Title: **Parameter Estimation by Optimisation of Interpretable Hyperparameters with Physics-Informed Gaussian Processes**


Authors : Nathan A. Hinchliffe
Matthew R. Jones
Elizabeth J. Cross

## ABSTRACT

In this work, the feasibility of a covariance-based parameter estimation approach utilising a physics-informed Gaussian process regression is considered and explored. There are a number of ways of incorporating physical assumptions or knowledge into a Gaussian process regression. One of the richest is through adapting the kernel to reflect some known behaviour of the system of interest. In some cases, it is possible to derive such covariance functions from stochastic differential equations, demonstrated for a single-degree-of-freedom (SDoF) oscillator in [1], where the result is a very flexible kernel that can be used in a variety of modelling situations. In this case, in aid of interpretability, the covariance function has hyperparameters that are the natural frequency and damping ratio of the SDoF system from which the covariance is derived. The current paper explores whether this model may feasibly be used for parameter estimation, and herein, an investigation into the SDoF kernel's parameter estimation behaviour is conducted. This involves the visualisation of optimisation surfaces, exploration of what signal characteristics affect the estimates found, and finally, the proposal of a method to efficiently produce reliable parameter estimates. The estimates produced by this method show good agreement with the known system parameters over the range of systems tested, as well as good robustness to noise in the measured data.

## INTRODUCTION

The process of system identification describes the development of mathematical models from data. Many possible approaches can be taken for system identification, broadly categorised into black-box, grey-box and white-box [2]. The black-to-white box spectrum describes how much prior physical knowledge can be used in the construction of the model form, where black-box models offer no insight into the physical system, while white-box models are derived purely from physics-based models. Grey-box methods then offer a compromise where some of the model form is inferred from physical understanding, while others are learned purely from data. Furthermore, the models developed through system identification are classified into two general forms: input-out

Nathan A. Hinchliffe, nahinchliffe1@sheffield.ac.uk. Dynamics Research Group, School of Mechanical, Aerospace and Civil Engineering, University of Sheffield, Sheffield, UK

models and output-only models [2]. Selecting the model form is predominantly determined by what data is available during model development and in the context in which it will be used. Input-output models are generally preferred when possible due to their increased predictive ability; however, the inputs to operational systems are often impossible or impractical to measure. In their absence, assumptions need to be made before any useful analysis can be conducted. Regardless of the approach taken, the resulting model will contain parameters which must be tuned using data in a process called parameter estimation.

In engineering contexts, it is desirable to use as much system knowledge as is available when developing models, but it is rare for any system to be fully characterised. This leads to the development of models based on assumptions and partial knowledge, containing both physical and tuning parameters. When modelling systems in this way, the discrepancies between the model and underlying physics introduce bias into parameter estimates and errors into any predictions [3]. One of the proposed methods of accounting for this is introducing a secondary black-box model to estimate and correct these discrepancies. While this can be a powerful tool for improving the accuracy of the predictions, it also introduces other problems. These include increased model complexity, non-unique solutions, and heavy reliance on priors provided about the system parameters [3].

In this work, a covariance-based grey-box method is proposed and assessed for parameter estimation of a single-degree-of-freedom (SDoF) oscillator using a physics-informed Gaussian process. Through the use of an auto-covariance function derived from the SDoF equation of motion, instead of the equation of motion directly, greater flexibility is granted to the model in that it describes a distribution of functions rather than a single one, while still being parametrised by the interpretable system parameters, allowing for insight into the physical system. The aim of this work is to assess the feasibility of parameter estimation via a covariance function under conditions where the system follows the assumptions made in the model's development. Should this prove successful, then the flexibility and sensitivity of the model can be assessed in future works, where less may be known about the system structure in advance.

## PHYSICS-INFORMED GAUSSIAN PROCESSES AND THE SDOF KERNEL

Physics-informed machine learning is a modelling framework that combines the flexibility of machine learning methods with the interpretability of physics-based models. Many methods have been explored for combining prior physical knowledge with most of the widely used machine learning algorithms [4]. This study will use a physics-informed Gaussian process regression to identify the parameters of the stochastic differential equation (SDE) defining an SDoF oscillator under white noise loading. Gaussian process regression is a powerful, Bayesian regression method that operates in the function space. The models are defined by kernels containing mean and covariance functions, which form a prior to be conditioned on data, resulting in a posterior distribution of the latent function. For further details, the reader is directed to the highly-cited textbook by Rasmussen [5]. While there are many ways to embed physical knowledge into a Gaussian process [6], in this work, a covariance function derived from the equation of motion of an underdamped SDoF oscillator under white noise forcing is used. This is achieved through manipulating the SDE into its power spectral density and auto-covariance function, outlined by Cross and Rogers in [1]. In order to aid in the implementation and optimisation of this as a Gaussian process kernel, the derived auto-covariance function is simplified and normalised, resulting in the form given below.

$$k(\tau) = \sigma_f^2 \, e^{-\zeta\omega_n|\tau|} \left[\cos\left(\omega_n\sqrt{1-\zeta^2}\,\tau\right) + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\left(\omega_n\sqrt{1-\zeta^2}\,|\tau|\right)\right] \tag{1}$$

In this form, the covariance is fully characterised by the hyperparameters $\sigma_f^2$, $\omega_n$ and $\zeta$, which are directly interpretable as the signal variance, system natural frequency and system damping ratio, respectively. This, in combination with a zero mean function, forms a prior for Gaussian process regression. In order to use this for parameter identification, a method of training the kernel hyperparameters is required. In this context, training refers to the maximisation of the marginal likelihood of the posterior prediction given a set of training data, achieved through tuning of the hyperparameter values. Practically, this is achieved through a minimisation of the negative log marginal likelihood (NLML) with respect to the hyperparameters, as outlined in [5].

## PARAMETER IDENTIFICATION

In this section, the methodology for assessing the feasibility of parameter estimation with the SDoF kernel is outlined, and corresponding results are presented. First, the optimisation surfaces of the hyperparameters were explored to qualitatively understand the expected behaviour during training. Following this, the source of variance in the estimates was investigated, exploring how features of the training signal affect the parameter estimates. Finally, a method following a multi-task learning approach is proposed and tested for robustness.

Throughout these trials, the data used was generated from the simulation of an SDoF system, with a natural frequency at approximately 316.2 rad/s (50.3 Hz) and a damping ratio of 0.158. The simulations were generated from a state-space model (zero-order hold), with the input force generated as realisations of white noise. The simulation was run at 4096 Hz, and the resulting signal was then downsampled to 512 Hz for use.

### Likelihood Surfaces

In order to explore the optimisation space, examples of both the NLML (acting as training loss) and marginal likelihood surfaces were generated by drawing natural frequency and damping ratio hyperparameter values from uniform distributions covering the ranges of 0-1256 rad/s (0-200 Hz) and 0-1, respectively. An example of these surfaces is shown in Figure 1. From these surfaces, it can be seen that a clear single optimum exists, indicated by the single peak in the likelihood surface, and that this optimum lies close to the values of the true parameters. Additionally, the NLML surface exhibits curvature that could be easily navigated by an optimisation algorithm, giving confidence that a gradient descent optimisation is suitable to employ in this case.

Although this was encouraging to see, when generating multiple instances of these surfaces using different realisations of data from the same system, the location of this

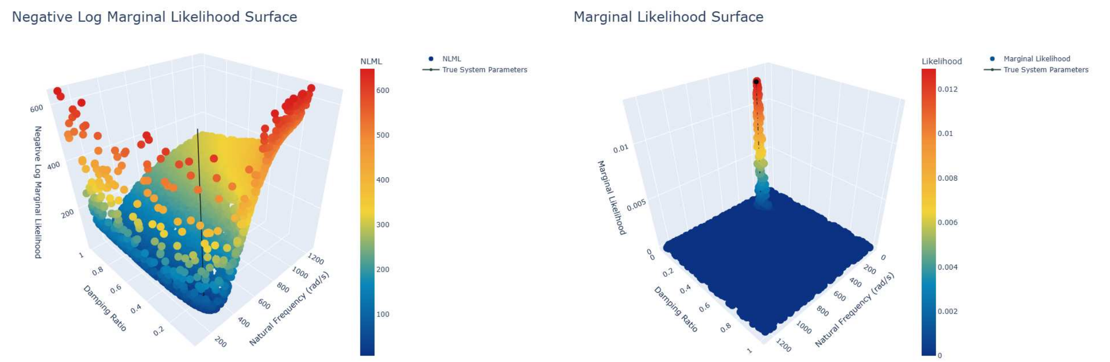


Figure 1. Negative log marginal likelihood surface (Left) and marginalised likelihood surface (Right), allowing for visualisation of the training optimisation surface and optimum location in the hyperparameter space

optimum was seen to shift relative to the true parameter values. This suggests that some stochastic nature of the signal propagates through to the optimum parameter location. In order to reliably extract an accurate estimate of the system parameters, this variance must first be understood.

**Parameter Distribution**

In order to investigate the source of the variance seen in the parameter estimates, the SDoF kernel's hyperparameters were trained on a large number of signals to produce an approximation of the parameter marginal distributions. This was repeated for training signals of different durations, the results shown in Figure 2. It is worth noting here that one of the primary limitations of Gaussian processes is their cubic order in computational time and cost as the number of data points increases, caused by an inversion of the covariance matrix. This imposes a practical limit on the duration of the signals that can be used for this training without employing a sparse scheme [7], which is considered as future work.

The results in Figure 2 show an apparent and expected reduction in variance observed in the parameter estimates as the training signal duration increases. Following this, approximations of marginal distribution were extracted by forming histograms of the parameter estimates generated with 0.5s training signals, selected as a representative signal duration compromising between reducing the variance in the distribution and training time. The resulting histograms are shown in Figure 3. From this, it can be seen that both parameter distributions are centred around the true parameter values, indicating that an accurate estimate would be reliably achieved with a long-duration signal to reduce the variance sufficiently.

Figure 3 also shows that the shape of the distributions differ. This suggests that the parameter estimates may be affected differently by components of the training signal. In order to investigate how signal components affect the training result in the case of shorter training durations, an example training signal was sectioned into short-duration signals that were labelled as: 'regions dominated by response close to resonance', 'regions dom-

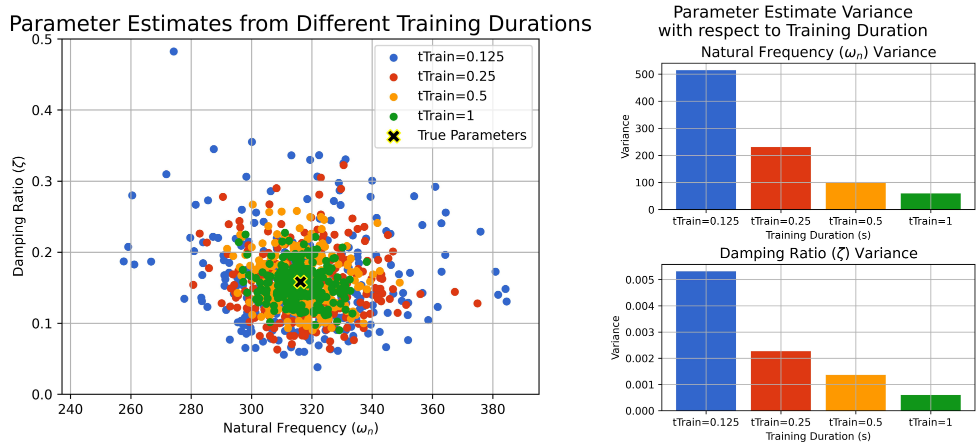


Figure 2. Trained parameter scatter plot (Left) and computed variance (Right) comparing training sets with signals of different durations, where tTrain indicates the duration of the training signal employed in seconds

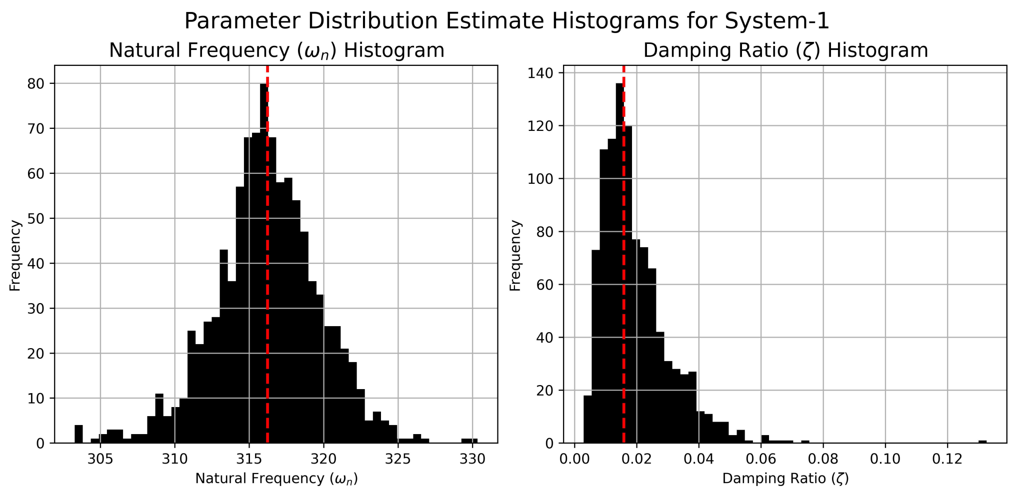


Figure 3. Histograms estimation of the marginal parameter distributions showing the distribution natural frequency (Left) and damping ratio (Right), where the red line indicates the system's true parameter values

inated by the high-frequency content', or 'regions containing some high-frequency content' where the natural frequency is still prominent. Training optimisations were then run on each of these sections, and the results plotted. Examples of the labelled signal regions and the resulting parameter estimates are shown in Figure 4.

From these results, it can be seen that the parameter estimates generated from signal sections dominated by the natural frequency exhibit the lowest variance in both parameters of interest and tend to underestimate the damping ratio. As the amount of high-frequency content in the signal increases, these variances both increase, and the damping ratio estimates trend upward. This further emphasises the importance of capturing a representative range of behaviours in the training signal to obtain accurate parameter estimates. This could be achieved through careful selection of the training signal, but also occurs naturally as the signal duration increases, as was seen previously.

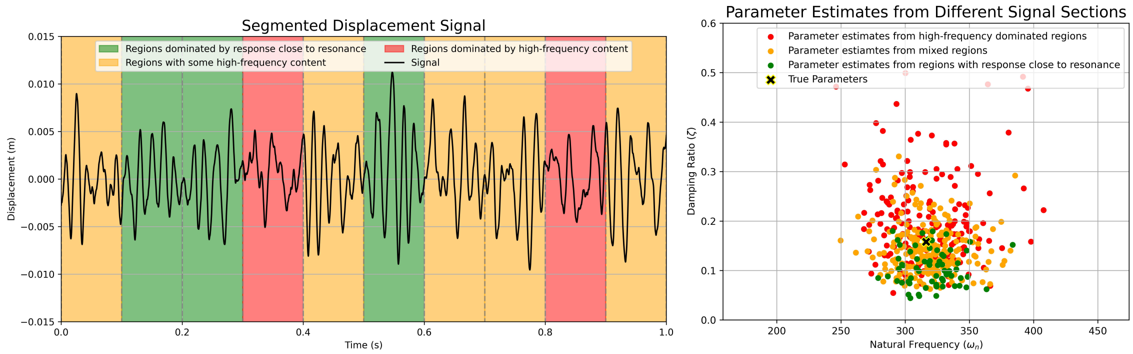


Figure 4. (Top) Example signal segmented for use in training, colour coded to loosely indicate frequency content of the signal section and (Bottom) the corresponding parameters that result from training on those signal sections

## Parameter Estimation

In order to train using an amount of data representative enough to reliably provide accurate parameter estimates without increasing the signal duration or considering sparse approximations, a training method adapted from the multi-task learning approach, as described by Rasmussen on pages 115-116 of [5], was adopted. Therein, a typical application of multi-task learning allows a model to learn hyperparameters shared between multiple datasets by summing the NLML calculated on each, and optimising on the resulting surface. Expanding on this idea, by splitting a long signal into many short signal segments, each can be thought of as an individual dataset, all of which have been generated by the same system, and thus share parameters. In separating the signal in this way and training on all the signals concurrently, the size of the required matrix inversion is significantly reduced. This allows for efficient training using a previously infeasible quantity of training data.

Applying this methodology, parameter estimates were obtained from signals generated by a range of example systems which share a natural frequency but vary in damping ratio. These signals were separated into 1000 0.5s signals, totalling 500s of training signal duration. The system parameters and parameter estimate results are shown in Figure 5. From these results, it can be seen that accurate parameter estimates were achieved across all systems tested, with almost all estimates having a very low percentage error. The exception to this was System 1, the very low-damping case. However, this still corresponds to a low error in the estimate due to the low magnitude of the damping ratio itself. The results also show that the accuracy of the estimates of natural frequency decreases as the damping ratio increases. This is likely a result of these signals containing more of the high-frequency content discussed previously, making identification of the natural frequency more difficult.

Finally, white noise was added to the signal and parameter estimates were extracted using the same method as above. The results can be seen in Figure 6. This shows that the parameter distribution and accuracy of the estimates did not significantly change as noise was added to the training signals. This was tested with various levels of signal noise, up to 25% (3:1 signal-to-noise ratio), showing an encouraging level of robustness to significant noise in the signal.

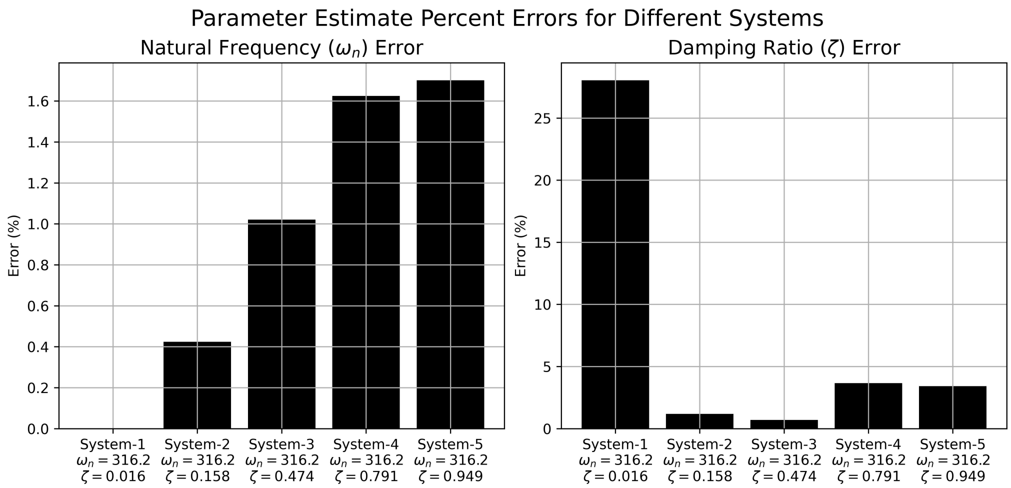


Figure 5. Error in parameter estimates for systems with varying damping ratios, trained over 1000 signals, 0.5s in duration

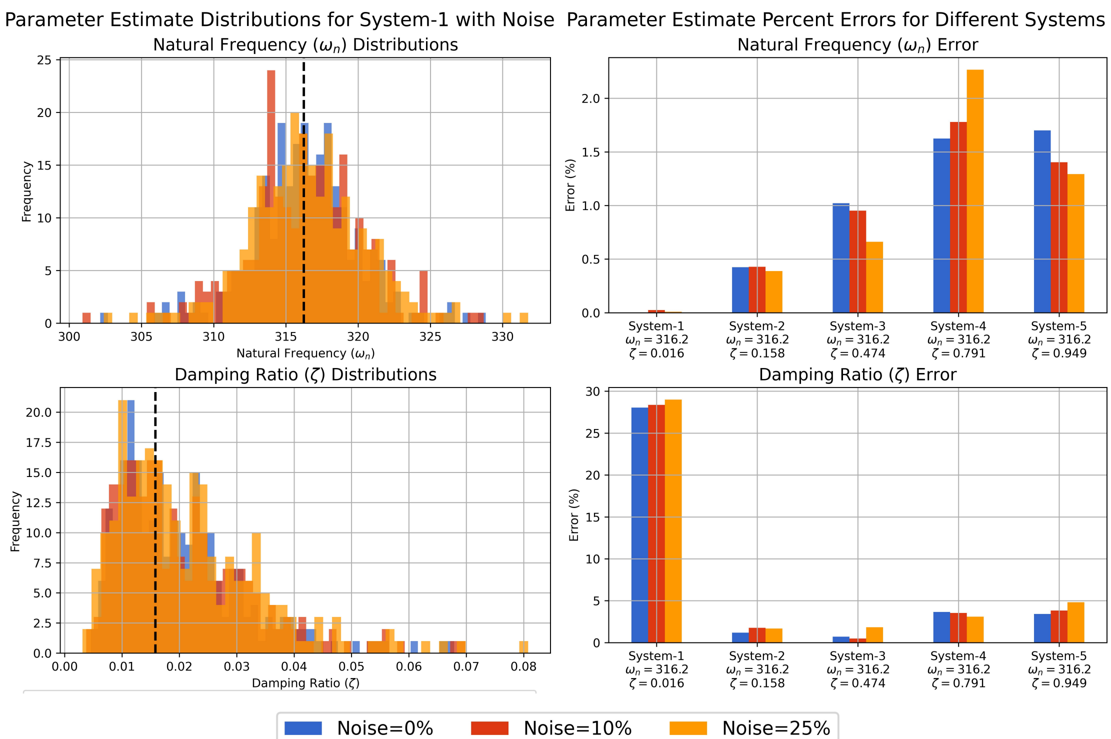


Figure 6. Histogram estimates of parameter distributions for the, where the true parameter values are marked by the dashed line (Left), and the error in parameter estimates (Right) arising from signals corrupted by various levels of noise

## CONCLUDING REMARKS

Physics-informed machine learning provides an interpretable yet flexible method of modelling systems from data. In most cases, this interpretability provides confidence in the model output, helping ensure the desired behaviour is learned. However, in the case of the physically-derived SDoF kernel, the hyperparameters themselves can be directly interpreted as parameters of an SDoF oscillator. This leads to the question of whether the values these hyperparameters take during training can be reliably interpreted for parameter identification.

In this work, the behaviour of the SDoF kernel's hyperparameters during training is explored. It was found that the likelihood surface of the interpretable hyperparameters has a single optimum, but the location of this optimum is highly dependent on the representativeness of the training data, which can be improved by increasing the amount of training data. However, this is problematic for Gaussian processes unless a sparse approach is considered due to the high computational cost. Therefore, a training methodology following a multi-task learning approach is proposed; by separating the training signal into multiple datasets and training over them concurrently, a large amount of data can be used in the training while maintaining computational efficiency. Following the proposed approach, the hyperparameter values identified closely agree with the true underlying system parameters and show good robustness to high levels of signal noise. This provides confidence that a Gaussian process covariance-based approach to parameter estimation is feasible and can be explored further. As such, future works will explore parameter estimation performance in more challenging situations, where information is more limited. This will include various unknown elements of the system structure, as well as data sparsity and downsampling.

## REFERENCES


1. Cross, E. J. and T. J. Rogers. 2021. "Physics-derived covariance functions for machine learning in structural dynamics," *IFAC-PapersOnLine*, 54(7):168–173, ISSN 2405-8963, doi: 10.1016/j.ifacol.2021.08.353.
2. Ljung, L. 1999. *System identification: Theory for the user*, Prentice-Hall information and system sciences series, Prentice Hall, Upper Saddle River, NJ [u.a.], 2. ed., 14. printing edn., ISBN 0136566952, literaturverz. S. 565 - 593.
3. Brynjarsdóttir, J. and A. O'Hagan. 2014. "Learning about physical parameters: the importance of model discrepancy," *Inverse Problems*, 30(11):114007, ISSN 1361-6420, doi: 10.1088/0266-5611/30/11/114007.
4. Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. 2021. "Physics-informed machine learning," *Nature Reviews Physics*, 3(6):422–440, ISSN 2522-5820, doi:10.1038/s42254-021-00314-5.
5. Rasmussen, C. E. and C. K. I. Williams. 2008. *Gaussian processes for machine learning*, Adaptive computation and machine learning, MIT Press, Cambridge, Mass. [u.a.], 3. print edn., ISBN 9780262182539.
6. Cross, E. J., T. J. Rogers, D. J. Pitchforth, S. J. Gibson, S. Zhang, and M. R. Jones. 2024. "A spectrum of physics-informed Gaussian processes for regression in engineering," *Data-Centric Engineering*, 5, ISSN 2632-6736, doi:10.1017/dce.2024.2.
7. Liu, H., Y.-S. Ong, X. Shen, and J. Cai. 2020. "When Gaussian Process Meets Big Data: A Review of Scalable GPs," *IEEE Transactions on Neural Networks and Learning Systems*, 31(11):4405–4423, ISSN 2162-2388, doi:10.1109/tnnls.2019.2957109.