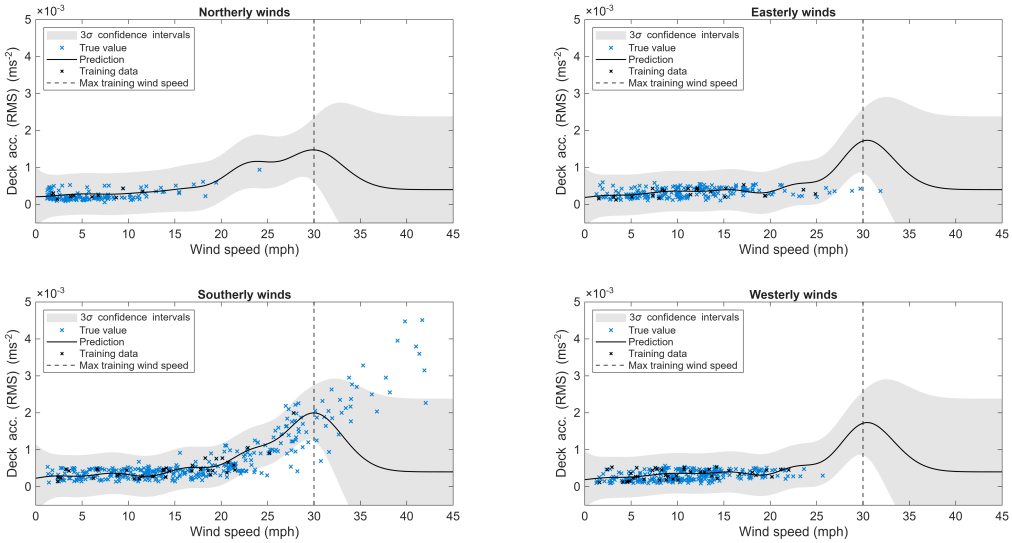

Figure A.11: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using a squared exponential kernel. The model was trained on a random scatter of 300 datapoints up a maximum of 30mph wind speeds

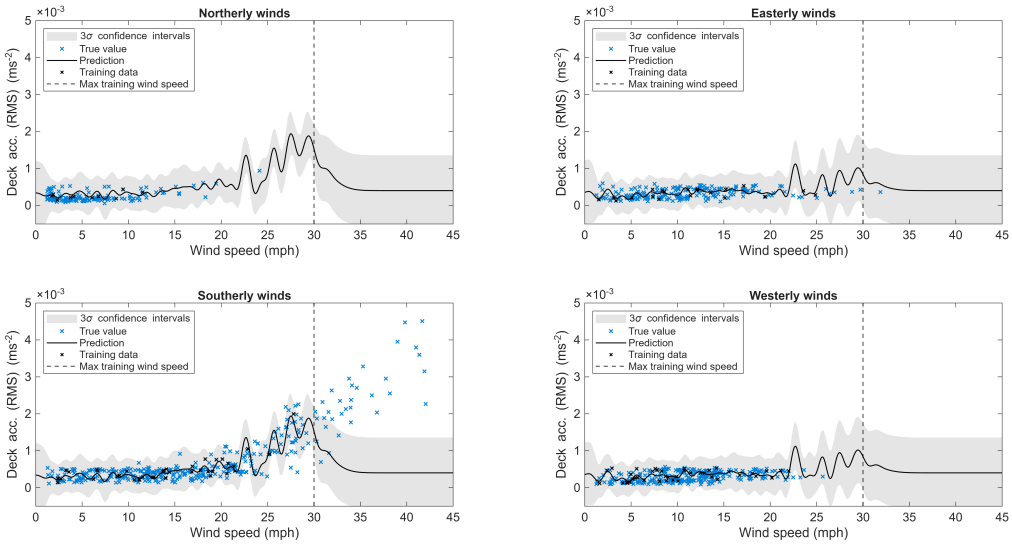

Figure A.12: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using a spectral mixture kernel with 3 components. The model was trained on a random scatter of 300 datapoints up a maximum of 30mph wind speeds

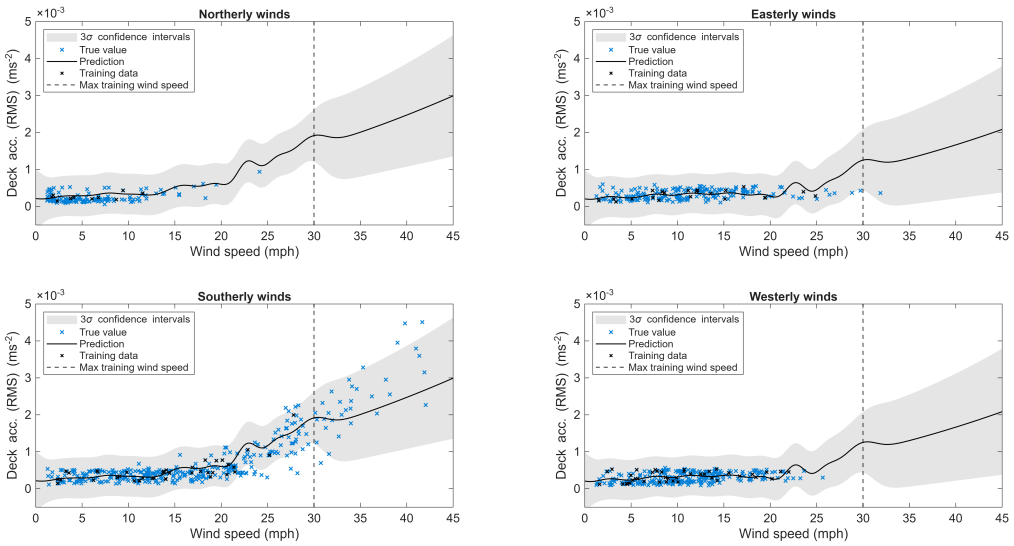

Figure A.13: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using a global physics-informed kernel. The model was trained on a random scatter of 300 datapoints up a maximum of 30mph wind speeds

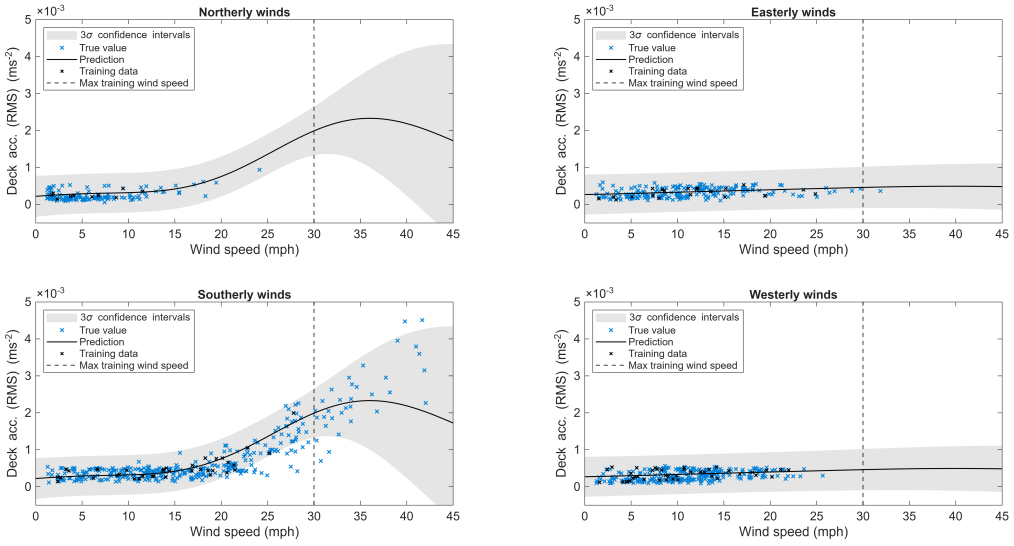

Figure A.14: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using a black-box change-point kernel. The model was trained on a random scatter of 300 datapoints up a maximum of 30mph wind speeds

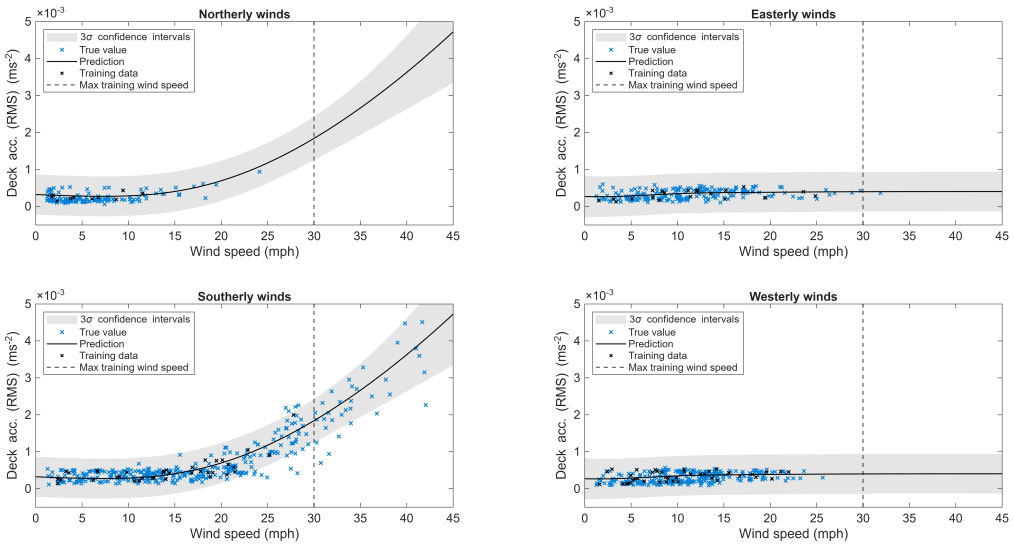

Figure A.15: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using a physics informed change-point kernel with a global noise variance. The model was trained on a random scatter of 300 datapoints up a maximum of 30mph wind speeds

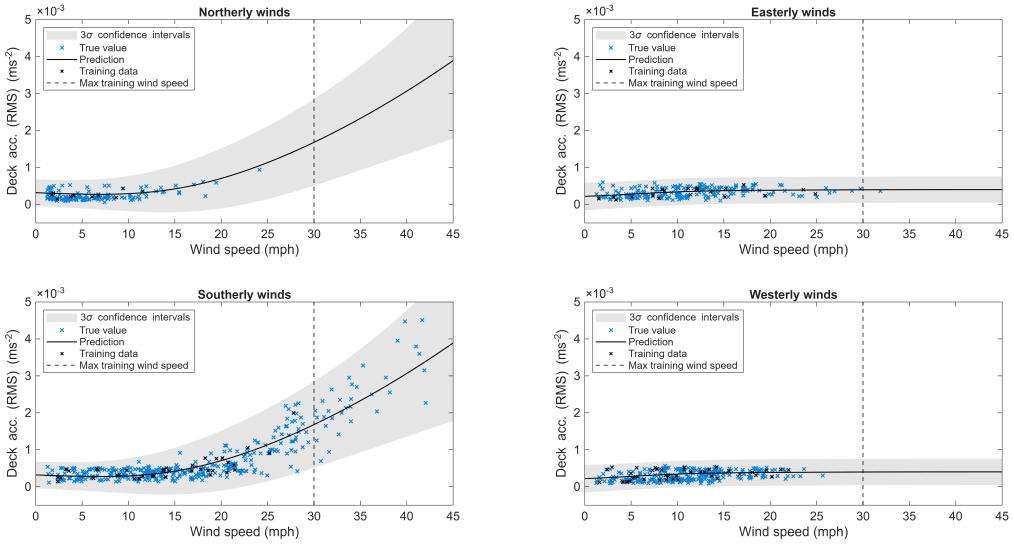

Figure A.16: Predictions of vertical deck acceleration on the Tamar bridge, separated by wind direction, using a physics informed change-point kernel with heteroscedastic regime-dependent noise. The model was trained on a random scatter of 300 datapoints up a maximum of 30mph wind speeds

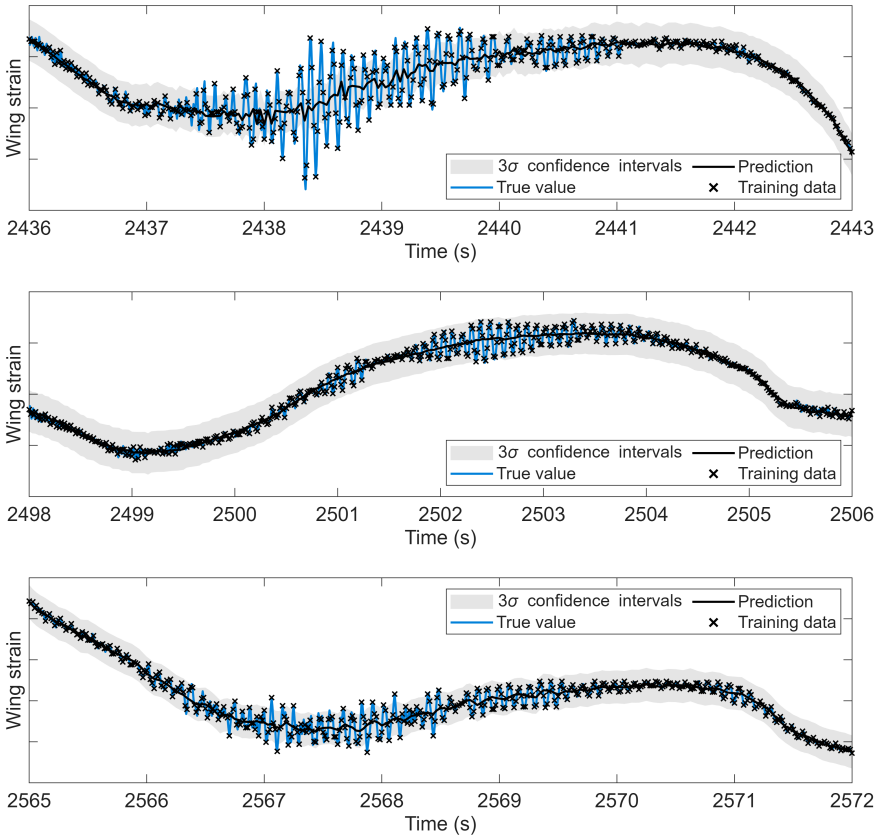

Figure A.17: Predictions of inner port side wing strain during the three in-flight turns using a squared exponential kernel.

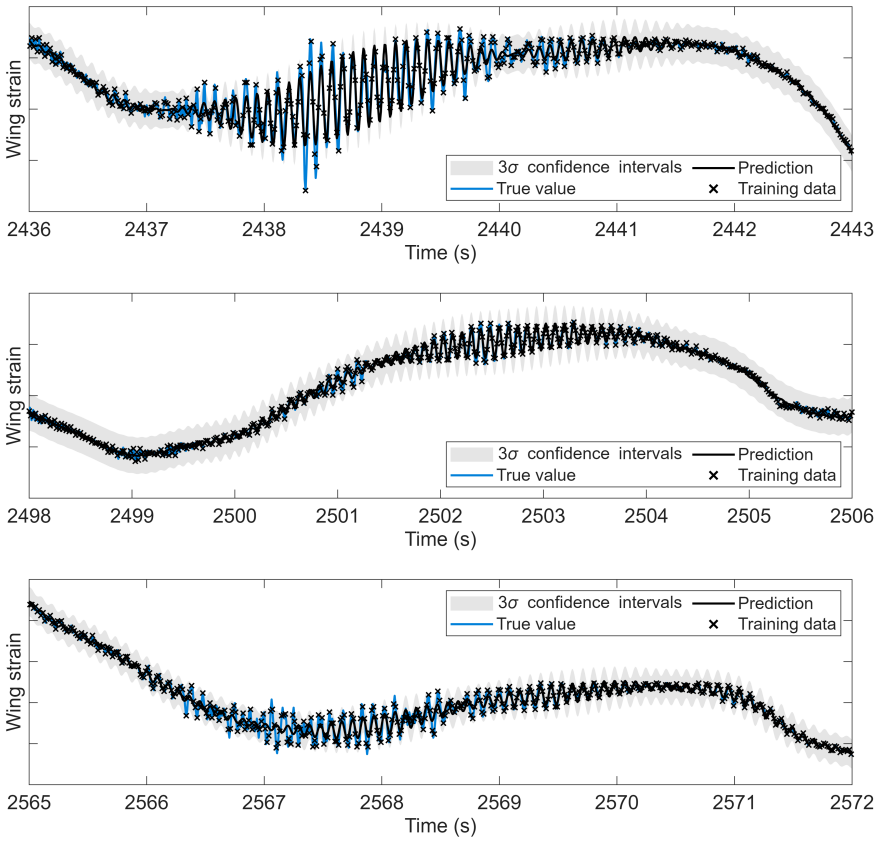

Figure A.18: Predictions of inner port side wing strain during the three in-flight turns using a spectral mixture kernel with 3 components.

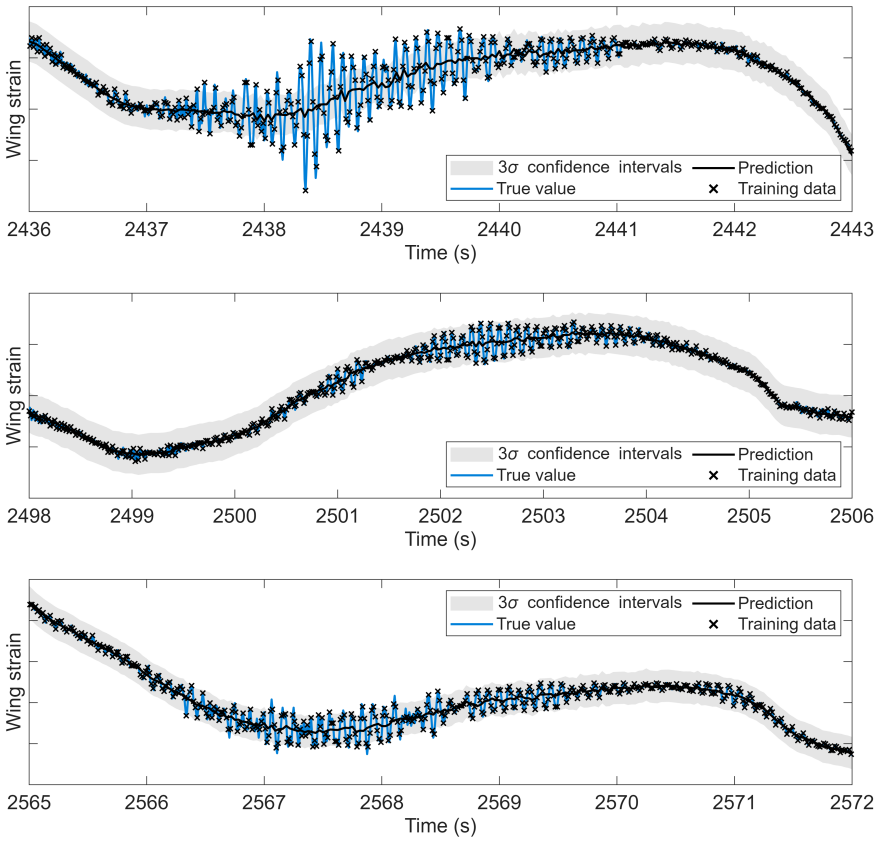

Figure A.19: Predictions of inner port side wing strain during the three in-flight turns using a global physics-informed kernel.

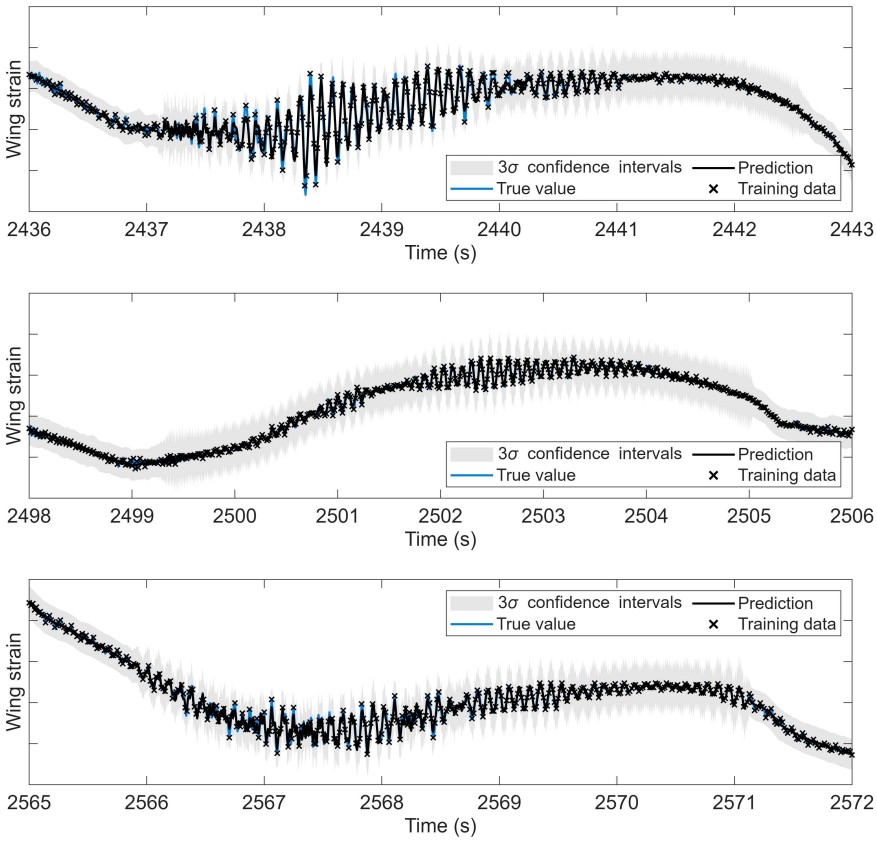

Figure A.20: Predictions of inner port side wing strain during the three in-flight turns using a SE kernel mixture.

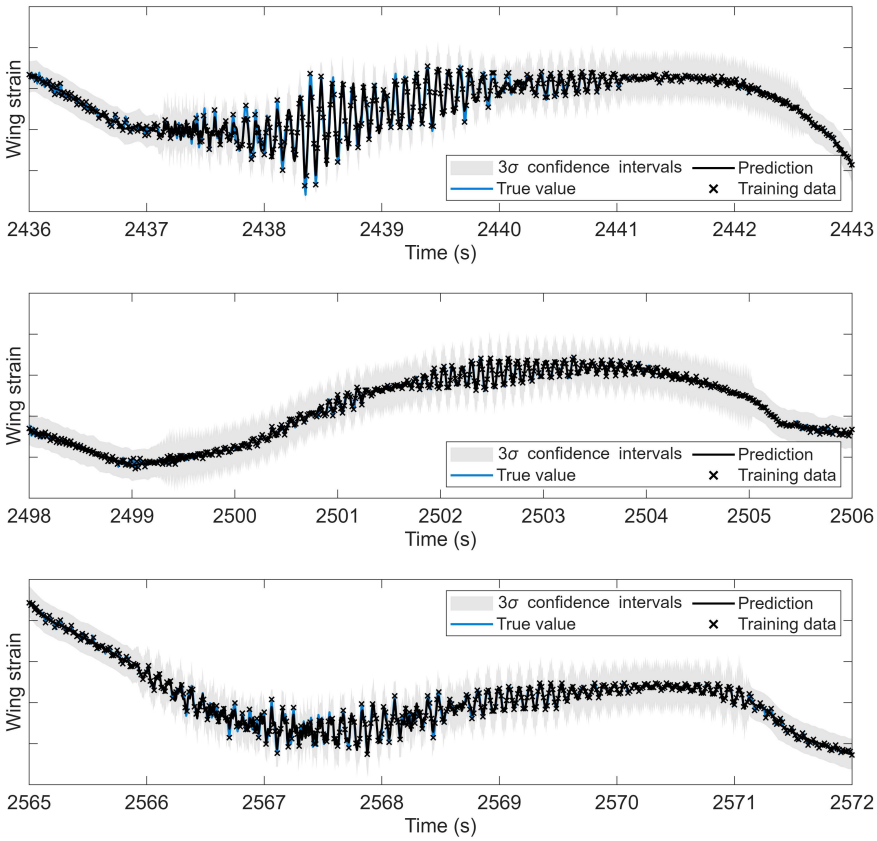

Figure A.21: Predictions of inner port side wing strain during the three in-flight turns using a physics-informed kernel mixture.

Appendix B. Non-parametric heteroscedastic noise

For the non-parametric, heteroscedastic noise model implemented in Section 4, the assumption (arising from standard GP theory) is made that the observation noise (see Equation 11) can be modelled as a draw from a Gaussian distribution with zero mean and variance σ_n^2. More succinctly, deck acceleration a is modelled as,

$$a = f(\theta, u) + \epsilon, \qquad f \sim \mathcal{GP}(0, K), \qquad \epsilon \sim \mathcal{N}(0, \sigma_n^2) \tag{B.1}$$

where the latent function f captures both the and lift and nominal/operational forces through the switching kernel. Such a model provides no mechanism for allowing the variance on the measurements to change dependent upon the input values (e.g. wind speed), instead imposing a noise model with

a fixed variance. This results in the uncertainty bounds that are returned by the global noise models models being too large at low wind speeds in northerly and southern directions, and is a consequence of having to compensate for the larger variance seen at higher wind speeds. A more appropriate model would allow for tighter confidence regions at lower wind speeds, that can widen at higher wind speeds.

To that end, we now consider a heteroscedastic noise process, where the observation noise itself becomes a function of the inputs, resulting in a model of the form,

$$a = f(\theta, u) + \epsilon(\theta, u), \qquad f \sim \mathcal{GP}(0, K), \qquad \epsilon \sim \mathcal{N}(0, r(\theta, u)) \tag{B.2}$$

This formulation adds an additional step into the inference procedure; estimating the new observation variance term $r(\theta, u)$. A common approach is to place a second Gaussian process prior over the term, effectively treating the noise variance as a latent function to be learnt instead of a fixed parameter, and can be expressed as,

$$r(\theta, u) \sim \mathcal{GP}(\mu_r, k_r). \tag{B.3}$$

where μ_r is the mean of the noise process, and k_r the covariance function. Here, we use a linear mean function to avoid imposing an arbitrary noise scale, with the squared exponential function used for the kernel. The overall model therefore now contains two GPs; one that governs the predictions of the noiseless underlying process, and another that captures the noise. Such a model is defined as a heteroscedastic Gaussian process (HGP). Although the use of a heteroscedastic noise model enhances the predictive power of the model, both the evidence and predictive distributions (the conditional mean and variance) are no longer analytically tractable. To perform inference here, we follow the variational approximation scheme as detailed in [53].

Appendix C. Computational complexity

The computational cost of Gaussian process regression is dominated by the inversion of the covariance matrix, which scales as $\mathcal{O}(n^3)$ with the number of training points n. In this study, model hyperparameters $\boldsymbol{\theta}$ were optimised by minimizing the Negative Log-Marginal Likelihood (NLML):

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2}\mathbf{y}^\top K_y^{-1}\mathbf{y} + \frac{1}{2}\log|K_y| + \frac{n}{2}\log 2\pi \tag{C.1}$$

where $K_y = K + \sigma_n^2 I$ is the covariance matrix for the noisy targets $\mathbf{y}$ [15]. To handle the many minima present in the surfaces associated with mixture kernels, a Quantum Particle Swarm Optimization

(QPSO) algorithm was employed. The optimizer was configured with a population size of 200 and a termination tolerance of 1×10^{-3}.

For the workstation used in this study (DELL Precision 5820, Intel Core i9-10900X CPU, 3.70GHz, 64GB RAM), the dataset sizes for case study B represent the practical upper limit for closed-form inference. Table C.4 and Table C.5 provide the training times for Case Studies A ($n = 300$) and B ($n = 12,800$), respectively.

Table C.4: Kernel structures and computation times for Case Study A: Tamar Bridge.

Model	**Kernel structure**	**Comp. time (s)**
SE kernel	k_{SE}	16.86
SM kernel	$\sum_{q=1}^{Q=3} k_{\text{SE}}^{q} k_{\text{Cos}}^{q}$	74.53
PI global	$k_{\text{Lift}} + k_{\text{SE}}$	20.57
SE mixture	$w_{\text{comp}} k_{\text{SE},1} + (1 - w_{\text{comp}}) k_{\text{SE},2}$	29.31
PI mixture	$w_{\text{comp}} k_{\text{Lift}} + (1 - w_{\text{comp}}) k_{\text{SE}}$	27.05

Table C.5: Kernel structures and computation times for Case Study B: Aircraft wing strain.

Model	**Kernel structure**	**Comp. time**
SE kernel	k_{SE}	09 h 02 min
SM kernel	$\sum_{q=1}^{Q=3} k_{\text{SE}}^{q} k_{\text{Cos}}^{q}$	19 h 39 min
PI global	$k_{\text{SE}} + \sum_{i=1}^{N=2} k_{\text{SDOF}}^{i}$	11 h 28 min
SE mixture	$k_{\text{SE}} + w_{\text{sig}} k_{\text{SE}}$	12 h 14 min
PI mixture	$w_{\text{sig},1} k_{\text{SE}} + w_{\text{sig},2} \sum_{i=1}^{N=2} k_{\text{SDOF}}^{i}$	14 h 42 min

As demonstrated in Case Study B, exact inference runtimes grow rapidly for moderate dataset sizes, with limits of 10,000-15,000 typical for a modern desktop computer. For a comparative case study, with emphasis on studying the effects of kernel structure, exact inference allowed a comparison of models without any additional variablity introduced due to approximate inference schemes. For larger datasets, and a more practical approach to implementation, the adoption of sparse GP formulations [65] would be necessary.