Incorporation of partial physical knowledge within Gaussian processes

D. J. Pitchforth [1], **T. J. Rogers** [1], **U. T. Tygesen** [2], **E. J. Cross** [1]
[1] University of Sheffield, Department of Mechanical Engineering,
Mappin Street, Sheffield, UK, S1 3JD
e-mail: **djpitchforth1@sheffield.ac.uk**

[2] Ramboll Energy, Global Division-Energy,
Bavnehøjvej 5, 6700 Esbjerg, Denmark

Abstract

The use of physical insight to assist data-based learning has the potential to help tackle two of the major problems faced by a purely data-driven model: a lack of insight into a model's operation, leading to mistrust, and the requirement for training data collection across all operational conditions. Physical knowledge can provide interpretability in to a modelled process, increasing confidence in predictions, whilst improved extrapolation capabilities of physically-informed models relax the need for expensive data collection. In many engineering applications, a well understood aspect of a process may be represented with a physical model with the remainder of the process (e.g. turbulence, nonlinearities, effects of mechanical joints etc.) better captured via a data-based component. This paper aims to show how partial knowledge may be incorporated within data-based Gaussian processes via manipulation of the covariance function (kernel) to improve performance and interpretability. The importance of model structure reflecting the underlying process is discussed and highlighted within the context of the free vibration of a cantilever beam.

1 Introduction

The modelling of modern engineering structures and accurate representation of phenomena within varying or extreme environments is challenging. The use of composite materials, effects of mechanical joints and changes in manufacturing tolerances, for example, can induce significant variability in dynamic behaviour, making the development and validation of physics-based models difficult. The challenge of modelling such phenomena, which may be hard to represent through traditional physics-based approaches, combined with an increasing availability of monitoring data, has lead to a surge in the adoption of data-based methods for many modelling tasks. These allow for direct learning of patterns and relationships within data without the need for complete knowledge of the underlying process, leading to highly flexible model structures.

Although a powerful tool when implemented correctly, data-based methods are not without limitations; performance of models far from observed conditions typically suffers, increasing the demand for expensive data collection. In many cases it may not be possible to obtain data for a given event e.g. structural response in extreme weather, forcing models to rely on extrapolation. The overfitting of models, particularly highly flexible ones, is also a problem [1]; this can lead to adequate or high perceived model performance within the training stage of model construction and a significant drop in performance when new data is presented.

The emerging field of physics-informed machine learning attempts to incorporate prior physical knowledge within the design of machine learning models, with the aim of benefitting from the advantages of both types of model structure. Insight, interpretability and improved extrapolation capabilities may be provided via a physics-based component, with flexibility and an ability to model unknown processes the task of data-based learning. An interesting focus point of this research field is how various forms of prior knowledge (algebraic equations, simulations, human feedback etc.) may be included within a model and how this affects the combined model structure. Von Rueden et al. [2] provide an overview of integrating prior knowledge within machine learning models whilst Cross et al. [3] focus more specifically on the field of SHM.

The work of this paper will focus on the integration of prior knowledge within Gaussian Process (GP) regression models. A GP is a non-parametric, flexible, Bayesian machine learning technique [4] and has shown to be effective within a variety of structural dynamics applications, ranging from the prediction of aircraft landing gear loads to wind turbine power curves [5, 6, 7]. The structure of a GP is defined through the selection of a mean function $m(x)$ and a covariance function (or kernel) $k(x, x')$, the modification of which provides two clear avenues for the inclusion of physical knowledge.

One may consider the construction of physics-informed machine learning models to lie on a scale between the extremes of a purely physics-based and purely data-based models, similar to that shown in Figure 1. In the case of complete physical understanding, a purely physics-based model would be most appropriate whilst zero knowledge of a process would suggest the implementation of a purely data driven learner.

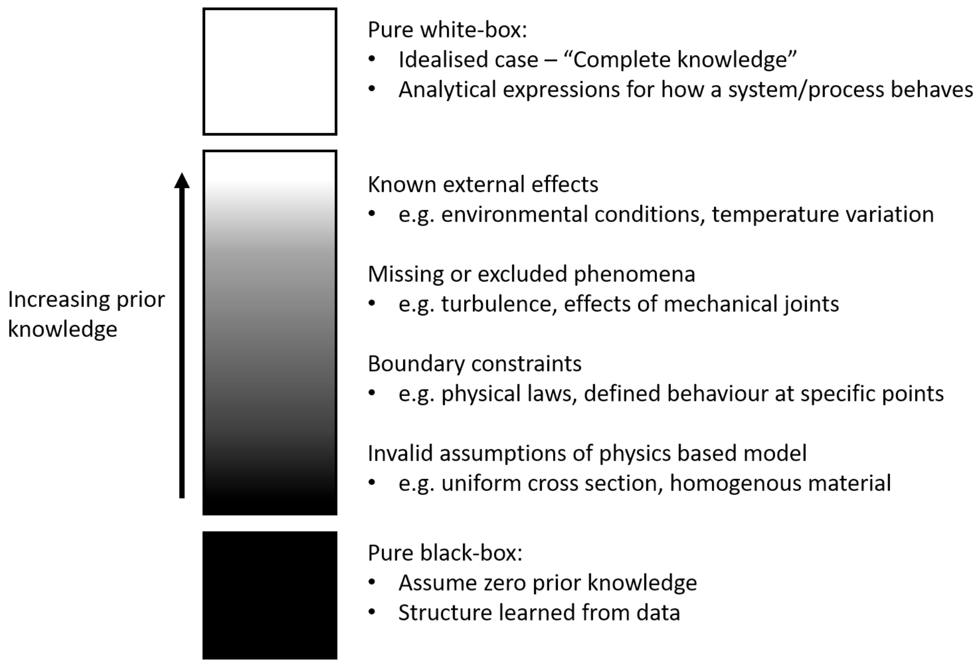

Figure 1: Sliding scale of prior knowledge inclusion within physics-informed machine learning. Examples of prior knowledge are placed in their approximate place on the scale for visualisation purposes, however there will be significant overlap between categories.

External effects such as environmental conditions and weather can change the dynamic behaviour of a structure and failure to account for this can lead to increasing model error and misclassification of damage states [8, 9]. Knowledge of how a structures behaviour changes with its environment can be useful to improve prediction quality within varying environments. Zhang et al. [10] incorporated the relationship between temperature and longitudinal deflection of a bridge deck within the mean function of a GP to improve prediction quality across varying seasons.

In many engineering applications, there is some aspect of a process that is not well understood and is not accounted for within a physics-based model. Previous work of the authors investigated the use of Morisons equation in the prediction of wave loads on offshore structures [11, 12]. Morisons equation relies on a number of simplifying assumptions and does not account for phenomena such as vortex shedding and turbulence [13, 14]. In [12], Morisons equation was used within the mean function, with the learning of the excluded processes the role of the data-based GP-NARX. This improved predictive performance, particularly in instances of reduced training data coverage.

The inclusion of boundary constraints within a model represents a scenario where some aspect of a systems behaviour is known for a given condition, often in the form of a physical law. Aspects of prior knowledge in the form of boundary constraints are strictly enforced within a model structure and a high degree of certainty that the constraints hold is therefore required. However, boundary constraints are often very basic knowledge and don't cover or summarise the main behaviour to be modelled. The work of Jidling [15] predicted strain fields with a Multiple Output Gaussian Process (MOGP) and enforced equilibrium constraints for linearly elastic materials through manipulation of the cross-covariance terms within the kernel. Jones [16] used Neumann boundary conditions to improve the localisation of acoustic emissions in instances of low data coverage.

The aim of this paper is to discuss how the level and type of prior knowledge available changes how to best manipulate model structure. The consideration of the scale within Figure 1 and where a particular model structure may lie will be a useful tool throughout the discussion of ideas. This concept is investigated in the context of free vibration of a cantilever beam where it is assumed only limited prior knowledge is available. The methods used within the paper will focus on how to incorporate partial physical knowledge within a GP through the use of kernel design.

The covariance function (kernel) of a GP defines a family of functions from which predictive samples may be drawn. Through the manipulation and selection of kernels, one may enforce desirable or physically derived behaviours within the predictions of a GP. A useful property of kernels for varying the inclusion of prior knowledge within a GP is the ability to be combined, through addition, multiplication and composition, with other kernels. The use of physically derived kernels, K_{Phy}, in combination with flexible, more generally applicable, kernels, such as a Squared Exponential, K_{SE}, allows for the creation of model structures where K_{Phy} aims to encode some aspect of prior knowledge and K_{SE} captures unknowns. For example, an additive kernel structure may aim to capture phenomena not represented within a physics based model:

$$K(x, x') = \underbrace{K_{Phy}(x, x')}_{\text{Understood process}} + \underbrace{K_{SE}(x, x')}_{\text{Excluded phenomena}} + \underbrace{\sigma_n^2 \delta_{ij}}_{\text{Noise}} \tag{1}$$

2 Case study: Free vibration of a cantilever beam

To demonstrate the proposed method of handling partial knowledge within the GPR framework, we employ here a simulated case study of a cantilever beam. This case study is useful for a number of reasons; a beam assumption is applicable to a wide range of engineering applications from the modelling of offshore structures to aeroplane wings, and importantly, it provides us with an opportunity to explore the embedding of various levels of physical insight. Where a structure of interest is well modelled by a cantilever beam, we have full analytical solutions available. If a structure is cantilever but not of uniform cross section or homogeneous material, knowledge of the mode shapes may no longer be assumed (this is the example we pursue here). Finally, we may face a situation where we are only confident about boundary conditions, in which case we may wish to employ constraints within the GPR, for example [16, 17].

The dataset used within the study was from a simulated cantilever beam as shown in Figure 2, where x is distance along the beam from the fixed end.

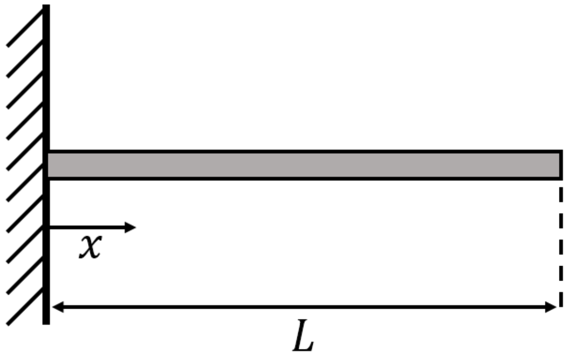

Figure 2: Cantilever beam of length L.

The free vibration response of a beam can be found through through separation of variables with a solution of the form $Y(x,t) = W(x)T(t)$. The response of a beam in free vibration may be calculated through the superposition of normal modes [18].

$$Y(x,t) = \sum_{i=1}^{\infty} W_i(x)T_i(t) = \sum_{i=1}^{\infty} W_i(x)e^{-\zeta\omega_n^i t}(A_i\cos(\omega_n^i t) + B_i\sin(\omega_n^i t)) \tag{2}$$

where A_i and B_i are determined from the initial conditions of the beam. For many sets of known boundary constraints it is possible to derive an exact analytical expression for the mode shapes $W_i(x)$. Following Blevins [19], for a fixed-free beam of uniform cross section in free vibration, we have

$$W_i(x) = \cos(\beta_i x) - \cosh(\beta_i x) - \frac{\cos(\beta_i L) + \cosh(\beta_i L)}{\sin(\beta_i L) + \sinh(\beta_i L)}(\sin(\beta_i x) - \sinh(\beta_i x)) \tag{3}$$

where L is the length of the beam and β_i is a constant specific to the boundary conditions and mode. The analytical solutions of Blevins [19] were used to create a simulated dataset of a fixed-free beam in free vibration. The first four modes of the beam were investigated, with plots of the mode shapes, $W_i(x)$, and oscillatory behaviour, $T_i(t)$, shown in Figure 3.

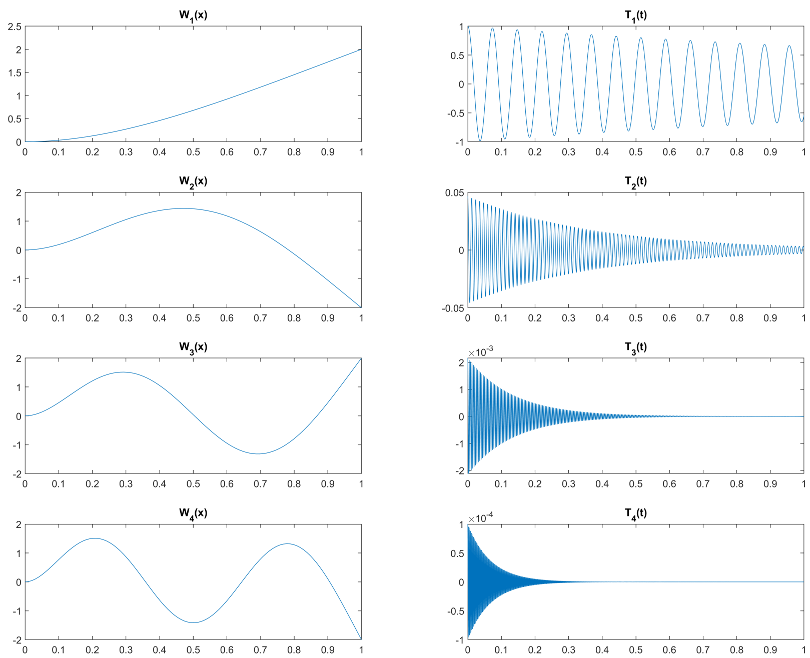

Figure 3: $W_i(x)$ and $T_i(t)$ plots of the first four modes of a simulated cantilever beam in free vibration.

It is worth noting that the main focus of this case study is not to minimise prediction error on a test set, as for the cantilever beam simulated this could be achieved through an analytical solution. The aim here is to investigate how to vary the level of physical knowledge incorporated within a machine learning model and discuss the effects this will have. At one extreme would lie “complete knowledge”, i.e. the analytical solutions of Blevins [19] and Rao [18], whilst a purely data-based approach would represent zero prior knowledge.

2.1 Relaxing assumptions through kernel design

The construction of physics-based (white-box) models requires the approximation of phenomena and a reliance on assumptions and simplifications remaining representative. In the case of a cantilever beam these may include a perfectly fixed beam, a uniform beam cross section or homogenous material properties. When assumptions hold, physics-based models often provide efficient, accurate solutions to a range of engineering problems. However, the usage of physics based models when their assumptions breakdown can lead to significant increases in error. For a cantilever beam, examples of this might range from a non-uniform cross section of a turbine blade or aircraft wing, marine growth along the length of an offshore monopile or a soil foundation being approximated as fixed. The relaxing or removal of assumptions within a physics-based model will therefore allow for the widening of conditions in which the model may be used. Examples of restricted modelling cases are summarised in Table 1.

Table 1: Summary of assumptions and modelling restrictions within the analytical expressions of Blevins [19] and Rao [18] for a fixed-free beam in free vibration.

Modelling assumption	**Restrictive modelling case**
Uniform cross section.	Non-uniform cross section of a turbine blade or aircraft wing.
Perfectly fixed base.	Deflection of structures within soil or sea bed foundations [20].
Homogeneous material properties.	Composite materials with directional material properties.
Equal mass distribution along the length of the beam.	Marine growth adding mass along the length of offshore monopiles [21].

Many key restrictions of the physics-based model of the cantilever beam come from the analytical expression for the mode shape, $W_i(x)$, which is specific to the boundary conditions of the beam. This paper suggests a means of removing these restrictions, accounting for partial knowledge by combining different covariance functions, each to capture different components of the combined target function. For the beam where we are assuming that mode shapes are unknown, we have the situation where the displacement we wish to predict is a product of oscillatory behaviour and those unknown modes, $Y(x,t) = W(x)T(t)$ where $W(x)$ is unknown. The covariance of Y in this case is equal to $cov(Y, Y')$=$K_W(x, x')K_T(t, t')$. Here we use the SE kernel to model the covariance, K_W, of the unknown mode shapes:

$$K_{SE}(\tau) = \sigma_f^2 \exp\left(-\frac{1}{2}(\tau)^T \Lambda^{-1}(\tau)\right) + \sigma_n^2 \delta_{ij} \tag{4}$$

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 $\tau = x - x'$ is the distance between a pair of input points x and x'. To capture the covariance of the oscillatory behaviour, K_T, we employ the derived SDOF kernel [22]:

$$K_{SDOF}(\tau) = \frac{\sigma^2}{4m^2\zeta\omega_n^3} e^{-\zeta\omega_n|\tau|}\left(cos(\omega_d\tau) + \frac{\zeta\omega_n}{\omega_d} sin(\omega_d|\tau|)\right) \tag{5}$$

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 a decaying SDOF oscillator, a useful property to encode.

For the prediction of the full beam response, the SE and SDOF kernels are multiplied together to predict the response contribution of a single mode i. These kernel products may then be summed for a given number of modes N to predict the combined response of the beam. This will lead to a combined kernel structure of

$$K(\tau_x, \tau_t) = \sum_{i=1}^{N} \underbrace{K^i_{SE}(\tau_x)}_{\substack{\text{Mode}\\\text{shape}}} \underbrace{K^i_{SDOF}(\tau_t)}_{\substack{\text{Oscillatory}\\\text{behaviour}}} + \underbrace{\sigma_n^2 \delta_{ij}}_{\text{Noise}} \tag{6}$$

where $\tau_x = x - x'$ and $\tau_t = t - t'$ represent distances between points in the spatial and temporal inputs respectively. The combined model has the flexibility to recover $W_i(x)$ and $T_i(t)$ with an important distinction; the analytical form of the cantilever mode shape has not been fixed and the corresponding assumptions used to construct it have been relaxed.

2.2 Recovery of mode shapes

An important test for the constructed model is the recovery of the mode shapes $W_i(x)$ and temporal functions $T_i(t)$ from the combined response of the beam. Although the simulated beam will obey the analytical expressions for $W_i(x)$, this constraint was removed from the model structure via the introduction of the SE kernel. As such, the capability of the model to recover the mode shapes is a useful property to measure. The use of a simulation is useful for this case as it allows the exact mode shapes used to construct the response to be compared with the predictions and performance levels to be measured. Eight evenly spaced data points were selected along the length of the beam as training points, with the response simulated at $8192Hz$. Model performance was measured on an unseen test set of 100 points along the beams length simulated for one second.

To extract the mode shapes from the combined beam response prediction, the posterior contributions for each mode i within the kernel sum will require decomposition. This was achieved through the use of Gaussian conditionals, following [23], with details provided in Appendix A. The predictive performance for each mode was measured using two metrics: the Normalised Mean Square Error (NMSE) and the Mean Standardised Log Loss (MSLL). The MSLL is a probabilistic measure, with superior models having more negative scores. A breakdown of predictive performance is shown in Table 2 and plots of the recovered $W_{1:4}(x)$ and $T_{1:4}(t)$ are shown in Figure 4

Table 2: Modal performance breakdown of beam response prediction.

Model target	Function	NMSE (%)	MSLL
First mode	$W_1(x)$	0.002	-5.472
	$T_1(t)$	0.411	-3.088
Second mode	$W_2(x)$	0.225	-3.520
	$T_2(t)$	0.285	-2.677
Third mode	$W_3(x)$	9.949	-1.118
	$T_3(t)$	0.240	-2.055
Fourth mode	$W_4(x)$	1.221	-2.522
	$T_4(t)$	0.541	-1.969

A key trend observed within Table 2 is that performance generally worsens for the higher modes of the beam. This is to be expected for a number of reasons; firstly, the mode shapes are more complex for the higher modes, meaning that the same eight spatial points measured along the beams length must be used to represent additional, sharper changes in direction. The magnitude of vibration of the higher modes was also significantly lower, making identifying a modes contribution from the full response more challenging. A

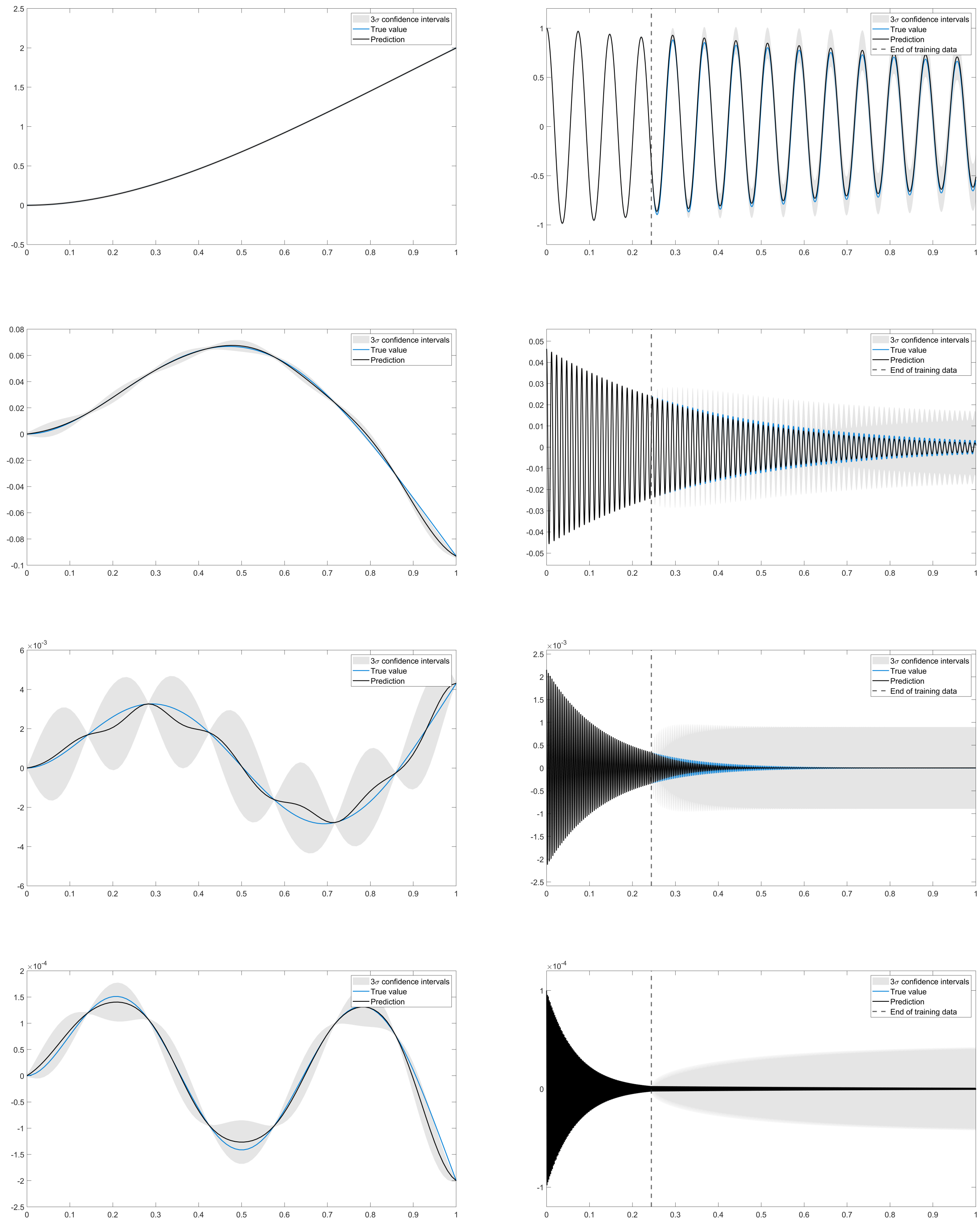

Figure 4: Breakdown of recovered $W_{1:4}(x)$ (left column) and $T_{1:4}(t)$ (right column) from a cantilever beam in free vibration by a combined kernel of the form $K(\tau_x, \tau_t) = \sum_{i=1}^{4} K_{SE}^{i}(\tau_x) K_{SDOF}^{i}(\tau_t) + \sigma_n^2 \delta_{ij}$. All models observed eight evenly spaced points along the beams length at time points [1:2:2000] as training data.

useful property of the SDOF kernel for identifying modal contributions was having physically interpretable hyperparameters, specifically the natural frequency ω_n. The first four natural frequencies of the beam were able to be identified and the corresponding hyperparameters fixed within the kernel. This reduced the number of required hyperparameters to optimise whilst encouraging the model to learn the response contributions at specific frequencies. Physically interpretable hyperparameters provide an avenue through which knowledge may flow in and out of a system. If parameters are learnt instead of fixed, optimisation can provide a means of retrieving this information from the system.

Within the plots of $W_{1:4}(x)$, the ability of the model to recover the mode shapes is observed. A useful property of the GP here is the quantification of uncertainty within predictions. For the first mode, where an excellent fit is achieved, a narrow confidence interval indicates the model is certain in its prediction. For $W_3(x)$, behaviour indicative of an underestimation of the lengthscale within the SE kernel is observed. The influence of observed points decays very quickly with distance and the model attempts to revert to its zero prior between the training points along the length of the beam. The confidence intervals of the GP expand quickly within these areas to reflect this however, preventing an overconfident, incorrect prediction.

One of the major advantages of including physical knowledge within a model is improving the ability to extrapolate. This can be seen in the $T_{1:4}(t)$ plots within Figure 4, where the model continues to predict beyond the end of observed training data (Dotted line). The behaviour of a decaying SDOF oscillator is encoded within the design of the SDOF kernel and has shown to be useful when conditioning models on a reduced quantity of data [22]. An important consideration when using prior knowledge to extrapolate however is any assumptions present within the construction of the model.

3 The balance between physics and data

Relating back the proposed model structure to Figure 1, it can be useful to consider how this changes the models relative position on the scale of prior knowledge inclusion. By relaxing the assumptions made within the physics-based model, the reliance on data has increased. If unsure about whether assumptions of a physics-based model may hold, e.g. the examples presented within Table 1, the inclusion of them has the potential to cause overly confident, erroneous model predictions. Were these assumptions maintained, for example via the inclusion of a mean function, a higher reliance on the physics-based component would be achieved, thereby moving the models position on the scale.

The balance between physics and data within a model is an important consideration within the construction of physics-informed machine learning models. Figure 5 highlights the concept of an optimum level of prior knowledge inclusion for a specific modelling scenario. A modelling scenario here is categorised by its relative levels of available knowledge and data. Ideally, one would wish for both high levels of data and prior knowledge, for example, a heavily sensored structure in well defined laboratory conditions; however in real life industrial applications this would be rarely achievable. In circumstances when neither knowledge or data is abundant, the predictions of models may often be unreliable or erroneous.

The concept of an “optimum” level of prior knowledge within a model should be considered when developing models along with the consequences of incorrect placement on the scale. The over inclusion of prior knowledge within a model can lead to a higher reliance on the physics-based component of the model which may not accurately represent the modelled system or phenomena. This can lead to reduced model flexibility and overly confident models. An under inclusion of prior knowledge may waste a key resource, useful for increasing model interpretability and capabilities in extrapolation. Models with reduced prior knowledge inclusion will rely more heavily on available data, increasing the demand for expensive data collection.

4 Further investigations

The work done so far has highlighted potential model structures for the relaxing of assumptions imposed on the vibration of a cantilever beam, however the initial testing of the model has focussed on a simulated dataset, representing an idealised case. The aim for future work is to implement and test the methods on a measured dataset, for which the assumptions made in the analytical formulas of Blevins [19] and Rao [18] would be less valid.

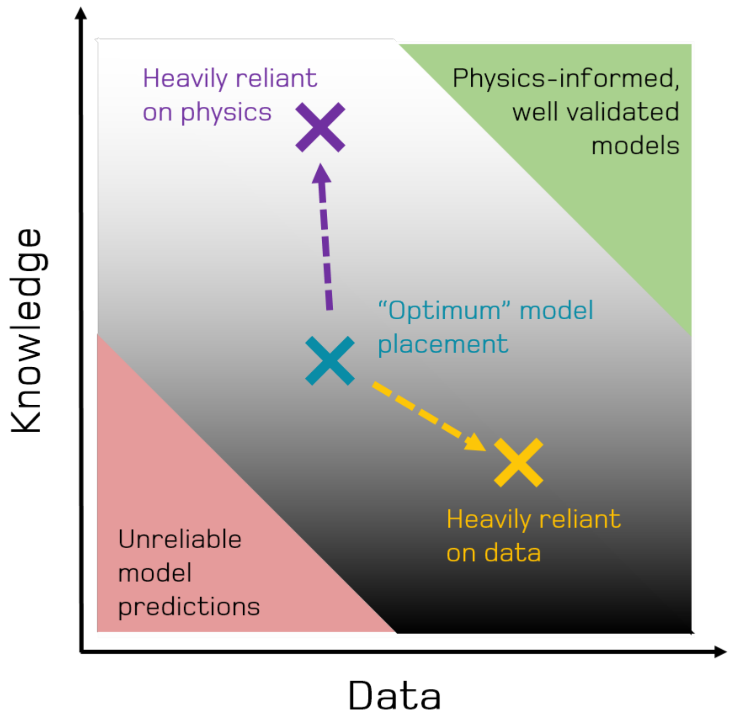

Figure 5: The concept of an optimum level of prior knowledge inclusion within a physics-informed machine learning model for a given modelling scenario. Examples of incorrectly estimating a models placement on the scale, leading to an heavy reliance on physics or data are shown.

The authors are currently running experiments on a monopile structure submerged within a wave tank. The response data should provide an interesting case study with which to test future model designs. For example, the presence of water around the structure has a significant impact on the relative damping. Testing the structure in dry conditions and within the tank could investigate how this affects the response of the structure and how well the model can cope with this change. There is also a mass at the top of the structure, to represent the rotor and nacelle of a wind turbine, which will change the natural frequencies and mode shapes. Gradually increasing the added mass at the top of the structure could investigate how the response of the structure changes.

5 Conclusions

The idea of an optimum level of prior knowledge to be included within a model changing depending on the modelling scenario was highlighted within the context of a cantilever beam in free vibration. The benefits of achieving this optimum level such as improved extrapolation, interpretability and reduced reliance on data collection were discussed. Consequences of being incorrectly placed on the scale of prior knowledge inclusion were also presented. These included reduced flexibility and potential reliance on unrealistic assumptions for a physics-heavy model and poor extrapolation, interpretability and increased demand for data collection for a data-heavy model.

The model developed within this paper explored a scenario where only partial prior knowledge was available and it was assumed the assumptions used to construct physics based models could not be guaranteed to hold. These assumptions were relaxed through the use of kernel design within a GP. The mode shapes of the beam, for which the analytical solutions were removed from the model as to relieve specific boundary conditions, were able to be recovered for the first four modes of the beam.

Acknowledgements

The authors gratefully acknowledge the support of Ramboll Energy, the University of Sheffield and the UK Engineering & Physical Sciences Research Council, through grant EP/S001565/1.

References

[1] D. M. Hawkins, "The problem of overfitting," *Journal of Chemical Information and Computer Sciences*, vol. 44, pp. 1–12, 2004.

[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, and 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, and 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] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

[5] G. Holmes, P. Sartor, S. Reed, P. Southern, K. Worden, and E. J. Cross, "Prediction of landing gear loads using machine learning techniques," *Structural Health Monitoring*, vol. 15, pp. 568–582, 2016.

[6] T. J. Rogers, P. Gardner, N. Dervilis, K. Worden, A. E. Maguire, E. Papatheou, and E. J. Cross, "Probabilistic modelling of wind turbine power curves with application of heteroscedastic Gaussian process regression," *Renewable Energy*, vol. 148, pp. 1124–1136, 2020.

[7] K. Worden, T. J. Rogers, and E. J. Cross, "Identification of nonlinear wave forces using Gaussian process NARX models," in *Nonlinear Dynamics, Volume 1*, 2017, pp. 203–221.

[8] K. Worden, C. Farrar, and G. Manson, "The fundamental axioms of structural health monitoring," *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, pp. 1639–1664, 2007.

[9] E. J. Cross, "On structural health monitoring in changing environmental and operational conditions," Ph.D. dissertation, University of Sheffield, 2012.

[10] S. Zhang, T. J. Rogers, and E. J. Cross, *Gaussian Process Based Grey-Box Modelling for SHM of Structures Under Fluctuating Environmental Conditions*, 2021, pp. 55–66.

[11] J. R. Morison, J. W. Johnson, and S. A. Schaaf, "The force exerted by surface waves on piles," *Petroleum Transactions*, vol. 189, pp. 149–157, 1950.

[12] D. J. Pitchforth, T. J. Rogers, U. T. Tygesen, and E. J. Cross, "Grey-box models for wave loading prediction," *Mechanical Systems and Signal Processing*, vol. 159, p. 107741, 2021.

[13] A. K. Agarwal and A. K. Jain, "Dynamic behavior of offshore spar platforms under regular sea waves," *Ocean Engineering*, vol. 30, pp. 487–516, 2003.

[14] E. Guilmineau and P. Queutey, "A numerical simulation of vortex shedding from an oscillating circular cylinder," *Journal of Fluids and Structures*, vol. 16, pp. 773–794, 2002.

[15] C. Jidling, J. Hendriks, N. Wahlström, A. Gregg, T. B. Schön, C. Wensrich, and A. Wills, "Probabilistic modelling and reconstruction of strain," *Nuclear Instruments and Methods in Physics Research*, vol. 436, pp. 141 – 155, 2018.

[16] M. R. Jones, T. J. Rogers, and E. J. Cross, "Constraining Gaussian processes for physics-informed acoustic emission mapping," *arXiv preprint arXiv:2206.01495*, 2022.

[17] E. J. Cross, T. J. Rogers, and T. J. Gibbons, "Grey-box modelling for structural health monitoring: Physical constraints on machine learning algorithms," in *12th International workshop on Structural Heath Monitoring*, F. Chang and F. Kopsaftopoulos, Eds., 2019.

[18] S. S. Rao, *Vibration of Continuous Systems*. John Wiley & Sons, Ltd, 2006.

[19] R. D. Blevins, *Formulas for Natural Frequency and Mode Shape*. R.E. Krieger, 1984.

[20] D. Lombardi, S. Bhattacharya, and D. Muir Wood, "Dynamic soil-structure interaction of monopile supported wind turbines in cohesive soil," *Soil Dynamics and Earthquake Engineering*, vol. 49, pp. 165–180, 2013.

[21] M. Martinez-Luengo, P. Causon, A. B. Gill, and A. J. Kolios, "The effect of marine growth dynamics in offshore wind turbine support structures," in *Progress in the Analysis and Design of Marine Structures*. CRC Press/Balkema, 2017, pp. 889–898.

[22] E. J. Cross and T. J. Rogers, "Physics-derived covariance functions for machine learning in structural dynamics," *IFAC-PapersOnLine*, vol. 54, pp. 168–173, 2021, 19th IFAC Symposium on System Identification SYSID 2021.

[23] D. Duvenaud, "Automatic model construction with gaussian processes," Ph.D. dissertation, University of Cambridge, 2014.

Appendix

A Conditional predictive distribution of additive kernel components

As derived in [23], the conditional predictive distribution of a kernel K_i within a kernel sum of the form $K = \sum_{i=1}^{i=n} K_i$ is expressed:

$$p(\boldsymbol{f_i^*}|\boldsymbol{f^*}, X^*, \boldsymbol{f}, X, \boldsymbol{\theta}) \sim \mathcal{N}(\mu_1^* + K_i^{*T}(\sum_{i=1}^{i=n} K_i)(\boldsymbol{y} - \mu_1 - \mu_2), K_i^{**} - K_i^{*T}(\sum_{i=1}^{i=n} K_i)K_i^*) \tag{7}$$

For the contribution of $K_{SE}^i K_{SDOF}^i$ within $\sum_{i=1}^{i=n} K_{SE}^i K_{SDOF}^i + \sigma_n^2 \delta_{ij}$, assuming a zero mean GP:

$$\begin{gathered} p(\boldsymbol{f_i^*}|\boldsymbol{f^*}, X^*, \boldsymbol{f}, X, \boldsymbol{\theta}) \sim \mathcal{N}((K_{SE}^i K_{SDOF}^i)^{*T}(\sum_{i=1}^{i=n} K_{SE}^i K_{SDOF}^i)\boldsymbol{y}, \\ (K_{SE}^i K_{SDOF}^i)^{**} - (K_{SE}^i K_{SDOF}^i)^{*T}(\sum_{i=1}^{i=n}(K_{SE}^i K_{SDOF}^i))(K_{SE}^i K_{SDOF}^i)^*) \end{gathered} \tag{8}$$

To obtain $W_i(x)$ and $T_i(t)$ from the posterior contribution of $K_{SE}^i K_{SDOF}^i$, the GP is used to predict whilst keeping either x or t fixed and varying the other. The result is then divided by the magnitude of the fixed function.