



This is a repository copy of *Constraining Gaussian processes for physics-informed acoustic emission mapping*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/208858/>

Version: Published Version

Article:

Jones, M.R. orcid.org/0000-0001-7446-0833, Rogers, T.J. orcid.org/0000-0002-3433-3247 and Cross, E.J. orcid.org/0000-0001-5204-1910 (2023) Constraining Gaussian processes for physics-informed acoustic emission mapping. *Mechanical Systems and Signal Processing*, 188. 109984. ISSN 0888-3270

<https://doi.org/10.1016/j.ymssp.2022.109984>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Constraining Gaussian processes for physics-informed acoustic emission mapping

M.R. Jones^{*}, T.J. Rogers, E.J. Cross

Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, S1 3JD, UK

ARTICLE INFO

Communicated by W. Ostachowicz

Keywords:

Physics-informed learning
Damage localisation
Acoustic emission
Constrained Gaussian processes

ABSTRACT

The automated localisation of damage in structures is a challenging but critical ingredient in the path towards predictive or condition-based maintenance of high value structures. The use of acoustic emission time of arrival mapping is a promising approach to this challenge, but is severely hindered by the need to collect a dense set of artificial acoustic emission measurements across the structure, resulting in a lengthy and often impractical data acquisition process. In this paper, we consider the use of physics-informed Gaussian processes for learning these maps to alleviate this problem. In the approach, the Gaussian process is constrained to the physical domain such that information relating to the geometry and boundary conditions of the structure are embedded directly into the learning process, returning a model that guarantees that any predictions made satisfy physically-consistent behaviour at the boundary. A number of scenarios that arise when training measurement acquisition is limited, including where training data are sparse, and also of limited coverage over the structure of interest. Using a complex plate-like structure as an experimental case study, we show that our approach significantly reduces the burden of data collection, where it is seen that incorporation of boundary condition knowledge significantly improves predictive accuracy as training observations are reduced, particularly when training measurements are not available across all parts of the structure.

1. Introduction

In the field of structural health monitoring (SHM), the objective is to implement monitoring strategies that seek to detect and assess damage that is present in engineering infrastructure [1]. One particular branch of SHM techniques are those that consider a *data-driven* perspective, treating the damage identification paradigm as a problem of pattern recognition. Under such an approach, the task is to first collect data from a structure of interest by means of sensing hardware. Features sensitive to damage are then extracted and used to learn a statistical model that can inform damage identification.

Within the last decade or so, approaches that adopt such a viewpoint have become increasingly popular [2–5]. One area of SHM that has benefited from the use of data-driven approaches is damage localisation via acoustic emission (AE) [6–10], enabling progression in scenarios that previously presented significant difficulty, such as for complex geometries or inhomogeneous material compositions. Whilst there are many physical mechanisms that can produce acoustic emission waves, AE activity can often be attributed to the formation and progression of damage. For example, as a crack begins to form and grow, the associated redistribution of internal stresses will cause a small amount of elastic energy to be released, manifesting as a high-frequency stress wave that will travel through the material. There is, therefore, great benefit from a health monitoring perspective in being able to spatially locate these signals, particularly for larger structures, providing the platform for more informed maintenance strategies.

^{*} Corresponding author.

E-mail address: mrjones2@sheffield.ac.uk (M.R. Jones).

A fundamental component in many of these data-driven approaches is learning acoustic emission difference-in-time-of arrival (ΔT) mappings [11–14], which spatially characterise the difference in arrival time between two sensors of an AE source across the surface of a structure. Although promising, this method is severely hampered by the requirement to collect an extensive set of artificial sources over the entire structure, making the action of constructing the maps costly from both a financial and temporal perspective. This cost is particularly pertinent as the size of the structures considered grows, which is often where an SHM practitioner gains the most insight from localising damage. Therefore, to ensure that the process of acquiring ΔT maps does not become too cumbersome, it is critical that the size of the training set is minimised, especially for larger structures. The challenge here is then scaling to structures that contain complex geometrical features such as holes and bolts, which will disrupt the propagation path of the ultrasonic waves. These irregularities will induce sharp discontinuities in the AE arrival time, and so a standard data-driven learner will require a dense training set in these regions to suitably capture the feature behaviour, conflicting with the desire to reduce the amount of data that needs to be acquired. An additional challenge is faced from the action of retrofitting a monitoring system, where gaining access to the complete coverage of all parts of a structure is often challenging, particularly at locations that contain joints and interfaces. As such, it is important to obtain good model generalisation in areas where measurements may be more sparse, which proves difficult when solely reliant on a data-driven learner. These problems are not unique to the application of acoustic emission localisation however, and instead arise consistently wherever black-box learners are deployed. Whilst their flexible nature enable a high level of performance in the presence of an abundance of data, when presented with sparse training sets or forced to extrapolate, there is often no guarantee on how the predictions will behave. In this work, we propose a means of addressing the above limitations of methods that rely on ΔT mapping by incorporating physical knowledge into the learning process. Methodologies that consider such an approach can broadly be grouped under the term physics-informed machine learning, where physical insight is embedded into data-driven algorithms — for a general overview of this field, see [15]. Also referred to as grey-box modelling [16], the objective when adopting such a model architecture is to combine the expressive power of machine learning tools with physically derived laws or constraints. Predictions that are drawn from these models are then guaranteed to be representative of some underlying physical laws that govern the dynamics of the system under consideration.

One particular way to embed physically-derived insight is by constraining the learning algorithm with physical constraints such that subsequent predictions comply with these assumptions. In the context of Gaussian process regression [17], a Bayesian machine learning tool often employed in SHM, there are numerous ways constraints can be applied, as extensively discussed in [18]. For example, if one has knowledge of the shape of the latent function, then monotonicity or convexity constraints can be applied [19–23]. General bounds such as nonnegativity can also be imposed on the GP prior [24], as well as constrained to satisfy linear operators [25]. Where more specific insight is available, for example, the underlying equation of motion, derivation of an exact autocovariance is possible [26]. Constraints also exist in the form of vector-output Gaussian processes, where relevant physical relationships between multivariate targets can be embedded into the cross-covariance terms. Under a multivariate output framework, the inclusion of both ordinary differential equations [27,28] and boundary conditions have been explored [29].

The focus of the work presented here will be on the application of constrained Gaussian processes for learning difference-in-time of arrival/ ΔT acoustic emission maps. The nature of the constraints that are considered are those of physical boundary conditions, embedded into the model by firstly rewriting the covariance of the GP prior as a Laplacian eigenfunction expansion. Given that the eigenfunctions are unique to a user-specified domain, the GP can then be naturally constrained to some boundary conditions along a physical domain. This results in a model that retains the flexibility of a traditional machine learner, whilst adhering to known physical conditions that exist at boundary locations. To the authors' best knowledge, up until now, all previous works that consider a data-driven approach to AE localisation have been purely black-box in nature. This paper demonstrates that by ensuring that the model is constrained to known physical laws, the process of generating ΔT maps is made feasible, improving the predictive performance in cases of sparse/few training observations and where training measurements only partially cover the full structure spatially, and therefore, only partial coverage of the input space for the data-driven learner. General discussions surrounding where one may implement the constrained Gaussian process are also considered.

It should be noted that where suitably dense training data is available across the whole structure, it is not expected that the constrained GP will significantly improve performance, particularly where training data is available on boundary locations. In the case where the training set includes observations on and around a boundary, then these points can be seen as constraints themselves in the sense that the resulting predictions will be constrained to these measurements. The purpose of this paper, however, is to explore how adding physical-consistency into the GP prior can assist where acquiring data is challenging and thus training data is limited, as often occurs in the ΔT mapping procedure.

The paper proceeds as follows; Section 2 offers a brief introduction to Gaussian process regression, and then outlines the necessary theory for constraining a GP. Section 3 details the general procedure of constructing ΔT maps, including details of the data set used throughout the work. A practical discussion of how one can implement constraints in this context is then given. Section 4 presents the results and corresponding discussion, with concluding remarks offered in Section 5.

2. Constrained Gaussian processes

Gaussian processes provide a Bayesian, non-parametric tool for solving regression problems. In a regression context, a Gaussian process specifies a prior distribution over a latent function, $f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^d$, which in addition to a noise term ϵ , is believed to represent the target value y such that,

$$y = f(\mathbf{x}) + \epsilon \quad \epsilon \sim \mathcal{N}(0, \sigma_n^2), \quad (1)$$

where,

$$f \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (2)$$

It can be seen in the above equation that a Gaussian process is fully defined by a mean function, $m(\cdot)$ and a covariance function $k(\cdot)$, which together, characterise one's prior belief about the behaviour of the latent function. The mean function can be specified as any basis function expansion of \mathbf{x} , whilst also having the ability to generalise to an input space $\mathbf{x}_m \in \mathbb{R}^q$, where q can differ from the input space dimension d . The covariance function then captures the covariance between two points in the input space.

As a set of training data $\{X, \mathbf{y}\}$ become available, it is possible to condition the Gaussian process prior on this set of known input/outputs to form a posterior y_* , over an unknown input, \mathbf{x}_* . Following [17], simple Gaussian machinery allows us to arrive at a closed form expression for the distribution over y_* ,

$$p(y_* | \mathbf{x}_*, X, \mathbf{y}, \theta) = \mathcal{N}(\mathbb{E}[y_*], \mathbb{V}[y_*]), \quad (3)$$

where,

$$\mathbb{E}[y_*] = m(\mathbf{x}_*) + K(\mathbf{x}_*, X)(K(X, X) + \sigma_N^2 \mathbb{I})^{-1} \mathbf{y}, \quad (4)$$

$$\mathbb{V}[y_*] = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, X)(K(X, X) + \sigma_N^2 \mathbb{I})^{-1} K(X, \mathbf{x}_*). \quad (5)$$

When specifying a covariance, there will often be a number of hyperparameters θ that are required to be specified (note that θ can also incorporate coefficients of a mean parametric function). These hyperparameters will alter the behaviour of the chosen kernel and often have a meaningful interpretation. For example, many popular kernels contain a length scale parameter, which in a practical sense, governs how close inputs in the same dimension should be to influence one another. Due to the framework in which the Gaussian process resides, it is possible to determine the value of these parameters in a systematic manner by optimising the marginal likelihood of the model (also known as the model evidence). Should the reader be interested, information detailing this procedure is included in [Appendix A](#).

2.1. Embedding physical constraints into a Gaussian process

Although Gaussian processes present a powerful tool, as with all machine learners, they are still black-box in the sense that their performance is entirely reliant on the data that the model is trained on. In cases where sufficient training data are scarce, then the resulting model may struggle to adequately learn the underlying behaviour of the features. As introduced in Section 1, one way to circumvent such issues is to incorporate physical insight into the machine learner. In this paper, focus is directed on embedding boundary condition knowledge into a Gaussian process prior through constraining the covariance function. This presents one view of a constrained process. From the perspective of the nature of the physics considered, the advantage here is that the level of insight into the governing mechanistic laws can be relatively shallow; knowledge of boundary conditions are generally easier to come by than an exact governing differential equation. Additionally, through directly constraining the form of the prior, the approach does not rely on the addition of artificial observations at the boundary. In fact, the method employed here is a sparse approximation, and so is computationally cheaper than the standard implementation of a Gaussian process, both in terms of complexity and storage demands [30].

To constrain a Gaussian process in this manner, it is first necessary to make use of the following covariance function approximation [30]:

$$k(\mathbf{x}, \mathbf{x}') \approx \sum_j^m S(\sqrt{\lambda_j}) \phi_j(\mathbf{x}) \phi_j(\mathbf{x}') \quad (6)$$

Under this representation, the covariance function is defined as a basis function expansion across m Laplacian eigenfunctions ϕ of a user-selected domain, projected onto the spectral density S of the covariance that has been evaluated in a point-wise manner at the corresponding Laplacian eigenvalues λ . To calculate the Laplacian eigenpairs, one is required to solve an eigenvalue problem of the form,

$$-\nabla^2 \phi_j(\mathbf{x}) = \lambda_j^2 \phi_j(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad (7)$$

where ∇^2 is the Laplacian operator and Ω represents the domain of interest. As for any differential equation, its solution may be sought given boundary conditions described generally by,

$$\mathcal{P}[\phi_j(\mathbf{x})] = H(\mathbf{x}), \quad \mathbf{x} \in \delta\Omega, \quad (8)$$

where \mathcal{P} denotes some operator and H is an arbitrary function that maps \mathbf{x} to a known value which exists on the boundary of a domain $\delta\Omega$. The solutions for λ and ϕ are, therefore, bound to the chosen domain Ω , and, consequently, are unique to the boundary conditions specified in Eq. (8). Upon substitution into Eq. (6), each draw from the prior is then guaranteed to abide by these constraints, returning an expression for the covariance that is dependent on the boundary conditions of the feature space.

To specify a suitable kernel spectral density, it is possible to employ Bochner's theorem which states that the covariance of a stationary function can be represented by the Fourier transform of a positive, finite measure [31]. If this measure has a corresponding

density S , then the spectral density and the covariance are Fourier duals of one another [32]. For example, the spectral density of the Matérn 3/2 kernel takes the following functional form,

$$S(\omega) = \sigma_f^2 \frac{4\pi(2\nu)^\nu}{l^{2\nu}} \frac{\Gamma(\nu+1)}{\Gamma(\nu)} \left(\frac{2\nu}{l^2} + \omega^2\right)^{-(\nu+1)}, \quad (9)$$

where $\nu = 3/2$ and Γ denotes the Gamma function. For a test point \mathbf{x}_* , the predictive posterior is then defined as:

$$\mathbb{E}[y_*] = \Phi_* (\Phi' \Phi + \sigma_n^2 \Lambda^{-1})^{-1} \Phi' \mathbf{y} \quad (10)$$

$$\mathbb{V}[y_*] = \sigma_n^2 \Phi_* (\Phi' \Phi + \sigma_n^2 \Lambda^{-1})^{-1} \Phi_*', \quad (11)$$

where,

$$\Phi = (\phi_1(X), \phi_2(X), \phi_3(X), \dots, \phi_m(X)), \quad (12)$$

$$\Lambda = \text{diag}(S(\lambda_1), S(\lambda_2), S(\lambda_3), \dots, S(\lambda_m)), \quad (13)$$

with the mean function now set to zero. The reasoning for this is not due to the mean being uninteresting; that is quite the opposite in a grey-box context. However, specifying a physics-based basis function often requires a significant level of knowledge regarding the dynamics of the system being modelled. The intention here is to present a general method that can be applied with known boundary condition information, which, in most cases, is readily available. For interesting examples of applying a physics-based mean function, the reader is referred to the following works [33,34].

Finally, as shown in Eq. (9), there are a number of hyperparameters to learn. Again, a type-II maximum likelihood approach as detailed in Appendix A can be followed.

3. Physical constraints for acoustic emission time of flight mapping

This section will introduce acoustic emission difference-in-time of arrival mapping, and how physically-relevant constraints may be incorporated into models that utilise such features. The data set used throughout the paper will then be detailed, which consists of ΔT measurements from a plate structure. The structure itself presents a challenging wavefield to model, containing a series of holes cut through the plate that adds significant complexity to the wave propagation behaviour. The section then concludes with a discussion on how constraints can be implemented.

3.1. Acoustic emission onset time mapping

Acoustic emissions characterise ultrasonic signals that are released as a structure undergoes some internal change. Often these changes are initiated by mechanisms such as crack propagation, spalling and delamination, making acoustic emission measurements highly suitable for use as features in damage monitoring strategies. Given that the time taken for an AE signal to arrive at a receiving sensor will be dependent on the distance travelled, it is possible to use acoustic emission measurements as a means for localisation [6,10]. In particular, methods that view AE localisation as a problem of spatially mapping ΔT information have been proven to perform well in challenging localisation environments [11,12,14]. There, some regression algorithm is first employed to learn spatial variations in ΔT for a structure of interest, where inputs to the model are a coordinate location with targets the corresponding ΔT , for a given sensor pairing. An estimate for the source origin of a future AE event is then inferred as the location that minimises the difference between the observed and predicted ΔT [12,35], or by considering the source location likelihood as part of a probabilistic framework providing that the spatial map has been learnt as a distribution [11]. In the latter case, not only is a maximum likelihood estimate of the source location provided, but also an associated confidence, which may be used to inform a maintenance engineer how large of an area to inspect, for example, in a wind turbine bearing [36].

3.2. Experimental case study

In this work, we will consider learning ΔT maps for a complex plate structure. First used in the work of Hensman et al. [7], the plate was manufactured to contain a series of holes to replicate the challenges found in real engineering infrastructure. These holes, which can be seen on the schematic and image of the plate in Fig. 1, induce a number of complex phenomena such as scattering and wave mode conversion. For large areas of the plate, a direct propagation path for many of the sensor pairs is also blocked, adding further complexity.

The dataset acquired in [7] is used for the entirety of this work. In the interest of conciseness, experimental details provided will be limited to key details. Should the reader be interested in a complete description of the experimental and data acquisition procedure, they should consult the original paper. However, in brief, artificial AE events were first generated in a uniform grid across the surface of the plate, where thermoelastic expansion induced by a high-power laser was used to excite the AE, recorded by eight piezoceramic Sonox P5 sensors. The time of arrival of the first arriving wave mode of every event was then estimated through the standard practice of implementing the Akaike information criterion (AIC) [7,37], returning an eight-dimensional onset time vector for each artificial excitation. Given that there are 28 possible pair-wise combinations, it is then trivial to transform this vector into a 28-dimensional vector of ΔT values. Note that although AE was artificially generated for the work presented here, the proposed methodology is appropriate for any type of AE source provided that a time of arrival can be extracted. As an example of the feature set, the full set of true ΔT values, which form the targets of the GP, for sensor pair 3 & 5 is shown in Fig. 2, consisting of 2277 measurement locations, each spaced 5 mm apart except for where holes in the plate are present.

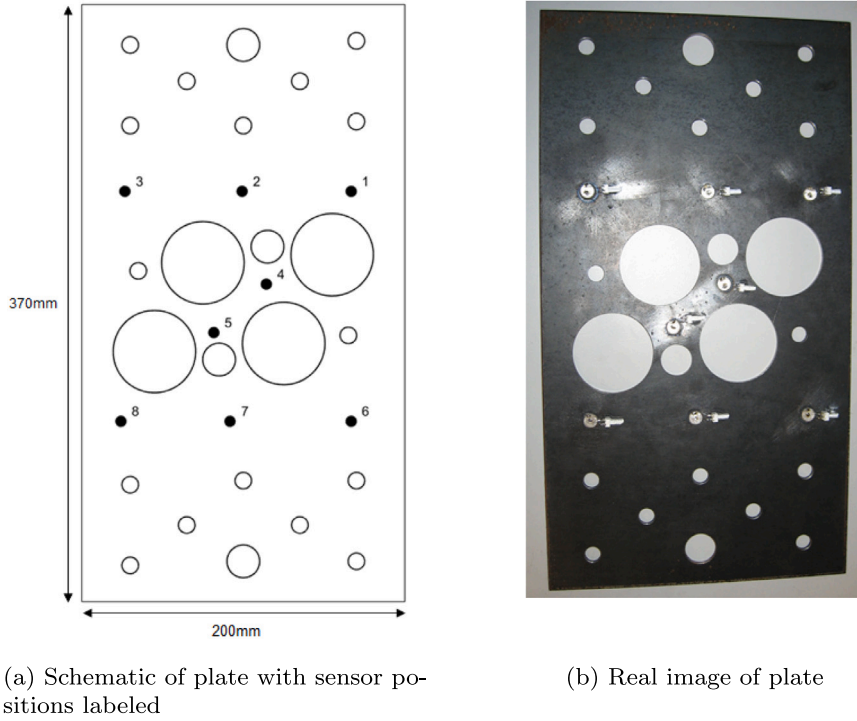


Fig. 1. Plate structure used as a case study.
Source: Recreated from [7]

3.3. Implementing constraints

To implement the constrained GP, it is firstly required that a physical domain is specified, and then for the associated Laplacian eigenfunctions to be solved. As these eigenfunctions are unique to a given domain, it is possible to directly encode boundary condition knowledge into the model. As we are using a Laplacian approximation, from the perspective of a dynamicist, this process can be seen as analogous to finding the wavenumbers and approximate mode shapes of the plate. For simple geometries, it is possible to arrive at closed form solutions for the Laplacian eigenvalues. However, due to the geometrical complexity of the plate, in this work the eigendecomposition is computed numerically. This is calculated by approximating the operator with a finite difference equation that is solved alongside the boundary conditions, with each boundary condition giving an equation that can be solved simultaneously with Eq. (7). To implement this numerical approximation, the Laplacian operator is first converted into its discrete counterpart by transforming the domain into a *grid mask*, u , which exists as a binary matrix where ones denote locations inside the domain, whilst zeros indicate the opposite. A discrete representation of the Laplacian can then be formed as a stencil matrix by applying a finite difference approximation of the Laplace operation on u ,

$$-\nabla^2 u(i, j) \approx \frac{1}{h^2} (-4u_{i,j} + u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}), \quad (14)$$

where (i, j) index the rows and columns of the grid mask, and h represents the step size between adjacent nodes within the grid. Eq. (14) can then be manipulated to reflect known information solution information in the form of boundaries conditions where (i, j) lie on the boundary of the grid mask. For example, in the case of Dirichlet conditions (process value is specified), the solution can be fixed at the boundary positions. For the onset time functions, the associated boundary condition is that of a first order spatial derivative equal to zero (Neumann boundary conditions). Eq. (8) can, therefore, be rewritten as

$$d\phi_j(\mathbf{x}) = 0 \quad \mathbf{x} \in \delta\Omega. \quad (15)$$

Following the construction of a numerical approximation of the negative Laplacian of Ω , for which the procedure of solving for first order boundary conditions is detailed in Appendix B, the leading m eigenvalues and eigenfunctions can then be calculated through a chosen numerical solver. For the work conducted in this paper, following [38] for large-scale spatial mapping problems, 256 eigenbases are used. The first 16 of these basis functions that incorporate all physical boundaries present on the plate are shown in Fig. 3. Note that although we only consider homogeneous first order derivative boundaries, it is possible to include other forms of boundary condition through proper treatment of Eq. (14) at boundary locations.

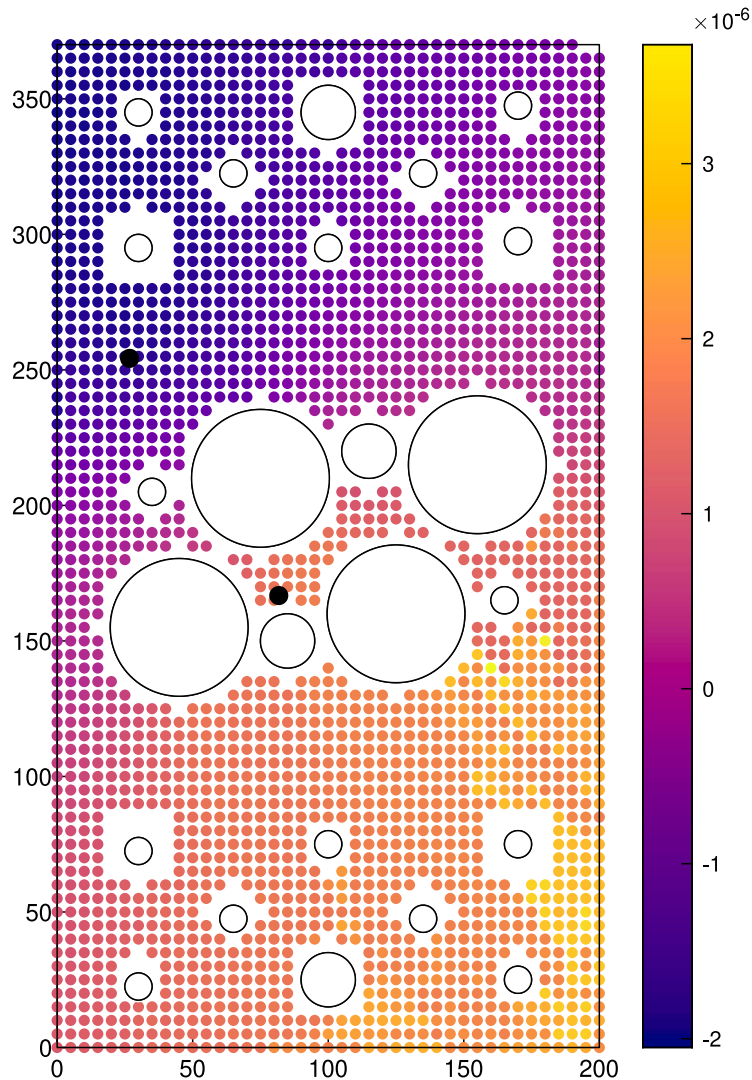


Fig. 2. True difference-in-time-of arrival (ΔT) values for sensor pair 3 & 5 across all of the positions on the plate that data was acquired. Sensor locations highlighted by the black circles.

4. Results and discussion

The feasibility of acoustic emission localisation for large and complex systems is severely hampered by the need to collect artificial AE events at locations on a dense grid across the structure. To explore how the constrained GP may help to alleviate this burden, in this section, results are shown that investigate scenarios where the number and location of training points are limited. To mimic the likely availability of training data from a measurement campaign on a real structure, we consider firstly the case where training measurements are available across the structure but with limited grid density. For most structures, however, particularly those with many connecting components, it is unlikely that access to the whole structure would be available to establish a training dataset (e.g. where access is obscured, or between closely spaced components). The second scenario investigated, therefore, limits the training dataset to a single part of the plate.

The investigation will compare the performance of the standard and constrained GPs. Naturally, the availability of measurements themselves from the boundary for GP conditioning will affect the performance of both methods. We quantify this by explicitly considering additional measurements at the available boundary for both scenarios. In the first case, where measurements are available across the full structure, we expect that the standard GP will outperform the constrained model when training grids are dense — the constrained GP is after all an approximation. We expect to see the benefit of the constrained model where training data are sparse and when data are only available from limited locations across the structure (as will likely be the case in operation).

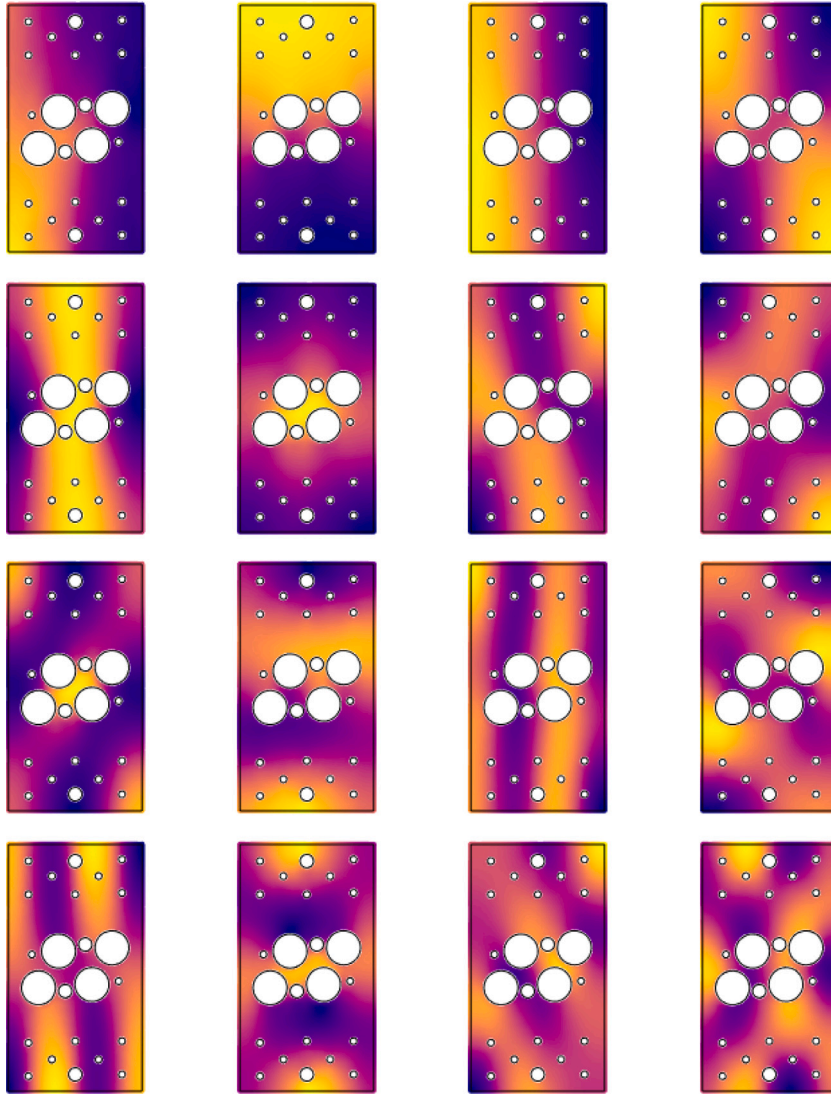


Fig. 3. First 16 Laplacian eigenfunctions (left to right) of the plate detailed in Section 3.2.

As a measure of model predictive performance, the normalised mean square error (nMSE) is considered across a test set of all 2277 data points, collected at a uniform spacing of 5 mm across the whole plate. The nMSE is defined as

$$nMSE = \frac{100}{N\sigma_y^2}(\mathbf{y}_* - \mathbf{y}^*)^T(\mathbf{y}_* - \mathbf{y}^*), \quad (16)$$

where N is the number of points, σ_y^2 is the variance of the true targets, \mathbf{y}_* are the model predictions, with \mathbf{y}^* the true targets. A score of 0 would be returned if the model predictions perfectly aligned with the true targets, whilst a score of 100 is identical to predicting at the mean.

4.1. Sparse training data available across the whole structure

When undergoing an AE data collection campaign, it is often not practical to collect a dense grid of training observations, particularly for larger structures or when setting up multiple monitoring regions. To consider how the constrained GP performs where the training data availability is reduced, a number of training sets containing varying numbers of observations are formed, with measurements available across the full spatial limit of the structure. For an individual set, the spacing between training points is uniform (excluding where the holes are located), with the value that the spacing is fixed at varying across the sets. The total range of training sets considered is outlined in Table 1.

Table 1
Spacing between observations and corresponding total number of datum points for each training set used.

Spacing between training points (mm)	10	15	20	25	30	40	50	60	70
Number of observations in training set	600	302	160	107	71	50	32	14	11

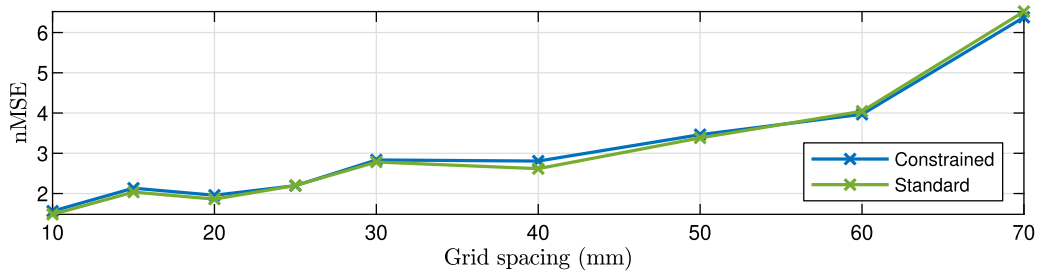


Fig. 4. nMSE of the predictions made by the constrained and standard Gaussian process as the number of training points are reduced. Here the training points fully cover the plate, but at different grid densities.

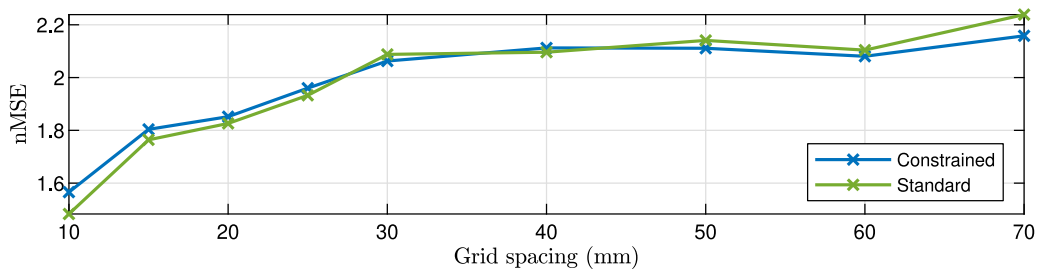


Fig. 5. nMSE of the predictions made by the constrained and standard Gaussian process as training points are reduced whilst retaining full boundary measurements.

The nMSE returned on the test set for each training set, averaged over all available sensor pairs, is plotted on Fig. 4 respectively for both the constrained and standard GP.

The figure demonstrates that as the number of training points reduces to the larger grid spacings, the constrained GP begins to offer a slightly greater accuracy. At the lower end of the grid spacings where the training set is denser, it can be seen that the standard GP is marginally favourable, which as introduced above, is expected. Where training data coverage and availability is good, sufficient boundary condition insight can be obtained from training measurements. That is not to say, however, that one should always seek to retain the full covariance where data are easily assessable. When computing predictions with the constrained GP, computational complexity reduces from $\mathcal{O}(n^3)$ to $\mathcal{O}(nm^2)$, with storage requirements moving from $\mathcal{O}(n^2)$ to $\mathcal{O}(nm)$. A benefit here, therefore, is that when the number of training points exceed several thousand, the use of the constrained GP presents a practical solution [39] without turning to more complex computing techniques such as parallelisation and/or the use of graphical processing units.

In the above example, it was assumed that boundary measurements were available inline with the overall training grid density. For example, at a spacing of 20 mm, boundary measurements were available every 20 mm. It may arise, however, that one would wish to gather more insight in regions on or around boundaries. For example, when mapping an acoustic emission wavefield, it is likely that one would want more insight around boundaries of the domain, where sharp discontinuities will be introduced into the propagation pattern. The first option available in this scenario would be to collect more measurements at the boundary location, which we consider by repeating the above experiment, but ensure that boundary measurements are available every 10 mm.¹ This results in a scenario where one may take a fairly sparse grid of training measurements, but adopt a more fine resolution at boundary locations. Clearly, it is also possible to combine more training measurements with physical constraints. Fig. 5 plots the results for both of these cases.

¹ For the outside boundaries, measurements were taken directly on the boundary. For the inside boundaries, measurements were not available directly on the boundary. Therefore, inside boundary measurements were defined as data points directly adjacent to a given hole in line with a 10 mm grid spacing.

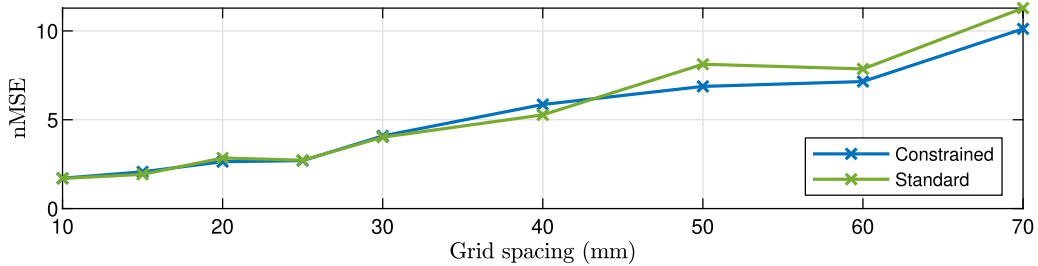


Fig. 6. nMSE of the predictions made by the constrained and standard Gaussian process as training points are reduced with no boundary measurements.

It can be seen that both models obtain similar error scores, with some improvement returned by the constraints where training data becomes more sparse (note the difference in the range of y axis values between Figs. 4 and 5). In the case that a measurement campaign has been specifically conducted with both extra measurements at the boundary and a dense training grid, then the benefit of using the constrained GP from the perspective of mean predictive accuracy is negligible. There are, however, a number of disadvantages to simply adding more training points at the boundary. As the structures we wish to represent become more complex, particularly when in two or three dimensions, the number of datum points required to sufficiently capture a continuous domain will quickly grow. Given the cubic and quadratic scaling of complexity and storage respectively for standard GPs, prohibitive computational demands can quickly be reached, and so for large or complex structures that will demand many boundary measurements, the constrained GP will be a more feasible solution. An additional limitation when collecting data is that many engineering structures simply prevent acquisition at boundaries, for example, at joints and connections that physically obscure generating an artificial signal at that location. We, therefore, now examine a second scenario, where one has access to no boundary measurements, repeating the procedure in the preceding paragraph, but removing all boundary locations from the training set. The predictive performance on the test set for both forms of model is plotted in Fig. 6. The figure demonstrates an improved performance from the constrained GP as training data become fewer. The significance here however, is that performance gain of the constrained GP at larger grid spacings is higher than that obtained in Figs. 4 and 5, where the error returned by both models is comparable for the denser training sets. As the standard GP now has no knowledge of boundary conditions, it is forced to predict at boundary locations with little information if the training grid is not dense. In the case of the constrained GP, the constraints provide the kernel function additional information to the training measurements, incorporating physically relevant structure into the covariance that can then be used when predicting at locations on and adjacent to boundary locations.

Overall, what can be deduced from the results in this section is that when fewer training points are available, the inclusion of boundary constraints improve the predictive capabilities of the model, particularly where boundary measurements are sparse or unavailable. Whilst there is no guarantee that an improved predictive accuracy will be obtained where training data are in more of an abundance, particularly at the boundaries, the use of the constrained GP will still be a consideration when one is limited by the computational demands of calculating the predictive equations in closed form through the standard GP implementation.

4.2. Training data from partial structure coverage

When undergoing an SHM data collection campaign, as previously discussed, it is often not possible to collect data across the entire input space. In a spatial mapping context, this limitation generally arises through being unable to collect data that fully covers the entire structure of interest. For example, when acquiring the artificial AE events that are used to learn a ΔT mapping, it may not be possible to gain full access of the structure, particularly when a health monitoring system is being retrofitted. For example, such a scenario may arise if trying to collect data from the drive train of a wind turbine gearbox, where the assembly of various interlocking gears and shafts will obscure access to many of the individual components that one may be interested in developing an AE mapping for. The fuselage of an aircraft is another example where full access is prevented without disassembly, particularly in the areas where the root of the wings are mounted.

To explore how the constrained GP can mitigate against a lack of training data coverage, the data points used to train the models are now restricted to the middle section of the plate. The three training conditions with respect to the inclusion of boundary points in the training set are again considered; these are (a) full boundary coverage, (b) partial boundary coverage (inline with the overall training grid density), and (c) no boundary measurements. Predictive performance on the test set for each of these three conditions for both the constrained and standard GP is shown in Fig. 7.

Across all three cases, the constrained GP offers either an improved or comparable predictive performance at all of the training point spacings. For the increased grid spacings, particularly where partial or no boundary measurements are available, a large performance gain is obtained by constraining the GP, regularly exceeding an nMSE reduction of greater than 10%.

To further investigate the performance of the constrained GP, we now consider just a single grid spacing of 30 mm as an example, where no boundary measurements are available (case c in Fig. 7). Case (c) was chosen as it presents the most challenging learning task. At a particular grid spacing, as there are a total of 28 sensor pair combinations, there are thus 28 total feature maps to learn. Fig. 8 plots the nMSE obtained on the test set for each of the individual sensor pairing. For reference, Appendix C lists the sensor pair numbers with the corresponding index used here.

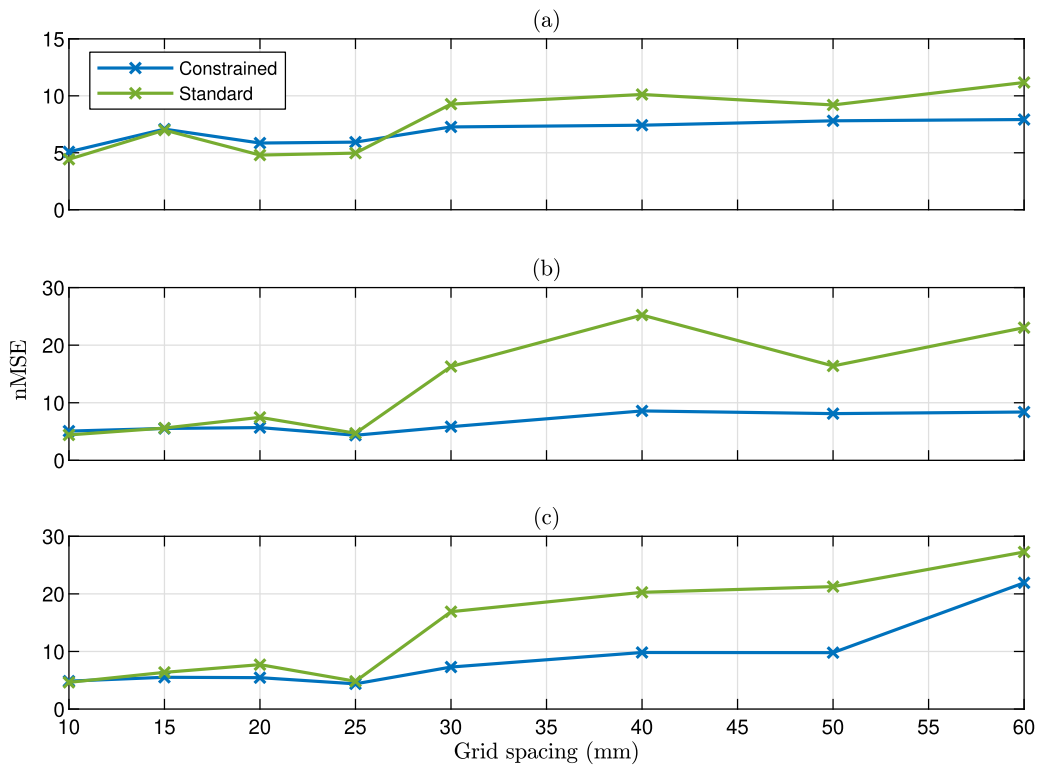


Fig. 7. Predictive performance with restricted training coverage for (a) full boundary location measurements (b) partial boundary measurements (c) no boundary measurements.

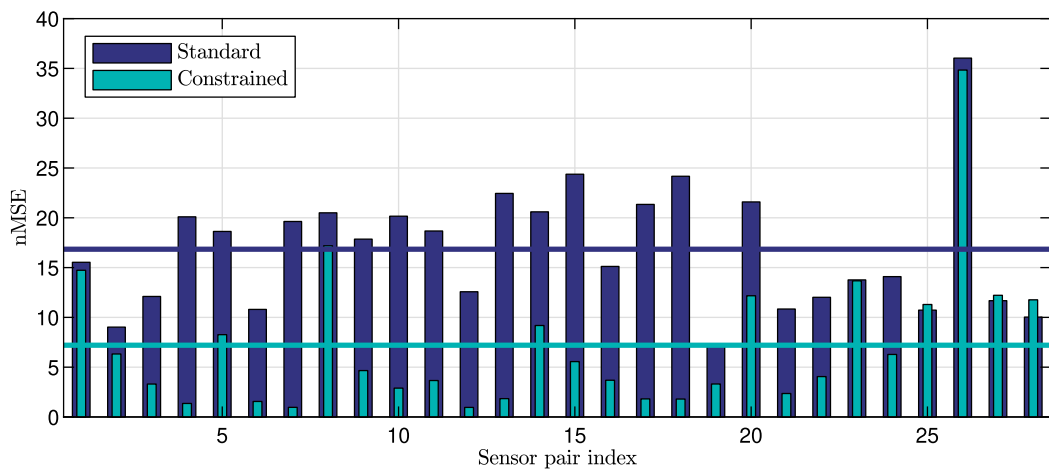


Fig. 8. nMSE returned by each sensor pair model on a training set with limited coverage (case c). Horizontal lines correspond to the average of the standard and constrained nMSE across all sensor pairs.

As an initial observation, it can be seen that for the majority of the sensor pairs, the predictions returned by the constrained GP are vastly improved. This can be quantified formally by considering the averaged nMSE across all of the sensor pairs for both models, where the constrained GP yields an averaged error of 7.30 in comparison to 16.91 from the standard GP. Considering examples of the constrained GP performing significantly better, sensor pairing 15, which corresponds to sensors 3 & 5, displays a large difference in error between the two models. Fig. 9 maps the mean predictions on the test set for both of the GP models. Comparing both of

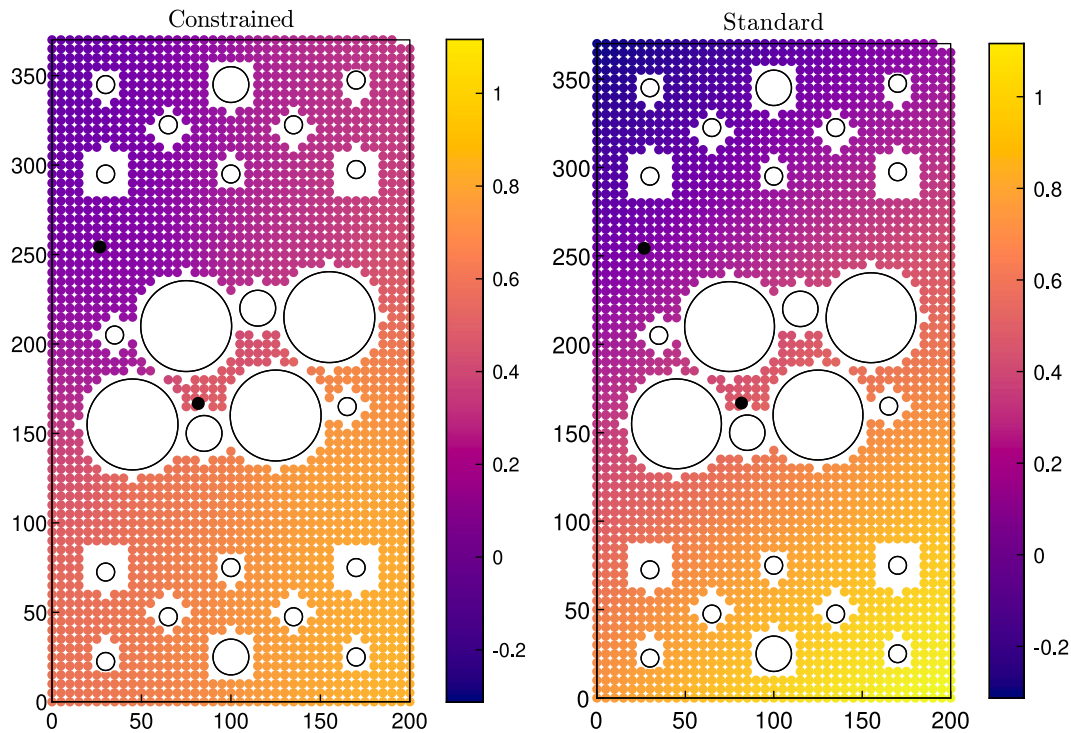


Fig. 9. ΔT predictive mean of both GP models on the test set for sensor pair 3 & 5. Sensor locations are highlighted by the black circles.

the plots, it can be seen that most of the variation between the ΔT predictions exists in the upper and lower part of the plate, away from the location of the training data. Where predictions are made closer to the training points, such as in the centre of the plate, both models return similar predictions.

As the GP learns a distribution at each test location, as well as a predictive mean, a predictive variance is also computed. This allows uncertainty to be quantified across the prediction space, which is often very desirable in SHM, offering a deeper level of insight to feed forward into assessments made regarding damage. Continuing with the analysis of sensor pair 3 & 5, Fig. 10 shows the predictive variance returned by both the standard and constrained GP on the test set. The results show that the constraints embedded into the GP generally reduce the uncertainty of the predictions made across the testing set compared to the standard GP, particularly in the upper and lower region of the plate. It is also possible for the predictive variance to be used to compute a log loss, providing a probabilistic error measure for the predictions made. However, as each of the training sets consist of uniformly spaced observations, differences in the predictive variance between the constrained and unconstrained model will generally only occur around the boundaries. When computing the log loss for each training scenario considered in this section so far, a similar trend will, therefore, be observed to that obtained for the nMSE. As such, log loss plots are not included in the main text, but for completeness, are provided in Appendix D.

Whilst these plots illustrate the learnt distribution over the AE features, they do not provide a direct measure of how well the predictions reflect the true target values. To analyse the discrepancy between the predicted and true targets for both model forms, a mapping of the squared error of the standard GP subtracted from the squared error of the constrained GP is considered. Under this metric, a positive value indicates that the error is larger in the standard GP, whilst negative values express a larger error in the constrained GP. Fig. 11 maps this error metric across the test set for sensor pair 3 & 5.

The figure clearly demonstrates that in the upper and lower segments of the plate, which are the regions away from the training points, the accuracy of the ΔT predictions are improved by the physics-informed GP. As the prediction locations move further away from where the training data points are placed, the level of improvement offered by the constraints generally grows, particularly towards the boundaries at the upper and lower edges of the plate. For regions where training data coverage is good, then the use of boundary constraints will not significantly affect the mean predictions, explaining the similar error score obtained for both models in the centre of the plate. If the testing set contained data points on (or closer to) the inner hole boundaries, then it is likely that an error decrease would have been observed at these positions for the constrained model. However, testing measurements were only collected at a minimum distance of around 5 mm from the inner holes.

A final observation that can be made is that the error reduction obtained in the upper left of the plate ($x = 0, y = 370$) is greater than that returned in the upper right ($x = 200, y = 370$), despite these two areas being geometrically symmetrical. To explain this behaviour, one must first recognise that the true ΔT values in a particular part of the plate will contain a level of variability that is dependent on the complexity of the propagation path between the locations of the sources and the receiving sensor. As the

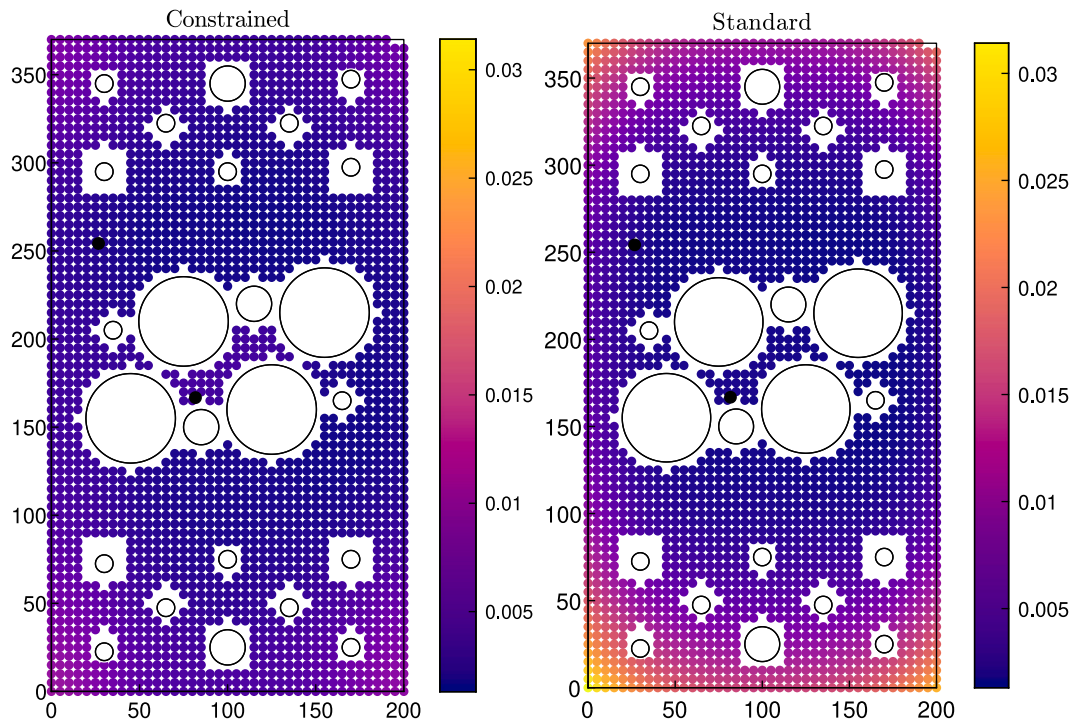


Fig. 10. Predictive variance of both GP models for sensor pairing 3 & 5. Sensor locations are highlighted by the black circles.

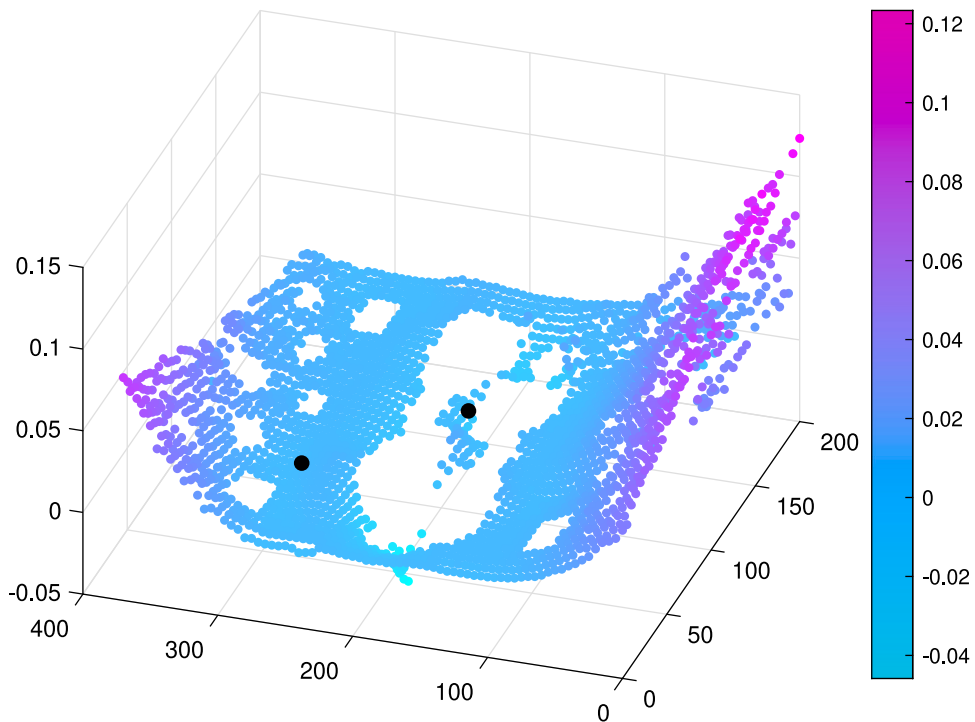


Fig. 11. Squared error difference between standard and constrained GP for sensor pair 3 & 5. Training observations taken from the middle section of the plate.

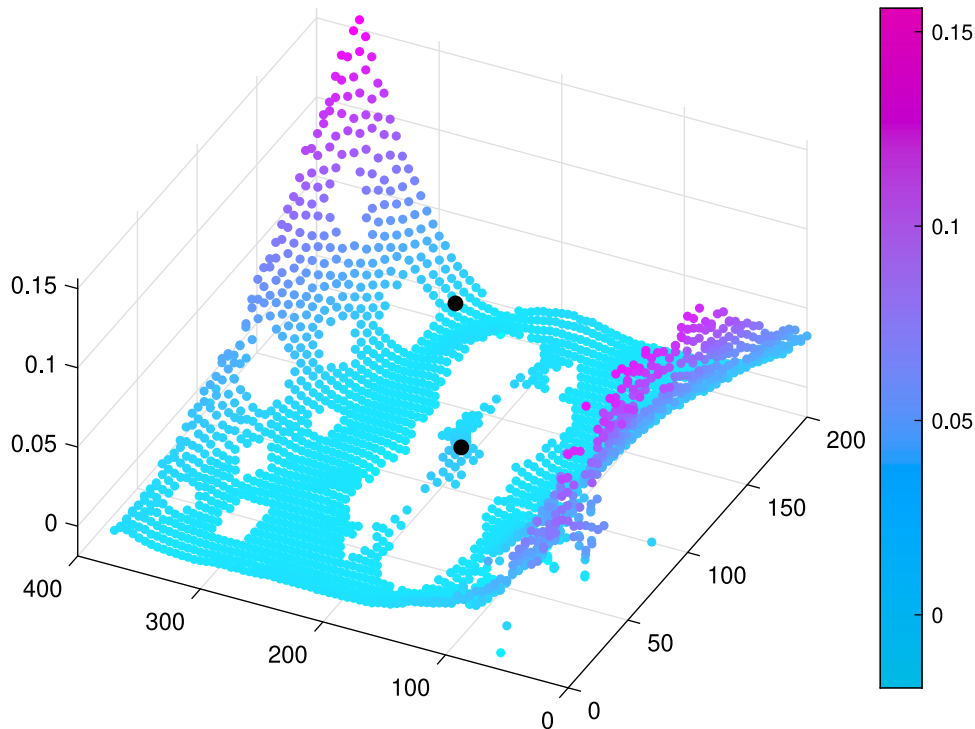


Fig. 12. Squared error difference between standard and constrained GP for sensor pair 1 & 5. Training observations taken from the middle section of the plate.

propagation path becomes more complex, whether that be due to the waves having to travel further to a receiving sensor, or because the propagation path is heavily obstructed (such as by multiple holes), then the variability of the onset times in a given region will increase, resulting in a more challenging feature map to learn in that part of the plate. For sensors 3 & 5, the sensor pairing is positioned closer to the upper left of the plate than the right, requiring AE sources from the upper right region to propagate further to the receiving sensors, leading to a significant increase in the level of variability in the onset times in the upper right region than the upper left. As the constraints implemented here act only in relation to boundary conditions, it should not be expected that they provide a means of capturing this variability that arises from sensor positioning, and as such, both models perform similarly in this region.

Examining a second sensor pair mapping, Fig. 12 plots the difference in square error for sensor pair 4 (sensors 1 & 5). Again, it can be seen that away from the training data, the constraints significantly reduce the error of the ΔT predictions, with the maximum improvement occurring on the upper and lower boundaries of the plate. However, unlike Fig. 11, there is now a significant error reduction on the upper right of the plate, with the previously seen improvement on the upper left now absent. Again, the position of the sensor pair explains this behaviour, with sensor coverage improved in the top right of the plate, but reduced towards the upper left region.

The dependence of the sensor coverage on the predictive performance also explains the reasoning for a number of sensor pairs returning a comparable error metric for the constrained GP with the standard GP. For example, sensors 1 & 2 (index 1) and sensors 6 & 7 (index 26). In these cases, the sensor coverage is poor, with both sensors generally lying adjacent to one another. This positioning results in large portions of the plate requiring waves to propagate further across the structure before being received, resulting in a more complex propagation path and, therefore, more variable ΔT features. If the interest is in improving predictions in locations with reduced sensor coverage, then implementing a constrained machine learner in isolation is not suitable. This is because constrained learners are still reliant on some baseline level of training data; for the constrained GP, the covariance structure of the features still needs to be learnt from input data, which are then used exclusively to make predictions. A potential approach to mitigate the effect of poor sensor coverage will be discussed in the following section, and forms a logical progression for future work.

5. Conclusions

This paper has demonstrated how known boundary conditions may be embedded into a Gaussian process regression model for learning acoustic emission onset time maps. For the time of arrival mapping problem, and also spatial modelling problems more generally, where there exists a lack of boundary measurements or restricted coverage of the total input space, it is shown that constraining the covariance function of the Gaussian process offers a significantly improved predictive performance. Due to the time and cost requirements of acquiring training data for real engineering structures, these are scenarios that consistently arise,

illustrating the benefit of incorporating physical insight into the Gaussian process model through the approach presented in this paper.

Boundary conditions are just one example of insight obtained through an understanding of the underlying physics of the problem that may be harnessed when learning a Gaussian process model. In future work, additional engineering knowledge of the propagation behaviour of the acoustic emissions will be exploited to derive a physics-based mean function, with a view towards improved model performance in even more sparse training data regimes and where sensor coverage may be limited.

CRediT authorship contribution statement

M.R. Jones: Conceptualisation, Methodology, Software, Formal analysis, Writing – original manuscript, Writing – review & editing. **T.J. Rogers:** Conceptualisation, Writing – review & editing. **E.J. Cross:** Supervision, Conceptualisation, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors are grateful for the support of grant reference numbers EP/S001565/1 and EP/R004900/1. Thanks is offered to James Hensman, Mark Eaton, Robin Mills and Gareth Pierce for their work in the generating the data used throughout this paper. The support of Keith Worden during the completion of this work is also gratefully acknowledged.

Appendix A. Marginal likelihood

To determine the hyperparameters, θ , of the covariance, a typical approach is to maximise the log marginal likelihood of the predictions with respect to θ . This is typically framed as a minimisation of the negative log marginal likelihood, which for the standard Gaussian process model, is expressed as,

$$\theta = \arg \min_{\theta} \left\{ \frac{1}{2} \mathbf{y}^T (K_{XX} + \sigma_n^2 \mathbb{I})^{-1} \mathbf{y} + \frac{1}{2} \log |(K_{XX} + \sigma_n^2 \mathbb{I})| + \frac{n}{2} \log(2\pi) \right\}. \quad (\text{A.1})$$

For the constrained GP, there exists a modified negative log marginal likelihood,

$$\begin{aligned} \theta = \arg \min_{\theta} \left\{ \frac{1}{2} (N - m) \log \sigma_n^2 + \frac{1}{2} \sum_{j=1}^m \Lambda_{j,j} + \frac{1}{2} \log |\sigma_n^2 \Lambda^{-1} + \Phi^T \Phi| + \frac{n}{2} \log(2\pi) \right. \\ \left. + \frac{1}{2\sigma_n^2} (\mathbf{y}^T \mathbf{y} - \mathbf{y}^T \Phi (\sigma_n^2 \Lambda^{-1} + \Phi^T \Phi)^{-1} \Phi \mathbf{y}) \right\}. \end{aligned} \quad (\text{A.2})$$

For optimising both the standard and constrained covariance hyperparameters, a Quantum-behaved Particle Swarm Optimisation (QPSO) [40,41] is employed, a gradient-free heuristic optimiser.

Appendix B. Laplace approximation with derivative boundary conditions

Solving for first order derivative boundary conditions is slightly more involved than Dirichlet conditions and requires the use of a first order approximation at the boundary location. However, before incorporating this boundary type, a series of “ghost points” must be added to every location directly adjacent to a boundary. These ghost points are purely artificial, and exist to help implement the finite difference approximation at endpoints of the domain. As an example, take the arbitrary 2D domain shown in Fig. B.1(a), where each circle represents a node in the grid mask. Consideration of the boundary on the left hand side of the domain along the j direction requires the addition of ghost points at $j = 0$, as shown in Fig. B.1(b).

Incorporation of a first order derivative at $j = 1$ can then be achieved by applying a backward difference approximation

$$u'_j = \frac{u_{i,j} - u_{i,j-1}}{h}, \quad (\text{B.1})$$

which for zero derivative boundary conditions, can be simplified to,

$$u_{i,j} = u_{i,j-1}. \quad (\text{B.2})$$

At this point, it is worth noting that although a higher order approximation would be obtained through the use of a central difference method, the use of a backward (and forward, as discussed shortly) difference scheme ensures that the resultant stencil matrix is

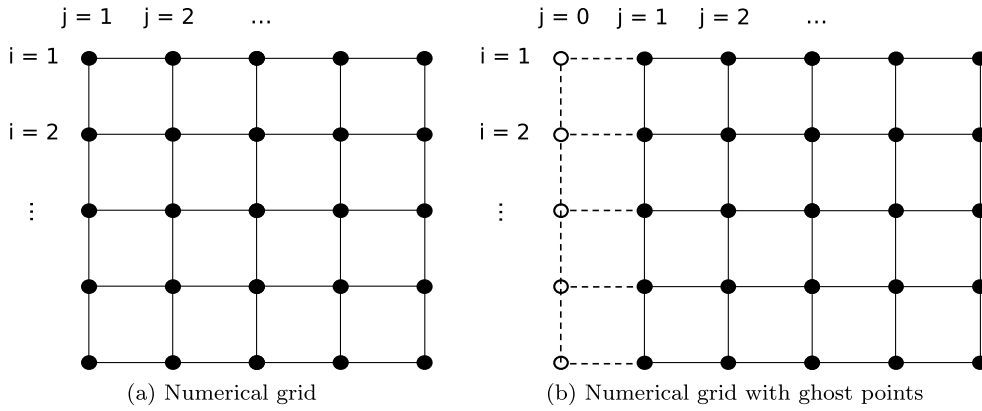


Fig. B.1. Numerical grid for arbitrary domain to depict how ghost points are introduced to include derivative boundary conditions.

Table C.1
Sensor pair index and corresponding individual sensors.

Sensor pair index	Sensors	Sensor pair index	Sensors
1	1-2	15	3-5
2	1-3	16	3-6
3	1-4	17	3-7
4	1-5	18	3-8
5	1-6	19	4-5
6	1-7	20	4-6
7	1-8	21	4-7
8	2-3	22	4-8
9	2-4	23	5-6
10	2-5	24	5-7
11	2-6	25	5-8
12	2-7	26	6-7
13	2-8	27	6-8
14	3-4	28	7-8

symmetric [42], and so the corresponding eigenvectors are real. Returning to Eq. (14), at $j = 1$, the ghost points that appear at $u_{i,0}$ can be removed by substituting in the above expression, yielding,

$$-\nabla^2 u(i, 1) \approx \frac{1}{h^2} (-3u_{i,1} + u_{i-1,1} + u_{i,2}), \tag{B.3}$$

and incorporating the boundaries into the stencil matrix. Where a ghost point lies at an index of +1 to the boundary, such as the right hand side of the domain in Fig. B.1(a), then it is necessary for a forward difference approximation to be applied

$$u'_j = \frac{u_{i,j+1} - u_{i,j}}{h}, \tag{B.4}$$

replacing Eq. (B.2) with,

$$u_{i,j+1} = u_{i,j} \tag{B.5}$$

The process detailed above is then identical for any given direction, providing the finite difference equation applied is consistent with the target direction. For example, in the i direction, Eqs. (B.2) and (B.5) become

$$u_{i,j} = u_{i-1,j} \tag{B.6}$$

$$u_{i,j} = u_{i+1,j}. \tag{B.7}$$

Iterating through each element of the grid mask, a stencil matrix can be computed that corresponds to the negative Laplacian of Ω . The leading m eigenvalues and eigenfunctions can then be calculated through a chosen numerical solver.

Appendix C. Sensor pair index

See Table C.1.

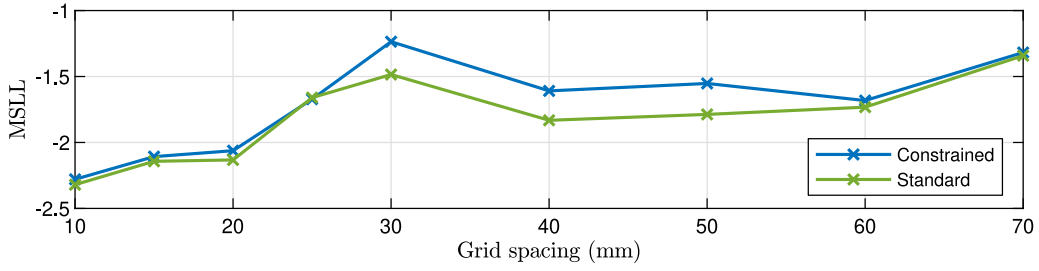


Fig. D.1. MSLL of the predictions made by the constrained and standard Gaussian process as the number of training points are reduced. Here the training points fully cover the plate, but at different grid densities.

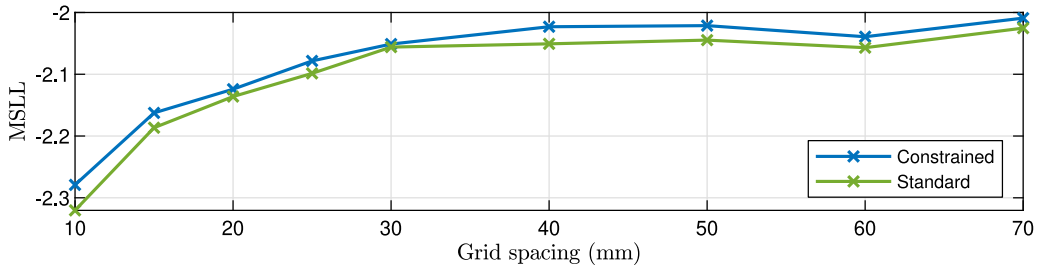


Fig. D.2. MSLL of the predictions made by the constrained and standard Gaussian process as training points are reduced whilst retaining full boundary measurements.

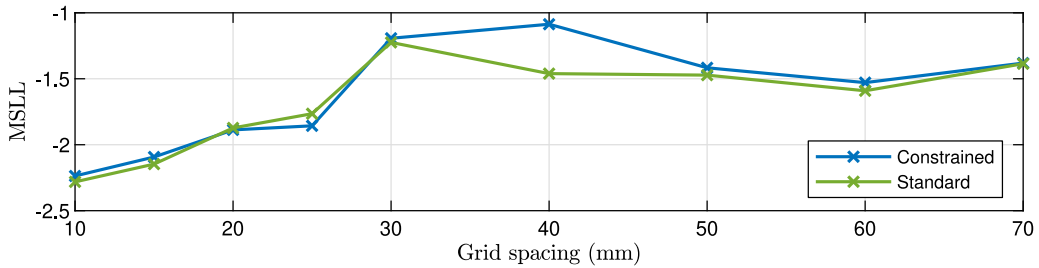


Fig. D.3. MSLL of the predictions made by the constrained and standard Gaussian process as training points are reduced with no boundary measurements.

Appendix D. Log loss results

The mean standardised log loss (MSLL) is expressed as,

$$MSLL = \frac{1}{N} \sum_k \{-\log p(\mathbf{y}_{\star,k} | X, \mathbf{y}, \mathbf{x}_{\star,k}) + \log p(\mathbf{y}_{\star,k}; \mathbb{E}(\mathbf{y}_k), \mathbb{V}(\mathbf{y}_k))\}, \tag{D.1}$$

where k indexes a particular test point. The log loss can be interpreted as the negative likelihood of the predictions relative to those made under the trivial model, i.e. the mean and variance of the training observations. As such, a larger negative MSLL reflects more favourable predictive distributions.

For each of the training scenarios considered in Section 4 (e.g. Figs. 4–7), equivalent MSLL plots are provided below (see Figs. D.1–D.4).

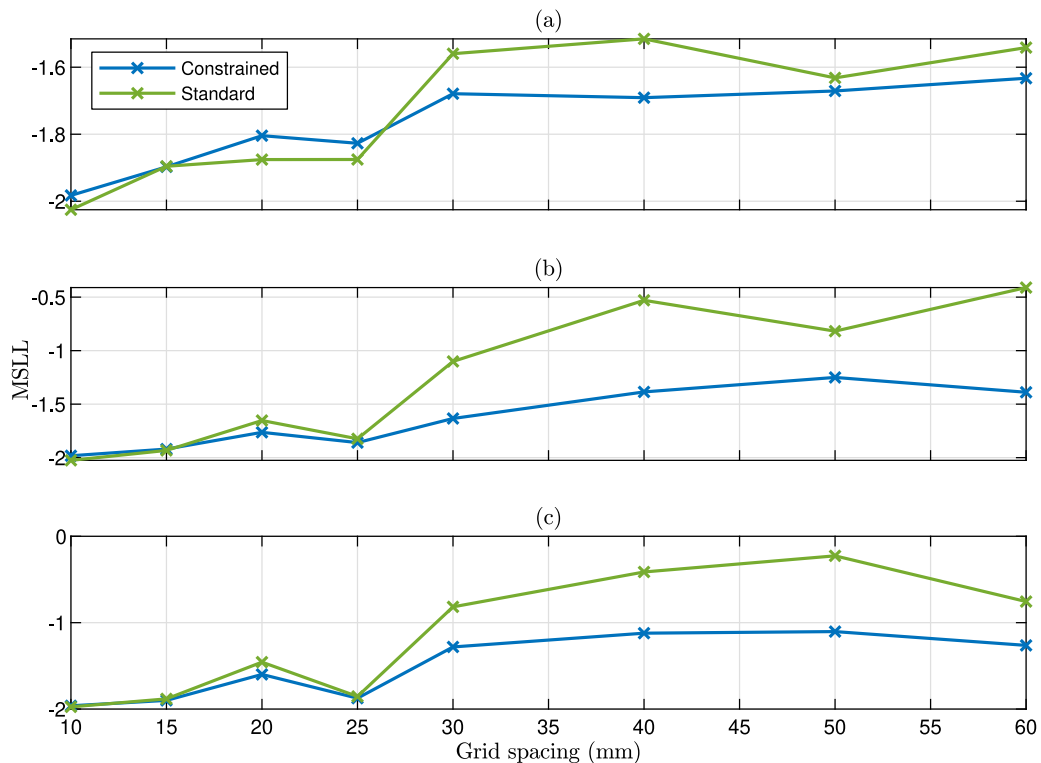


Fig. D.4. MSLL with restricted training coverage for (a) full boundary location measurements (b) partial boundary measurements (c) no boundary measurements.

References

- [1] C.R. Farrar, K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*, John Wiley & Sons, 2012.
- [2] T. Rogers, K. Worden, R. Fuentes, N. Dervilis, U. Tygesen, E. Cross, A Bayesian non-parametric clustering approach for semi-supervised structural health monitoring, *Mech. Syst. Signal Process.* 119 (2019) 100–119.
- [3] L. Bull, K. Worden, G. Manson, N. Dervilis, Active learning for semi-supervised structural health monitoring, *J. Sound Vib.* 437 (2018) 373–388.
- [4] L.D. Avendaño-Valencia, E.N. Chatzi, K.Y. Koo, J.M. Brownjohn, Gaussian process time-series models for structures under operational variability, *Front. Built Environ.* 3 (2017) 69.
- [5] E.J. Cross, *On Structural Health Monitoring in Changing Environmental and Operational Conditions* (Ph.D. thesis), University of Sheffield, 2012.
- [6] A. Tobias, Acoustic-emission source location in two dimensions by an array of three sensors, *Non-Destructive Testing* 9 (1) (1976) 9–12.
- [7] J. Hensman, R. Mills, S. Pierce, K. Worden, M. Eaton, Locating acoustic emission sources in complex structures using Gaussian processes, *Mech. Syst. Signal Process.* 24 (1) (2010) 211–223.
- [8] F. Ciampa, M. Meo, Impact detection in anisotropic materials using a time reversal approach, *Struct. Health Monit.* 11 (1) (2012) 43–49.
- [9] E.D. Niri, A. Farhidzadeh, S. Salamone, Nonlinear Kalman filtering for acoustic emission source localization in anisotropic panels, *Ultrasonics* 54 (2) (2014) 486–501.
- [10] T. Kundu, Acoustic source localization, *Ultrasonics* 54 (1) (2014) 25–38.
- [11] M.R. Jones, T.J. Rogers, K. Worden, E.J. Cross, A Bayesian methodology for localising acoustic emission sources in complex structures, *Mech. Syst. Signal Process.* 163 (2022) 108143.
- [12] M.G. Baxter, R. Pullin, K.M. Holford, S.L. Evans, Delta T source location for acoustic emission, *Mech. Syst. Signal Process.* 21 (3) (2007) 1512–1520.
- [13] M.R. Pearson, M. Eaton, C. Featherston, R. Pullin, K. Holford, Improved acoustic emission source location during fatigue and impact events in metallic and composite structures, *Struct. Health Monit.* 16 (4) (2017) 382–399.
- [14] A. Ebrahimkhanlou, S. Salamone, Single-sensor acoustic emission source localization in plate-like structures using deep learning, *Aerospace* 5 (2) (2018) 50.
- [15] J. Willard, X. Jia, S. Xu, M. Steinbach, V. Kumar, Integrating physics-based modeling with machine learning: A survey, 2020, arXiv preprint arXiv:2003.04919.
- [16] E.J. Cross, S.J. Gibson, M.R. Jones, D.J. Pitchforth, S. Zhang, T.J. Rogers, in: A. Curry, D. Ribeiro, F. Ubertini, M. Todd (Eds.), *Structural Health Monitoring based on Data Science Techniques*, Springer - IN PRESS, 2022.
- [17] C.K. Williams, C.E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 3, MIT press Cambridge, MA, 2006.
- [18] L.P. Swiler, M. Gulian, A.L. Frankel, C. Safta, J.D. Jakeman, A survey of constrained Gaussian process regression: Approaches and implementation challenges, *J. Mach. Learn. Model. Comput.* 1 (2) (2020).
- [19] C. Agrell, *Gaussian processes with linear operator inequality constraints*, 2019, arXiv preprint arXiv:1901.03134.
- [20] J. Riihimäki, A. Vehtari, Gaussian processes with monotonicity information, in: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings*, 2010, pp. 645–652.
- [21] S. Da Veiga, A. Marrel, Gaussian process modeling with inequality constraints, in: *Annal. Facult. Sci. Toulouse*, 21, (3) 2012, pp. 529–555.
- [22] H. Maatouk, *Finite-dimensional approximation of Gaussian processes with inequality constraints*, 2017, arXiv preprint arXiv:1706.02178.

- [23] A.F. López-Lopera, F. Bachoc, N. Durrande, O. Roustant, Finite-dimensional Gaussian approximation with linear inequality constraints, *SIAM/ASA J. Uncertain. Quantif.* 6 (3) (2018) 1224–1255.
- [24] A. Pensoneault, X. Yang, X. Zhu, Nonnegativity-enforced Gaussian process regression, *Theor. Appl. Mech. Lett.* 10 (3) (2020) 182–187.
- [25] C. Jidling, N. Wahlström, A. Wills, T.B. Schön, Linearly constrained Gaussian processes, 2017, arXiv preprint arXiv:1703.00787.
- [26] E.J. Cross, T.J. Rogers, Physics-derived covariance functions for machine learning in structural dynamics, *IFAC-PapersOnLine* 54 (7) (2021) 168–173.
- [27] M. Alvarez, D. Luengo, N.D. Lawrence, Latent force models, in: *Artificial Intelligence and Statistics*, PMLR, 2009, pp. 9–16.
- [28] M. Raissi, G.E. Karniadakis, Hidden physics models: Machine learning of nonlinear partial differential equations, *J. Comput. Phys.* 357 (2018) 125–141.
- [29] E.J. Cross, T.J. Rogers, T.J. Gibbons, Grey-box modelling for structural health monitoring: physical constraints on machine learning algorithms, 2019, *Structural Health Monitoring* 2019.
- [30] A. Solin, S. Särkkä, Hilbert space methods for reduced-rank Gaussian process regression, *Stat. Comput.* 30 (2) (2020) 419–446.
- [31] M.L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Science & Business Media, 2012.
- [32] C. Chatfield, H. Xing, *The Analysis of Time Series: An Introduction with R*, CRC Press, 2019.
- [33] D. Pitchforth, T. Rogers, U. Tygesen, E. Cross, Grey-box models for wave loading prediction, *Mech. Syst. Signal Process.* 159 (2021) 107741.
- [34] S. Zhang, T.J. Rogers, E.J. Cross, Gaussian process based grey-box modelling for SHM of structures under fluctuating environmental conditions, in: *European Workshop on Structural Health Monitoring*, Springer, 2020, pp. 55–66.
- [35] S.K. Al-Jumaili, M.R. Pearson, K.M. Holford, M.J. Eaton, R. Pullin, Acoustic emission source location in complex structures using full automatic Delta T mapping technique, *Mech. Syst. Signal Process.* 72 (2016) 513–524.
- [36] M.R. Jones, T.J. Rogers, I.E. Martinez, E.J. Cross, Bayesian localisation of acoustic emission sources for wind turbine bearings, in: *Health Monitoring of Structural and Biological Systems XV*, vol. 11593, International Society for Optics and Photonics, 2021, p. 115932D.
- [37] J.H. Kurz, C.U. Grosse, H.-W. Reinhardt, Strategies for reliable automatic onset time picking of acoustic emissions and of ultrasound signals in concrete, *Ultrasonics* 43 (7) (2005) 538–546.
- [38] A. Solin, M. Kok, Know your boundaries: Constraining Gaussian processes by variational harmonic features, in: *The 22nd International Conference on Artificial Intelligence and Statistics*, PMLR, 2019, pp. 2193–2202.
- [39] M.R. Jones, T.J. Rogers, P. Gardner, E.J. Cross, Constraining Gaussian processes for grey-box acoustic emission source localisation, in: *ISMA 2020 - International Conference on Noise and Vibration Engineering and USD 2020 - International Conference on Uncertainty in Structural Dynamics*, 2020.
- [40] J. Sun, W. Xu, B. Feng, A global search strategy of quantum-behaved particle swarm optimization, in: *IEEE Conference on Cybernetics and Intelligent Systems*, 2004, vol. 1, IEEE, 2004, pp. 111–116.
- [41] K. Worden, R. Barthorpe, E. Cross, N. Dervilis, G. Holmes, G. Manson, T. Rogers, On evolutionary system identification with applications to nonlinear benchmarks, *Mech. Syst. Signal Process.* 112 (2018) 194–232.
- [42] R.J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, SIAM, 2007.