Multifidelity Data Fusion Applied to Aircraft Wing Pressure Distributions

Mehdi Anhichem* and Sebastian Timme†
University of Liverpool, Liverpool, L69 3GH, United Kingdom

Jony Castagna‡
UKRI-STFC Hartree Centre, Warrington, WA4 4AD, United Kingdom

Andrew J. Peace§ and Moira Maina¶
Aircraft Research Association Ltd, Bedford, MK41 7PF, United Kingdom

Designing an aircraft and analysing its performance requires uncertainty-aware and robust aerodynamic data. The three principal ways of acquiring such aerodynamic data are flight testing, wind tunnel testing and numerical analysis. They all can be expensive and are subject to multiple sources of uncertainty. A multifidelity data fusion framework applied to surface pressure data of a large aircraft wing model obtained from experiment and simulation is presented herein with the ambition to enhance its intended use in optimising wind tunnel campaigns and eventually in high-value design. Static pressure tapping and dynamic pressure-sensitive paint are exploited to obtain experimental data sets of different quality and quantity. These are complemented by numerical data ranging from a linear potential panel method to non-linear Reynolds-averaged Navier–Stokes simulations. The data fusion approach introduced by Lam, Allaire and Willcox (2015) and revisited here is non-hierarchical, meaning there is not an absolute hierarchy in terms of accuracy between the information sources. The confidence in an information source over the parameter space is defined through a fidelity function. The approach uses Gaussian processes to enable the fusion of experimental and numerical data into a single multifidelity surrogate model. The generated (and required) volume of data to study the surface flow and distributed aerodynamic loads on an aircraft wing leads to scalability issues. A suitable extension of Gaussian process regression based on stochastic variational inference has been adopted alongside the use of GPU architecture to enable the application of the data fusion framework on large data sets. The approach provides a surrogate model with a quantified uncertainty from uncertainty-aware disparate data sources. We explore (and adapt) the framework for a high-dimensional practical data set generated through industrial wind tunnel testing and numerical flow analysis.

Nomenclature

C_p	=	pressure coefficient
$\mathcal{D}$	=	input space
d	=	dimension of the input space
f	=	scalar quantity of interest
$\mathbf{f}$	=	vector of observation values
$\mathbf{f}^*$	=	prediction vector
I	=	number of information sources
$k(\mathbf{x}, \mathbf{x})$	=	covariance function
l_k	=	lengthscale

*PhD Student, Centre for Doctoral Training in Distributed Algorithms, m.anhichem@liverpool.ac.uk, Student Member AIAA.
†Senior Lecturer, School of Engineering, sebastian.timme@liverpool.ac.uk. Member AIAA.
‡Computational Scientist, jony.castagna@stfc.ac.uk
§Chief Scientist, Senior Member AIAA.
¶Technical Consultant.

$m(\mathbf{x})$ = mean function
$\mathbf{m}$ = variational distribution mean
M = Mach number
N_{ind} = number of inducing points
N = number of training points
p = probability distribution
q = variational probability distribution
$\mathbf{S}$ = variational distribution standard deviation
$\mathbf{u}$ = vector of observation at inducing point locations
$\mathbf{x}$ = input vector
X, Y = Cartesian coordinates
$\mathbf{X}$ = matrix of training points
y = noisy observation scalar function
$\mathbf{y}$ = noisy observation vector
$\mathbf{Z}$ = matrix of inducing points
α = angle of attack
ε = Gaussian noise random variable
η = dimensionless spanwise coordinate
μ_{GP} = Gaussian process mean
$\overline{\mu}$ = multifidelity mean estimate
σ = output variance
σ_{GP} = Gaussian process standard deviation
σ_f = fidelity function
σ_t = total standard deviation
$\overline{\sigma}$ = multifidelity of the total variance
τ = Gaussian noise variance
$\boldsymbol{\theta}$ = hyperparameters vector

I. Introduction

The study of aerodynamics is crucial for estimating aircraft performance characteristics during the various design phases. Aerodynamic analysis and design rely heavily on three main sources of information; flight testing, wind tunnel testing and computational fluid dynamics (CFD). Flight testing is carried out once high confidence in the final design is established due to its extremely high cost. In wind tunnel experiments, a scaled model is manufactured to collect data about the air flow around the model through measurements. Those measurements can, for example, be acquired by pressure taps, force balance or by more advanced optical methods such as pressure sensitive paint and particle image velocimetry. Aerodynamic data can also be computed with deterministic numerical simulations such as using an incarnation of the Navier–Stokes equations. With the ever-increasing computing power, such numerical simulations have become a must in aerodynamics. However, when dealing with complex configurations and phenomena, particularly near the edge of the flight envelope, numerical simulations tend to lack the required accuracy and come with a significant cost burden, particularly when choosing an eddy-resolving approach. It remains essential for aerodynamic analysis and design to carry out physical wind tunnel experiments. Wind tunnel testing can be expensive and, just like numerical analysis, is subject to multiple sources of uncertainty. It may be debated within the aeronautics community that the digital revolution, especially through the introduction of CFD, has not had the expected impact on wind tunnel usage for flight vehicle development programmes [1]. For instance, the dependency on wind tunnel testing for developing military aircraft has not diminished [2]. In practice, CFD has been integrated into the industrial testing routine to complement the theoretical, numerical and experimental methods already in use.

Nevertheless, digital engineering can be considered as an opportunity to rethink current practices and processes. For instance, current practice evaluates the aerodynamic data accuracy by comparison to other available data sets which are all subject to uncertainties of varying nature. It is argued in [2] that this practice lacks relevance in the way of defining and using uncertainties within future design paradigms. Machine learning offers practical and efficient methods for merging the information from different data sets and for incorporating real world variability and probabilistic behaviour into engineering analysis. There is a significant challenge to leverage information from the different aerodynamic data sources while controlling the definition and propagation of uncertainty towards the decision-making level. Uncertainty

quantification takes place when making high-consequence decisions within the risk management step of the design cycle. Thus, our ambition is to advance aerodynamic design of future aircraft through the application of machine learning algorithms that enable experimental and numerical data to be combined while accounting for uncertainty.

Methodologies based on multiple sources of varying fidelity first involved techniques called trust region and variable-complexity modelling. A first-order approximation and model management optimisation based on the trust region approach was developed in [3] where the method was demonstrated on a three-dimensional wing optimisation problem using Euler analysis on meshes of varying refinement to establish a set of variable-fidelity models. Variable-complexity modelling, also called variable-fidelity modelling, uses an additive and/or multiplicative correction factor built with the low- and high-fidelity data to approximate an objective function. Variable-complexity strategy was applied to optimise the wing and fuselage shape and nacelle placement of high-speed civil transport vehicles [4]. A crew transfer vehicle flight envelope was studied in [5] by applying a variable-complexity modelling approach coupled with an increment function to merge solutions of varying fidelity into a single database. The main disadvantages of this method are that it assumes a hierarchy in the levels of accuracy and that the corrections made are linear. This last point is overcome by using a method called kriging, also known as Gaussian process regression [6]. The popularity of Gaussian processes for surrogate modelling is rising in the machine learning community as they are able to model complex, multidimensional surfaces by accounting for the correlation of the data in the design space while providing a measure of confidence in their prediction [7]. For instance, kriging interpolation applied within the Schur complement eigenvalue framework is used in [8] to study transonic aeroelastic instability, while looking into aerodynamic models of varying fidelity. The most common variation when applying Gaussian processes in a data-fusion context is hierarchical kriging, often called co-kriging. Hierarchical kriging is applied in [9] investigating an aerofoil and an industrial transport aircraft configuration. Co-kriging is a hierarchical multifidelity methodology that combines data from two or more models to train a single Gaussian process surrogate model [10–12]. A kriging-based function is used in [13] to correct low-fidelity models with sparse high-fidelity data from an experimental design to study wing optimisation. A radial basis function was used to approximate low-fidelity data and kriging was used as a correction model in order to construct a hybrid variable-fidelity global approximation model in [14]. In [15], Bayesian hierarchical Gaussian process models perform location and scale adjustments to predict high-fidelity output at untried points from a low-fidelity model. Gaussian process models can also be associated with a Bayesian optimisation application, for instance, to evaluate the next design point using the expected improvement criterion [16]. Gaussian processes are a very popular approach for inference on functions. However, its associated cubic complexity and quadratic storage requirements limit its application to relatively small data sets such as information on forces or discrete pressure data over the design space.

Other studies have used the dimensionality reduction technique called proper orthogonal decomposition (POD) for multifidelity surrogate modelling of velocity or pressure fields. It has been used with additional constraints in [17] to improve the overall prediction of the pressure distribution over two aerofoils exposed to noise, bias and incompleteness, in addition to a method based on a Bayesian framework to study the transonic pressure distributions on an aerofoil and the NASA Common Research Model wing. Similar work has been reported in [18] with application in reconstruction of particle image velocimetry data. An extension, called Gappy POD, was explored for data fusion. The principle is to combine the POD basis vectors through the solution of a least-squares problem calibrated with reference data at discrete locations. Variations of this method were used to reconstruct unsteady flow [19, 20] or to fuse CFD with experimental pressure data [21, 22]. In [23], Gaussian process regression was applied to solve the least-squares problem, needed for data reconstruction from the POD modal basis. Therefore, it provides an estimate of the prediction confidence to the regression problem involved in the data reconstruction. The limitations of these techniques stem from the restriction to the data fusion of two information sources, with the experimental data typically taken as 'ground truth' making it difficult to consider its own associated uncertainties. Also, as mentioned earlier, the use of standard Gaussian processes is severely limited by its scalability and therefore restricts its use to discrete data points.

This paper revisits a non-hierarchical multifidelity approach previously presented in [24, 25] where each information source is modelled through a Gaussian process regression and fused in a fidelity-weighted combination. This approach is composed of three main ingredients; the intermediate surrogate model obtained from Gaussian process regression on each information source, the definition of fidelity functions over the input space for each information source and the fidelity-weighted combination of all intermediate models. It is necessary to adapt the regression step to 'big data' applications in order to be able to implement this framework on fine-grained aerodynamic data sets, such as surface pressure distributions from CFD simulations or advanced optical measurement techniques. A method based on stochastic variational inference developed in [26] has been adopted alongside the use of graphics processing unit (GPU) architecture to enable the application of the data fusion framework on large data sets. Each intermediate surrogate model is built independently of the others and in terms of fidelity the model hierarchy varies in the input space. It allows

the associated framework to be flexible on the number of information sources and to benefit from each model's strengths. Our contribution lies in the assessment of the multifidelity data fusion method on higher dimensional problems by applying this approach to the pressure distribution over a wing. Specifically, the information sources considered here provide pressure data from static pressure taps, time-averaged dynamic pressure-sensitive paint, Reynolds-averaged Navier–Stokes simulations and a potential-flow panel method. We also analyse the formulation of the Gaussian process in the construction of the intermediate surrogate models, particularly in terms of scalability.

The different ingredients of the multifidelity approach are described in Section II. The test case and the data used herein are given in Section III, specifically the RBC12 half wing-fuselage configuration [27] and four different sources that will be used to apply multifidelity data fusion on the pressure distribution. Section IV discusses results of the application of the multifidelity data fusion framework at fixed and variable flow conditions.

II. Multifidelity Surrogate based on Gaussian Processes

A. Gaussian process regression

A Gaussian process is a stochastic process with Gaussian finite dimensional distributions. If a design space $\mathcal{D} \subseteq \mathbb{R}^d$ of dimension d and an input vector $\mathbf{x} \in \mathcal{D}$ are considered, the function $f : \mathbf{x} \in \mathcal{D} \longmapsto f(\mathbf{x}) \in \mathbb{R}$ represents a scalar quantity. According to [7], a Gaussian process $\mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ is characterised by the mean function $m(\mathbf{x})$ and the covariance function $k(\mathbf{x}, \mathbf{x}')$. In some cases, noisy observations are considered and written $y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon$. The noise is an independent identically distributed Gaussian variable ε with zero mean and variance τ^2. Let $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ ...\ \mathbf{x}_N]$ be the matrix of N training points $\mathbf{x}_j \in \mathcal{D}$ and $\mathbf{f} = f(\mathbf{X}) = [f_1, f_2, ..., f_N]^\top$, the latent variable vector containing the values $f_j = f(\mathbf{x}_j)$. The observable variable vector is defined in the same way by $\mathbf{y} = y(\mathbf{X}) = [y_1, y_2, ..., y_N]^\top$. The Gaussian process prior, $p(\mathbf{y}, \mathbf{f}^*)$, can be written as a joint normal distribution over the random variable vectors $\mathbf{y}$ and $\mathbf{f}^* = f(\mathbf{X}^*)$ where $\mathbf{X}^*$ represents the matrix of testing points in $\mathcal{D}$.

$$p(\mathbf{y}, \mathbf{f}^*) = \mathcal{N}\left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{X}^*) \end{bmatrix}, \begin{bmatrix} k(\mathbf{X}, \mathbf{X}) + \tau^2\mathbf{I} & k(\mathbf{X}, \mathbf{X}^*) \\ k(\mathbf{X}, \mathbf{X}^*)^\top & k(\mathbf{X}^*, \mathbf{X}^*) \end{bmatrix} \right) \tag{1}$$

The notations $m(\mathbf{X})$ and $k(\mathbf{X}, \mathbf{X})$ generalises the component-wise evaluation of $\mathbf{X}$ by m and k, e.g. $m(\mathbf{X}) = [m(\mathbf{x}_1)\ m(\mathbf{x}_2)\ ...\ m(\mathbf{x}_N)]$. Standard Gaussian process regression uses a zero prior for the mean, $m(\mathbf{x}) = 0$, and the Gaussian-form radial basis function, also called exponentiated quadratic, for the covariance function, written here:

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\sum_{k=1}^{d} \frac{\|x_k - x'_k\|_2^2}{2l_k^2} \right) \tag{2}$$

where σ and l_k are, respectively, the output standard deviation and a length scale. They are the model hyperparameters and are denoted by the vector $\boldsymbol{\theta}$. The covariance uses automatic relevance determination [28] to learn a length scale for each dimension of the design space. It was decided to take a universal standard formulation for the Gaussian process because the objective is to study the broader multifidelity data fusion framework. Future work will aim to improve the intermediate surrogate models. To make predictions of the function outputs $f(\mathbf{X}^*) = \mathbf{f}^*$, the posterior distribution $p(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y})$ can be derived by applying the multivariate Gaussian conditional rule which gives $p(\mathbf{f}^*|\mathbf{X}^*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(\mathbf{f}^*|\mu_{GP}(\mathbf{X}^*), \sigma_{GP}(\mathbf{X}^*))$ with posterior mean $\mu_{GP}(\mathbf{x})$ and posterior standard deviation $\sigma_{GP}(\mathbf{x})$ defined as

$$\mu_{GP}(\mathbf{X}^*) = k(\mathbf{X}, \mathbf{X}^*)^\top [k(\mathbf{X}, \mathbf{X}) + \tau^2\mathbf{I}]^{-1}\mathbf{y} \tag{3}$$

$$\sigma_{GP}(\mathbf{X}^*) = k(\mathbf{X}^*, \mathbf{X}^*) - k(\mathbf{X}, \mathbf{X}^*)^\top [k(\mathbf{X}, \mathbf{X}) + \tau^2\mathbf{I}]^{-1} k(\mathbf{X}, \mathbf{X}^*) \tag{4}$$

The standard deviation σ_{GP} quantifies the uncertainty arising from the approximation process and corresponds to the uncertainty in the model prediction away from a training point. The posterior distribution strongly depends on the formulation used for $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$. The method often used for learning the hyperparameters is based on the evaluation of the likelihood function $p(\mathbf{y}|\boldsymbol{\theta})$. Using the standard form for a multivariate Gaussian distribution, the log-likelihood function is given by

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^\top [k(\mathbf{X}, \mathbf{X}) + \tau^2\mathbf{I}]^{-1}\mathbf{y} - \frac{1}{2}\log |k(\mathbf{X}, \mathbf{X}) + \tau^2\mathbf{I}| - \frac{N}{2}\log (2\pi) \tag{5}$$

Gradient-based optimisation algorithms can be used to maximise this log-likelihood function. It can be seen in Eqs. (3–5) that the regression model involves the matrix inversion of $[k(\mathbf{X},\mathbf{X})+\tau^2\mathbf{I}]$. The number of operations for this matrix inversion scales with N^3 operations. Considering the size of the data sets for the test case introduced later in this paper, it leads to scalability limits of the Gaussian process regression and requires additional approximations.

B. Stochastic variational inference for Gaussian process models

A comprehensive review of scalable Gaussian processes is presented in [29]. Stochastic variational inference is used herein to enable the application to large data sets. To reduce the computational complexity, a set of random variables $\mathbf{u}$ at some locations $\mathbf{Z}$ is introduced in [26, 30]. The relation between $\mathbf{f}$ and $\mathbf{u}$ is established with a multivariate Gaussian distribution considered as the prior distribution of the stochastic variational Gaussian process (SVGP) model

$$p(\mathbf{f},\mathbf{u}) = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} k(\mathbf{X},\mathbf{X}) & k(\mathbf{X},\mathbf{Z}) \\ k(\mathbf{X},\mathbf{Z})^\top & k(\mathbf{Z},\mathbf{Z}) \end{bmatrix}\right) \tag{6}$$

where $k(\mathbf{Z},\mathbf{Z})$ is the covariance matrix associated to $\mathbf{u}$ and is built by evaluating the covariance function at the pairs $(\mathbf{z},\mathbf{z}') \in \mathbf{Z}\times\mathbf{Z}$. The previous joint distribution can be formulated as

$$p(\mathbf{f},\mathbf{u}) = p(\mathbf{f}|\mathbf{u})p(\mathbf{u}) = \mathcal{N}\big(\mathbf{f}|k(\mathbf{X},\mathbf{Z})k(\mathbf{Z},\mathbf{Z})^{-1}\mathbf{u}, k(\mathbf{X},\mathbf{X})-\mathbf{Q}\big)\,\mathcal{N}\big(\mathbf{u}|\mathbf{0}, k(\mathbf{Z},\mathbf{Z})\big) \tag{7}$$

where $\mathbf{Q} = k(\mathbf{X},\mathbf{Z})k(\mathbf{Z},\mathbf{Z})^{-1}k(\mathbf{X},\mathbf{Z})^\top$. Using the observable variable $\mathbf{y}$, it is now written as

$$p(\mathbf{y},\mathbf{f},\mathbf{u}) = p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{u})p(\mathbf{u}) \tag{8}$$

The marginal likelihood $p(\mathbf{y})$ measures how well the model explains the training data and is used to set the hyperparameters. However, it can be demonstrated that the expression of $p(\mathbf{y})$ for the SVGP is the same as Eq. (5) and implies the problematic matrix $k(\mathbf{X},\mathbf{X})$ inversion. Starting from Jensen's inequality applied on the conditional probability

$$\log p(\mathbf{y}|\mathbf{u}) = \log \mathbb{E}_{p(\mathbf{f}|\mathbf{u})}[p(\mathbf{y}|\mathbf{f})] \geq \mathbb{E}_{p(\mathbf{f}|\mathbf{u})}[\log p(\mathbf{y}|\mathbf{f})] \tag{9}$$

It is substituted into the expression $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{u})p(\mathbf{u})\mathrm{d}\mathbf{u}$ in [31] to obtain a tractable bound on the marginal likelihood

$$\log p(\mathbf{y}) \geq \log \mathcal{N}\big(\mathbf{y}|\mathbf{0}, k(\mathbf{X},\mathbf{Z})k(\mathbf{Z},\mathbf{Z})^{-1}k(\mathbf{X},\mathbf{Z})^\top + \tau^2\mathbf{I}\big) - \frac{1}{2\tau^2}\mathrm{tr}(k(\mathbf{X},\mathbf{X})-\mathbf{Q}) \tag{10}$$

This bound can then be used as an alternative maximisation objective function to optimise model hyperparameters and the inducing points $\mathbf{Z}$. In the case where all the training points are considered as inducing points, i.e. $\mathbf{Z}=\mathbf{X}$, the bound is equal to the marginal likelihood corresponding to Eq. (5), specifically $\log \mathcal{N}\big(\mathbf{y}|\mathbf{0}, k(\mathbf{X},\mathbf{X})+\tau^2\mathbf{I}\big)$. To obtain a more computationally scalable bound, a variational distribution $q(\mathbf{u}) = \mathcal{N}(\mathbf{y}|\mathbf{m},\mathbf{S})$ is introduced such that $q(\mathbf{f},\mathbf{u}) = p(\mathbf{f}|\mathbf{u})q(\mathbf{u})$. Indeed, in order to make predictions, it is necessary to compute the posterior distribution $p(\mathbf{f},\mathbf{u}|\mathbf{y})$ from Bayes' formula which requires the computation of $p(\mathbf{y})$. Thus, the idea is to approximate the posterior with $q(\mathbf{f},\mathbf{u})$. It is shown in [26] that applying a standard variational bound leads to the Evidence Lower BOund (ELBO)

$$\begin{aligned}\log p(\mathbf{y}) \geq \log \mathcal{N}\big(\mathbf{y}|k(\mathbf{X},\mathbf{Z})k(\mathbf{Z},\mathbf{Z})^{-1}\mathbf{m}, \tau^2\mathbf{I}\big) - \frac{1}{2\tau^2}\mathrm{tr}(k(\mathbf{X},\mathbf{X})-\mathbf{Q}) \\ - \frac{1}{2\tau^2}\mathrm{tr}(k(\mathbf{X},\mathbf{Z})k(\mathbf{Z},\mathbf{Z})^{-1}\mathbf{S}k(\mathbf{Z},\mathbf{Z})^{-1}k(\mathbf{X},\mathbf{Z})^\top) - \mathrm{KL}[q(\mathbf{u})||p(\mathbf{u})]\end{aligned} \tag{11}$$

where $\mathrm{KL}[.||.]$ is the Kullback–Leibler divergence. The ELBO has a unique optimum over the parameters $\mathbf{m}$ and $\mathbf{S}$ at which point it equals the bound given by Eq. (10). The advantages of this formulation come from the use of stochastic optimisation and mini-batch learning which reduces the computational complexity to $O(N_{\mathrm{ind}}^3)$. In this paper, the ELBO is optimised by the Adam Optimizer designed in [32] for stochastic objective functions. Predictions of $\mathbf{f}^*$ are made in test points $\mathbf{X}^*$ by substituting the approximate posterior into $p(\mathbf{f}^*|\mathbf{y})$:

$$\begin{aligned} p(\mathbf{f}^*|\mathbf{y}) &= \int p(\mathbf{f}^*|\mathbf{f},\mathbf{u})p(\mathbf{f},\mathbf{u}|\mathbf{y})\mathrm{d}\mathbf{f}\mathrm{d}\mathbf{u} \\ &\approx \int p(\mathbf{f}^*|\mathbf{f},\mathbf{u})p(\mathbf{f}|\mathbf{u})q(\mathbf{u})\mathrm{d}\mathbf{f}\mathrm{d}\mathbf{u} \\ &= \int p(\mathbf{f}^*|\mathbf{u})q(\mathbf{u})\mathrm{d}\mathbf{u} \end{aligned} \tag{12}$$

This expression is tractable because $q(\mathbf{u}) = \mathcal{N}(\mathbf{y}|\mathbf{m}, \mathbf{S})$ is known and $p(\mathbf{f}^*|\mathbf{u})$ can be derived by applying the multivariate Gaussian conditional rule to the prior $p(\mathbf{f}^*, \mathbf{u})$. It only depends on the inducing variables $\mathbf{u}$ and does not depend on the training data. All the information is incorporated in the variational distribution q and the hyperparameters are obtained during the training process.

C. Fidelity function

In order to quantify the physical uncertainty associated with the generation of the data by an information source, we use the fidelity function σ_f defined in [24]. It is a representation of the confidence in an information source over the design space and comes from expert knowledge, past experience and/or empirical data. The total variance is then

$$\sigma_t^2 = \sigma_{GP}^2 + \sigma_f^2 \tag{13}$$

which captures both the quality of the Gaussian process model and the expert's confidence in the underlying data.

D. Variance-weighted combination of intermediate surrogate models

As presented in [24, 25], a surrogate model is built for every information source, i.e. a Gaussian process regression model is trained from every $\mathbf{f}_i$ and $\mathbf{X}_i$ produced by the i-th among I information sources. The resulting intermediate posterior mean and posterior standard deviation are denoted $\mu_{GP,i}$ and $\sigma_{GP,i}$. As it is defined earlier, the total variance of the i-th information source is $\sigma_{t,i}^2 = \sigma_{GP,i}^2 + \sigma_{f,i}^2$ where $\sigma_{f,i}$ is the fidelity function defined for the i-th information source. The multifidelity fused estimation of $f(\mathbf{x})$ is computed with a weighted sum of $\mu_{GP,i}$ and $\sigma_{GP,i}$. The weighting established in [33] to give more importance to high-confidence models is defined by the multifidelity mean estimate and the multifidelity total variance as

$$\overline{\mu}(\mathbf{x}) = \overline{\sigma}^2(\mathbf{x}) \sum_{i=1}^{I} \frac{\mu_{GP,i}(\mathbf{x})}{\sigma_{t,i}^2(\mathbf{x})} \tag{14}$$

$$\overline{\sigma}^2(\mathbf{x}) = \left(\sum_{i=1}^{I} \frac{1}{\sigma_{t,i}^2(\mathbf{x})} \right)^{-1} \tag{15}$$

This methodology combines the non-hierarchical multifidelity formulations to fuse disparate data and the definition of fidelity function to quantify information source uncertainty.

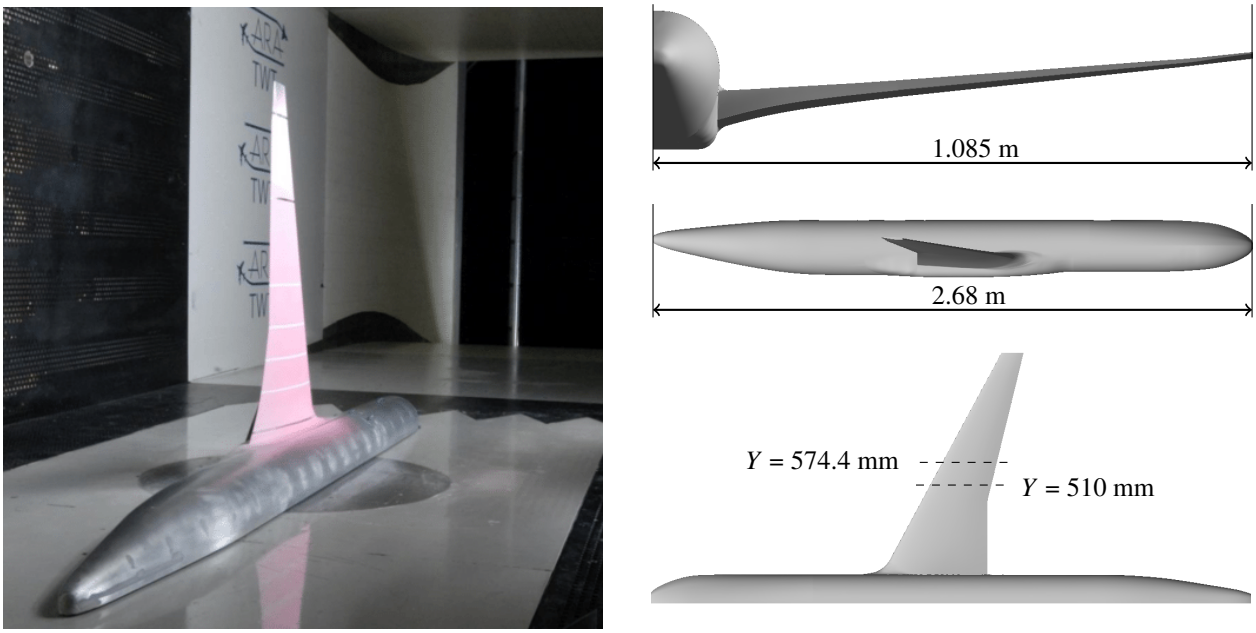
III. Civil Aircraft Wing Configuration

A. Half-wing-fuselage geometry and previous test campaign

Figure 1 presents the RBC12 half-wing-fuselage wind tunnel model which is representative of a typical 1970s/1980s commercial aircraft design and has been used in previous work to investigate the shock-buffet phenomenon [27, 34, 35]. The wind tunnel model has been scaled down by a factor of 17.5. Its mean aerodynamic chord is 0.279 m, the wing tip chord is 0.098 m, the semispan is 1.085 m, the aspect ratio is 7.78 and the reference area is 0.2959 m^2. The model wing is twisted, tapered and swept back with a constant quarter-chord sweep angle of 25° and a leading-edge sweep angle of 28.3°. For visualisation purposes, a dimensionless spanwise coordinate is defined as $\eta = (Y - 0.019)/1.085$. It is derived from the spanwise coordinate Y from which the plinth thickness (0.019 m) is subtracted.

The work presented here uses data from experiments conducted in the Aircraft Research Association Transonic Wind Tunnel in 2015 within the European Commission Clean Sky buffet control of transonic wings project (BUCOLIC) [36]. The wind tunnel is of a closed-circuit, continuous type. It has a test section of 2.74 m by 2.44 m, an operating Mach number of up to 1.4, and the stagnation pressure can be varied between 80 and 120 kPa. The RBC12 half-model installed on the floor of the test section is shown in Fig. 1a. Although the previous test entry focused on shock buffet, the data considered hereafter are assumed to be steady in pre-buffet-onset conditions. A more detailed description of the experimental study concerning the unsteady phenomenon can be found in [27].

Herein we focus on pressure features, through the pressure coefficient, over the upper surface of the RBC12 wing. An overview of all available data in this study is shown in Fig. 2 with details discussed in the following. The chosen flow conditions follow the available data from the previous test campaign including reference Mach numbers between $M = 0.7$ and 0.84, a Reynolds number (with respect to mean aerodynamic chord) of approximately 3.4×10^6 to 3.7×10^6 and angles of attack approximately between $\alpha = -1.2°$ and 8.4°.



(a) Model mounted on wind tunnel floor **(b) Front, side and top views of geometry**

Fig. 1 Perspective of installed physical model as well as front, side and top views of computational model.

B. Experimental data

A large number of static pressure taps (chosen as the first data source) were installed on the wing and the fuselage to observe the shock movement and flow development around shock-buffet onset with varying Mach number and angle of attack. Figure 3 (left) shows a total of 124 static pressure taps located on the upper surface of the wing as well as the measured pressure coefficient at an arbitrarily chosen experimental condition (Mach number $M = 0.74$ and angle of attack $\alpha = 3°$). In addition, the second data source is dynamic pressure-sensitive paint (DPSP) recorded with a high-frame-rate imaging system and previously processed to analyse the unsteady behaviour of the shock dynamics with plenty of detail presented in [27]. Such an advanced optical measurement technique can provide pressure information over a large region of the wing surface, which offers a significant advantage over discrete pressure measurements. For instance, the set-up of the previous test campaign resulted in a resolution of 1.3 pixel/mm. The idea is to exploit time-averaged DPSP measurements to benefit from its spatial abundance (and indeed to address the impact of this abundance in a machine learning context using Gaussian processes) and to complement static pressure tap data.

C. Reynolds-averaged Navier–Stokes simulations

The geometry shown in Fig. 1 was reconstructed from the wind tunnel model with a laser-equipped scanner. A hybrid mesh was created from it with the mesh generator SOLAR [37], following industry best-practice guidelines. The resulting mesh is composed of approximately 12.2×10^6 points. The unstructured finite volume flow solver DLR-TAU [38], developed by the German Aerospace Center (DLR), is used to perform Reynolds-averaged Navier–Stokes (RANS) simulations (i.e. the third data source). The turbulence model applied is the negative Spalart-Allmaras model [39]. The convective fluxes of the mean flow equations are computed through the central scheme with scalar artificial dissipation and those of the turbulence model use a Roe scheme. To the extent possible, numerical flow conditions have been matched to the available wind tunnel experimental conditions. The reference Mach number varies between 0.7 and 0.8, the Reynolds number (based on mean aerodynamic chord) is fixed to 3.75×10^6 (in accordance with previous work by the authors), the reference temperature is 266.5 K, and the reference pressure varies between 66 and 76 kPa. For the steady-state simulations computed here, the angle of attack varies between $\alpha = 0°$ and 5° depending on Mach number to stay below the onset of the shock-buffet instability.

Table 1 Overview of different data sets and attributed constant fidelity functions.

Information source	Flow conditions	Spatial coordinates	Total number of points	Fidelity function value
Pressure tappings	183	150	27,450	0.0
DPSP	30	289,215	8,676,450	0.025
RANS	14	52,491	734,874	0.05
XFLR5	42	900	37,800	0.8

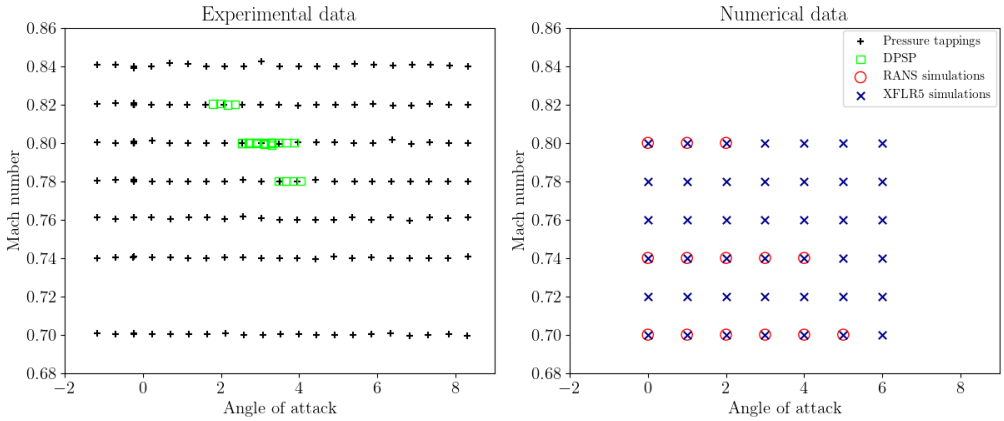


Fig. 2 Available data according to Mach number and angle of attack.

D. Three-dimensional panel method

Panel methods normally solve the linear, inviscid and irrotational (i.e. linear potential) flow equations on a suitable surface discretisation of the geometry. Such a method developed in [40] and implemented in XFLR5 has been used to generate the pressure distribution over the model wing. Obviously, the potential flow hypothesis lacks accuracy for the transonic Mach numbers and viscous flow addressed herein, but the flexibility and execution speed justify its inclusion, while its associated fidelity function will take its inaccuracy into account. Rather than only considering the wing planform for the discretisation, the model geometry used in XFLR5 accounts for the main geometrical information of the RBC12 half-model configuration (going beyond the wing span and local chord length to describe the planform), such as the aerofoil profile, the twist and dehidral angles. The fuselage has been replaced by a flat wing extension with the same profile as the wing-root profile for modelling purposes. XFLR5 computes the pressure coefficient at the centroid of each panel. The flow solution has been computed for angles of attack varying between $\alpha = 0°$ and $6°$ with an increment of $1°$ and Mach numbers varying between $M = 0.7$ and 0.8 with an increment of 0.2, cf. Fig. 2.

IV. Results for Surface Pressure Distribution

An initial phase of the work concerned the preparation of standardised pressure data from the four information sources presented in Section III. Each information source has been pre-processed to get arrays of unique pressure coefficient data points identified by their spatial location on the wing and the associated flow conditions (limited to Mach number and angle of attack herein). The experimental and numerical data are summarised with respect to the flow conditions in Fig. 2. All numerical data were produced as part of this work, in order to go through the design space considering the computational cost and the shock-buffet onset limit. From a data analysis perspective, Table 1 summarises the number of flow conditions studied and the number of spatial coordinates at which a pressure coefficient value is known. The latter gives an idea of the spatial abundance of each information source. In practice, the data availability can differ significantly, particularly for wind tunnel testing, and it can fairly be said that the amount of available data is correlated with the cost of data generation. The test case studied in this paper offers a great example of such variety, specifically accuracy and resolution on the wing. In terms of dimension, the input space is limited to four dimensions herein, defined by Mach number and angle of attack, and the spatial coordinates (X, Y) in stream- and spanwise directions, respectively. The third spatial coordinate has been discarded (in agreement with the information

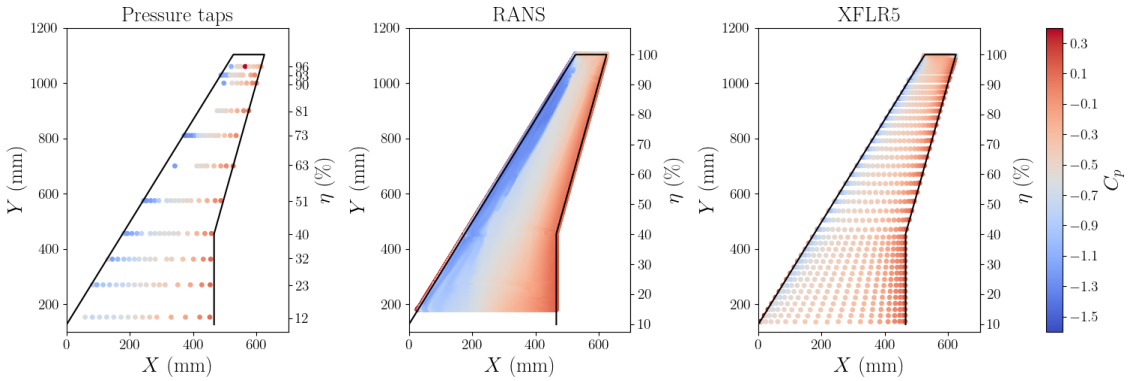


Fig. 3 Pressure coefficient on the wing surface.

available from the experimental data sets) while projecting each numerical data set equally onto the wing planform. The table also gives the constant values of the fidelity function used herein. They are case dependent and have been set somewhat arbitrarily to reflect the confidence and desired influence of each of the information sources. Pressure tappings are considered as the 'ground truth' and therefore have a null value of fidelity function over the whole design space. DPSP and RANS have low fidelity-function values as they are considered the best aerodynamic information sources in this study whereas XFLR5 is considered a low quality information source for the reasons given in Section III, and therefore has a significantly higher value in the fidelity function. The relative difference between DPSP and RANS is explained by the fact that DPSP is calibrated on pressure tapping data. Going forward, we will explore more sophisticated fidelity functions which take into account variation in the design space and more refined expert opinion. It can be argued about the data fusion framework, albeit being presented as non-hierarchical, that the values of the fidelity function create a hierarchy between the information sources. However, this hierarchy is absent in the construction of the multifidelity model because the total standard deviation balances the fidelity function with the posterior standard deviation of each intermediate surrogate model. It leads to values of the total standard deviation that vary over the design space and whose hierarchy between information sources changes over the same design space.

Application of the multifidelity approach to high-fidelity (and hence high-dimensional) data sources is challenging because it increases significantly the number of data points considered in addition to facing the 'curse of dimensionality'. Initially, the framework is applied to the design space of spatial coordinates at fixed flow condition, and then the investigation is extended to include also the variable flow conditions. Hence, the first objective (addressed in Section IV.A) is to demonstrate the feasibility of the approach in a way that allows its explanation and a visual representation of the SVGP regression. The second objective (addressed in Section IV.B) seeks to understand the capabilities and limitations on a higher-dimensional design space. The numerical library adopted in the implementation of the data fusion framework is GPflow, a Python module based on TensorFlow [41, 42]. It has Graphics Processing Unit (GPU) enabled capability. All GPU computations make use of a dedicated compute node equipped with four Nvidia Tesla V100 GPUs.

A. Two-dimensional input space of coordinates (X, Y) at representative fixed flow conditions

At fixed flow conditions, the model input space $\mathcal{D}$ has two dimensions (spatial coordinates in stream- and spanwise directions). Pressure coefficient distributions on the wing from experimental pressure tappings as well as numerical RANS and XFLR5 simulations at a representative Mach number $M = 0.74$ and angle of attack $\alpha = 3°$ are considered. This first demonstration is limited to three information sources because DPSP data are not available at this flow condition. The standardised (raw) data sets used to train the model are shown on Fig. 3. Standard Gaussian process regression can be applied to train two (out of the three) intermediate surrogate models. The number of spatial coordinates given in Table 1 rules out the use of a standard Gaussian process in practice when leveraging the information obtained from RANS simulations, even at a single fixed flow condition. Thus, SVGP regression is applied instead and its ability to form a suitable surrogate model from a big data set for such aerodynamic application is demonstrated. The two important points to consider when setting up the SVGP method are the choice of the number of inducing points and their initial location. It is worth remembering that the SVGP regression method seeks to optimise the hyperparameters, the parameters of the variational distribution and the inducing points. The impact of both the number of inducing points and their initial distribution in space is presented in Table 2 (and later in Table 3 in Section IV.B). It was found that the

Table 2 Accuracy of SVGP regression on a two-dimensional input space for RANS data.

Random initialisation				Uniform initialisation			
N_{ind}	Training time (s)	MSE (training)	MSE (test)	N_{ind}	Training time (s)	MSE (training)	MSE (test)
50	115	0.237	0.241	50	113	0.218	0.218
100	122	0.172	0.174	101	119	0.136	0.135
500	145	0.075	0.076	523	158	0.103	0.103
1000	239	0.051	0.052	1121	288	0.077	0.078
2000	713	0.035	0.037	2057	783	0.051	0.052
3000	1673	0.028	0.030	3055	1785	0.040	0.041
5000	5873	0.022	0.023	5059	6055	0.034	0.035

initialisation has an influence on the results. Indeed, it was observed that the optimisation acts predominantly on the hyperparameters and the parameters of the variational distribution. This tends to reinforce with an increasing dimension of the input space. Thus, the inducing points do not move significantly and have to be initialised wisely. In this example, two different initialisation methods are presented. The first one randomly selects N_{ind} inducing points in the training data set while the second uniformly distributes points over the wing in both spatial dimensions. We will assess the SVGP approach with an evaluation of the relative accuracy of the model with respect to the training time required. Then, we will use SVGP with the RANS training set to enable the application of the multifidelity data fusion framework. The mean squared error to assess the relative accuracy of the surrogate model is used

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{16}$$

where y_i and $\hat{y}_i$ are the i-th original known and predicted values, respectively, and n is the total number of predictions. The RANS data set at $M = 0.74$ and $\alpha = 3°$ has been randomly divided for training (80% of the data) and testing purposes (the remaining 20%). In Table 2, the MSE values are similar for the training and testing data set. This suggests that the model has been set up correctly and that the model also predicts well for points not used for training. In other words, it indicates that overfitting has been avoided. The quality of the model increases with the number of inducing points. However, the computation time required also increases significantly due to the cubic complexity with respect to N_{ind}. The choice of the number of inducing points is a trade-off between calculation time and accuracy. Non-linear transonic features of the flow such as shock waves are difficult to model and this explains the need to take into account a sensible number of data points. Despite a slightly better accuracy of the intermediate surrogate model with the random initialisation, the uniform initialisation was found to give better results when utilised for multifidelity data fusion and will be chosen in the following. To apply the framework on the three information sources mentioned, it was decided (after consulting Table 2) to initialise 2,480 inducing points uniformly over the wing for the RANS pressures.

Figures 4 and 5 present the resulting intermediate surrogate models showing mean and standard deviation. It can be observed immediately for the intermediate surrogate model based on pressure-tap data that the posterior mean presents an inaccurate pressure prediction and that the associated posterior standard deviation is high away from the training points. The abundance of data in the design space for the two other information sources, particularly from RANS simulations, explains a more realistic pressure distribution predicted by the surrogate models. For instance, the RANS prediction based on 2,480 inducing points is almost indistinguishable from the raw data with 52,491 surface mesh points. It also explains the low standard deviation values for these two information sources. However, this does not give any information on the physical accuracy of the intermediate models. Low values of standard deviation indicate that a predicted location is close to a training point but do not give any particular information on its relevance. For the RANS data set, the corresponding intermediate surrogate model exhibits an interesting (spotty) pattern in the standard deviation in which we can see the influence of the inducing points. In our case, we consider the data from pressure tappings as the most accurate source of information whereas the data obtained from the linear potential panel method lacks accuracy at the flow conditions considered. The idea is to benefit from the spatial abundance while balancing with the assumed accuracy of and confidence in each information source. The fidelity function is exactly defined for this purpose. It aims to benefit from the highest confidence in the discrete pressure taps where they are available while

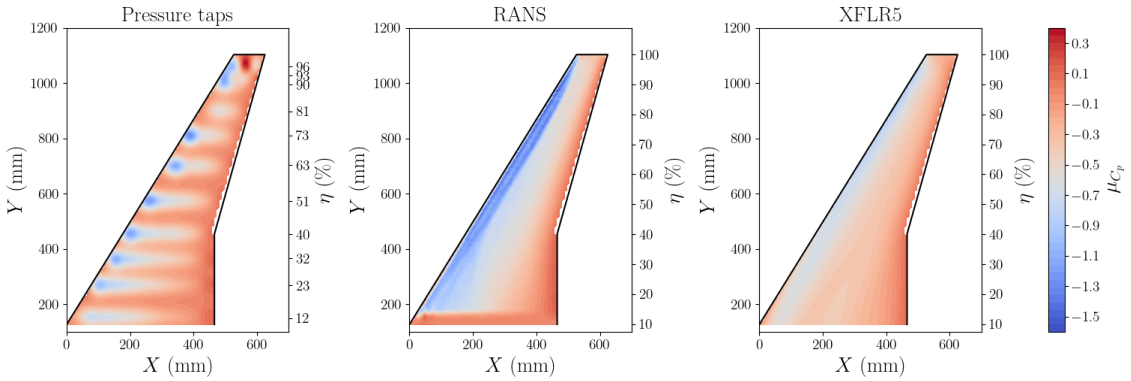


Fig. 4 Mean of each intermediate surrogate model.

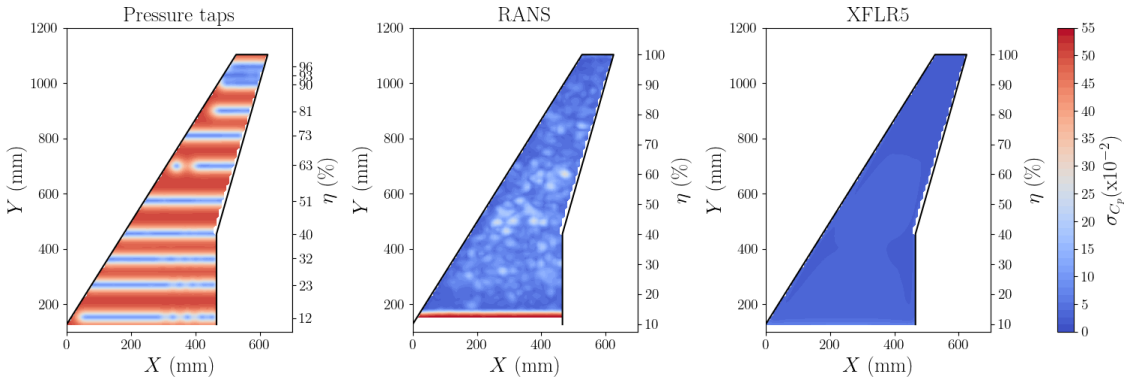
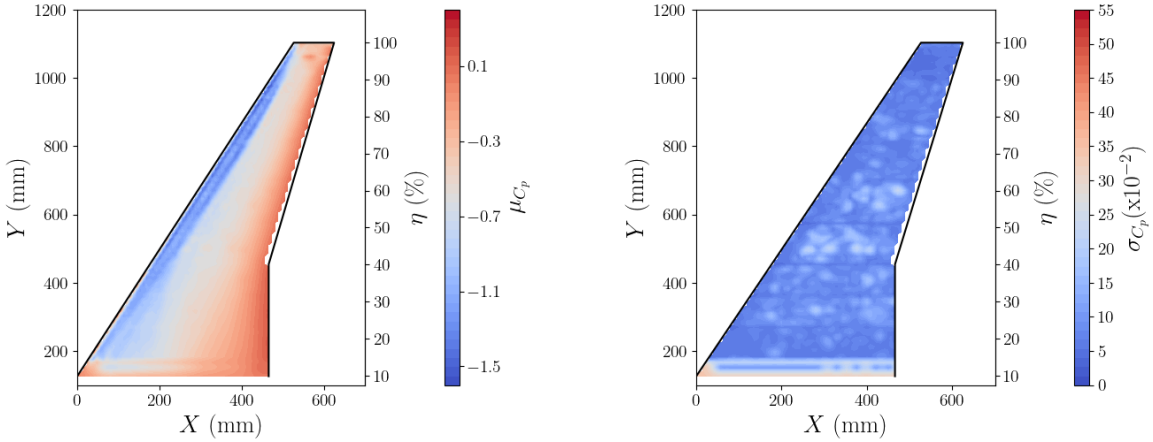


Fig. 5 Standard deviation of each intermediate surrogate model.

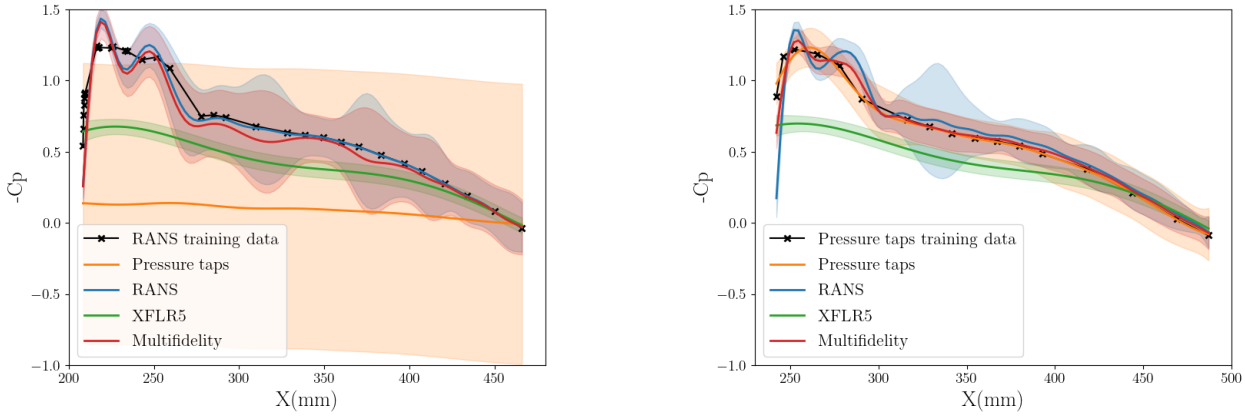
giving more importance to the RANS results compared with the potential-flow data where the standard deviation of the pressure tap surrogate model is high. Of course, in the absence of relatively expensive RANS simulations (e.g. at benign flow conditions with low Mach number and angle of attack), the potential-flow data would fill the gap.

A multifidelity surrogate model is then obtained by applying Eqs. (14) and (15), and its output is plotted in Fig. 6. The figure demonstrates the feasibility of the multifidelity approach for such realistic application. It shows that the resulting model respects the 'ground truth' pressure-tap data and learns the spatial distribution of the pressure field from the spatially more resolved data sets to obtain a physically reliable surrogate. The standard deviation plotted in Fig. 6b may seem higher than that of the intermediate models but accounts for the total standard deviation, including the fidelity function. As required, the standard deviation is relatively low in the whole design space benefiting from all data available, except for the wing root where there is little information and, in particular, no CFD data. The spanwise region for extraction of CFD data was a deliberate choice to demonstrate the characteristics of the resulting intermediate surrogate model. Locations with little or no training data have a resulting mean that visually appears false and higher standard deviation values. An isolated pressure tap located near the wing tip seems to be out of the standard range and could be considered as an outlier, as seen in Fig. 3. This highest value of pressure coefficient (deep red spot) is found in its intermediate surrogate model in Fig. 4 and is propagated to the multifidelity model in Fig. 6a. The high confidence attributed to the pressure taps entails an extensive discussion of automatically and reliably dealing with possible outliers in the construction of the intermediate and multifidelity surrogate models. The benefits of the data fusion approach are also showcased in Fig. 7 for pressure distributions at two selected spanwise stations, The figure compares the predictions (and raw data where available) showing the posterior mean and the Gaussian process confidence intervals ($\pm 1.96\sigma_{GP}$) superimposed. Pressure-tap data is available for spanwise station $Y = 574.4$ mm. In contrast, at the other station $Y = 510.0$ mm, pressure-tap data is not available and instead the Gaussian process will heavily rely on RANS data included as reference. Away from training data, the standard deviation of the intermediate surrogate model based on pressure-tap data is high. Yet, in both locations, owing to its construction, the multifidely model behaves similarly to the model with the lowest total standard deviation. The prediction follows the RANS intermediate surrogate when



(a) Mean **(b) Standard deviation**

Fig. 6 Mean and standard deviation of multifidelity surrogate model.



(a) η = 45%, Y = 510.0 mm **(b) η = 51%, Y = 574.4 mm**

Fig. 7 Comparison of data and predictions at two chosen spanwise stations. Solid lines describe the posterior mean and the colour-matched transparent areas give the corresponding confidence intervals. Note the difference in available training data and its impact on the prediction of the multifidelity model.

the posterior standard deviation of the pressure-tap model is high due to a lack of its training data. It is clear that the potential-flow intermediate surrogate model is inaccurate and the fidelity function accounts for that by a relatively high constant. Despite the large number of inducing points, the confidence interval of the RANS intermediate surrogate (transparent blue area in Fig 7) is higher than other intermediate surrogate models. It is due to the specific use of SVGP regression for this information source.

The preceding demonstration at fixed flow conditions shows both the capability of the multifidelity data fusion framework for a realistic case and the feasibility of using SVGP regression to incorporate information from spatially resolved sources leading to very large data sets. However, the motivation of using SVGP regression is also to explore higher-dimensional design spaces. Indeed, if we consider variable flow conditions, Table 1 suggests that the cubic complexity of standard Gaussian processes with respect to the data size makes it prohibitive even for the spatially coarse data sources (i.e. discrete pressure tappings and potential-flow modelling). This is investigated next.

B. Four-dimensional input space (including variable flow conditions)

For the four-dimensional design space, the data set sizes range from 27,450 to 8,676,450 input points. The latter results from highly-resolved, time-averaged DPSP measurements. Indeed, this optical measurement technique gives a

Table 3 Accuracy of SVGP regression on a four-dimensional input space for DPSP data.

Random initialisation				Uniform initialisation			
N_{ind}	Training time (s)	MSE (training)	MSE (test)	N_{ind}	Training time (s)	MSE (training)	MSE (test)
500	157	0.034	0.036	504	154	0.186	0.209
1000	205	0.031	0.031	984	205	0.051	0.055
2000	511	0.028	0.027	1992	517	0.035	0.037
5000	3747	0.026	0.023	4992	3646	0.031	0.030
10000	24811	0.025	0.022	10464	30023	0.028	0.027

pressure value at each pixel of the captured image. It is an advantage over traditional pressure-tap measurements but can be problematic computationally. In general, possible solutions to overcome the computational cost arising from the size of the data set could be to reduce the resolution of the image or to sub-sample the data set. It is considered here that it may lead to the loss of valuable information; cf. [27] where the quantity of the data set was instrumental in studying the details of the flow physics featuring unsteady shock motions. For reference, in [26], SVGP regression has been applied on a data set with over a million points for the study of apartment prices in the United Kingdom and enabled the surrogate model to be approximated with a compromise on accuracy and computation time. First, we will assess the SVGP approach for our aerodynamic test case in higher dimension. Second, we will exploit the capabilities of the SVGP approach when applying it to the multifidelity data fusion framework in the four-dimensional parameter space.

Concerning the SVGP assessment, the data set composed of the DPSP measurements has been randomly divided for training (80% of the data) and testing purposes (the remaining 20%). This division has been carried out only with respect to the experimental conditions of Mach number and angle of attack to conform to the experimental reality, i.e. for each flow condition it is assumed that the full spatial resolution is available. The two initialisation methods presented in Section IV.A are used again. The first one randomly selects N_{ind} inducing points in the training data set while the second uniformly distributes over the wing at each experimental condition considered for training. For the uniform point distribution, the total number of inducing points N_{ind} is divided by the number of experimental conditions. The resulting number of inducing points per experimental condition corresponds to the number of points distributed uniformly over the wing. Table 3 summarises the MSE computed on the training and testing subsets for various values of N_{ind} and for the two initialisation methods. Evidently, the conclusions as stated in the discussion surrounding Table 2 remain the same and are not repeated.

For the SVGP exploitation, all four information sources have been used. As stated previously, pressure tappings are considered as the most accurate (albeit spatially sparse) source of information of the pressure distribution on the wing. Its corresponding data set has been divided in equal halves with respect to the experimental conditions for training and testing purposes. SVGP regression is applied to build the four intermediate surrogate models. For pressure tappings, only the training data set is used to train the intermediate surrogate model and the inducing points are initialised randomly. For the RANS, DPSP and XFLR5 information sources, a number of around 5,000 inducing points is initially shared between all available flow conditions and then distributed uniformly over the wing. For example, considering the number of flow conditions for RANS and DPSP data sets, these information sources result in 350 and 164 inducing points per flow condition, respectively. The number of inducing points was chosen as a trade-off between accuracy and computing time, based on the insight gained earlier. The total time required to train the four intermediate surrogate models was just under four hours in total. To compute the resulting multifidelity model, the four fidelity functions (previously stated in Table 1) are defined as different constants in the design space with values characterising the relative confidence that the user has in the information source. The MSE of each intermediate model as well as the multifidelity model is computed on the defined training and testing subsets. Specifically, the MSE compares the predictions of each surrogate model at the training and testing subsets of the pressure tappings with known pressure-tap values at these locations. Table 4 summarises the computed results and, importantly, shows a relatively low MSE for the multifidelity model in comparison with the intermediate surrogate models. Indeed, it is comparable with that of the intermediate surrogate model for the pressure tappings itself. To explain, the pressure-tap data are used to compute the MSE, i.e. the original known value y_i in Eq. (16), explaining the low MSE of the pressure-tap surrogate. Hence, calculating the MSE of a model with respect to the data that trained the model results in a low MSE (training) value. Also, the pressure-tap

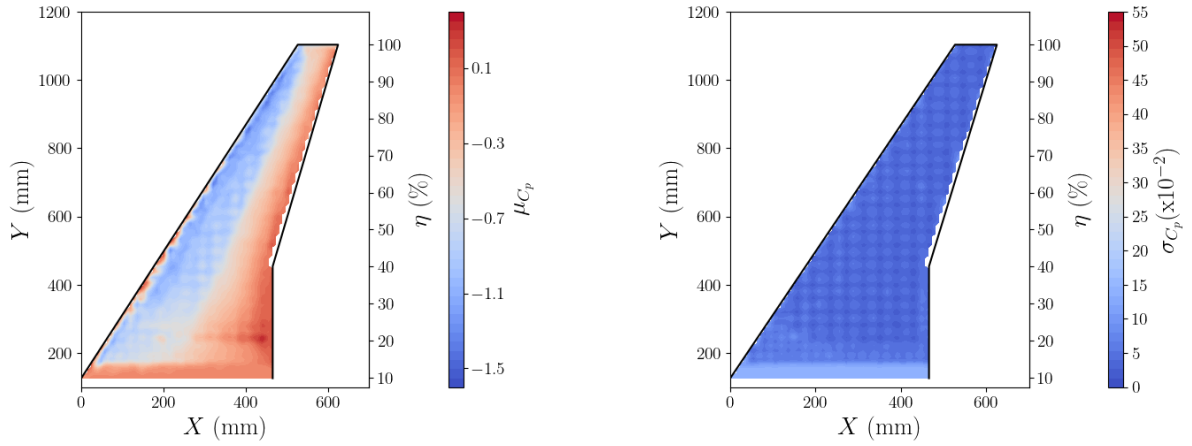
Table 4 Relative accuracy of intermediate and multifidelity surrogate models.

Information source	N_{ind}	MSE (training)	MSE (testing)	final ELBO
Pressure tappings	5000	0.003	0.006	1×10^4
DPSP	4900	0.191	0.188	-1×10^6
RANS	4920	0.067	0.063	-4×10^6
XFLR5	4704	0.093	0.100	-1×10^4
Multifidelity	-	0.008	0.010	-

training data are widely spread over the design space in terms of flow conditions (see Table 1), resulting in a globally well-suited model and the value of the MSE (testing) remains low. If an information source does not have data at the pressure-tap locations or if the data provided is significantly different, the resulting intermediate surrogate model will have high MSE. For example, the values from DPSP are similar to those from pressure tappings because the latter are used to calibrate the DPSP. However, the available DPSP data are concentrated at similar flow conditions (near Mach number $M = 0.8$ and angle of attack $\alpha = 3.0°$) and therefore do not allow sufficiently accurate predictions on the entire design space. On the other hand, as seen earlier, XFLR5 predictions are not physically accurate for the transonic Mach numbers involved because of the underlying assumptions of the aerodynamic model, but the distribution of data over the design space allows a global prediction closer to the training and test data sets. The intermediate surrogate based on RANS data has balanced characteristics between these two aspects (accuracy with respect to the value of pressure taps and exploration of the input space), which explains a lower MSE value. The multifidelity model thus benefits from the accuracy of the different models and the presence of data from different sources on the design space considered.

Rather than assessing the multifidelity model over the entire design space, specific predictions at desired user-defined flow conditions are sought. The multifidelity posterior mean and standard deviation of the surface pressure coefficient for Mach number $M = 0.72$ and angle of attack $\alpha = 3.0°$ are plotted in Fig. 8. This set of flow parameters is deliberately chosen as it is away from most training flow conditions. Nevertheless, the resulting multifidelity posterior mean has a contour with physically coherent shape. The uniform grid used for initialisation is visible as patterns in the posterior mean but especially in the posterior standard deviation. For instance, the multifidelity mean in Fig. 8a presents irregular patterns at the leading edge. The lack of inducing points at some locations of the leading edge prevents the model from taking into account the strong pressure gradient at those same locations. As observed with fixed flow conditions in Section IV.A, the position of the inducing points has an influence on the model, which is more accurate around these points. This explains why the uniform initialisation is chosen. Indeed, with random initialisation, the uneven distribution of data on the wing for some information source intensifies this phenomenon and makes predictions unevenly distributed over the design space. This phenomenon diminishes with the number of inducing points. It is therefore important to consider a sufficient number of inducing points, especially in a high-dimensional problem. A possible improvement would be to distribute the training time and thus the number of inducing points according to the number of data. Each source information would have a different number of inducing points depending on the size of the training data set.

As explained in Section II, the ELBO function is the lower bound of the marginal likelihood and describes how well the model explains the training data. The idea is then to maximise the ELBO to derive the set of model hyperparameters that explains best the training data. Stochastic gradient descent is used to find optimal values of those parameters. The ELBO evolution over 20,000 iterations of the Adam optimiser [32] during the training of the four intermediate surrogate models with SVGP regression is plotted in Fig. 9. An estimate of the final value is computed as the average over the last 2,000 iterations and included in Table 4. According to the table, the two largest data sets have significantly lower values of final ELBO. Indeed, each intermediate surrogate model used the same number of inducing points but some intermediate surrogate models have a significant higher number of training data. It is thus harder for them to capture the variation of all the training data. For instance, the intermediate surrogate model associated with the RANS data has a lower value than that of the DPSP data, which may be attributed to the wider spread of the RANS data over the design space. The calculation time of the method depends on the number of iterations specified. Figure 9 shows that 20,000 iterations is a reasonable choice for obtaining a converged ELBO value from these data, while convergence is reached in a tenth of that for the pressure taps. The strong oscillations of the ELBO can be explained by a large learning rate in the Adam optimiser. It is planned in future work to study the influence of the formulation and the settings of the stochastic optimiser used for SVGP regression. The converged value of ELBO, and thus the model accuracy with



(a) Mean **(b) Standard deviation**

Fig. 8 Multifidelity surrogate model predicted at $M = 0.72$ and $\alpha = 3.0°$

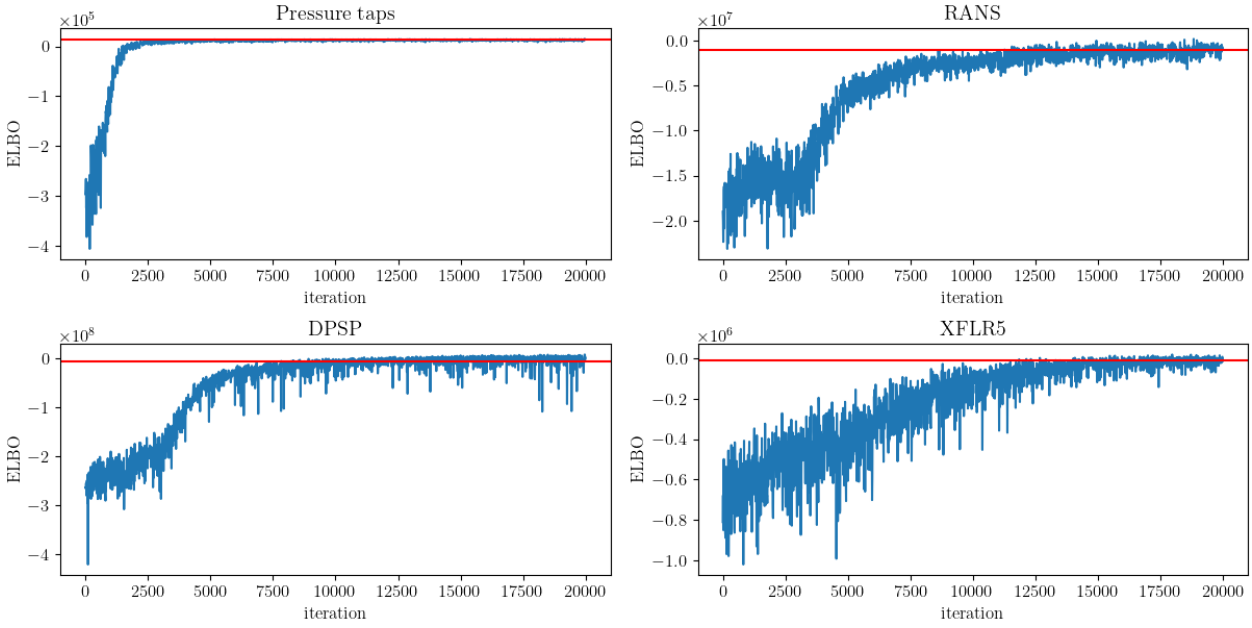


Fig. 9 ELBO evolution over 20,000 iterations of the Adam optimiser. The blue lines give the ELBO values every ten iterations while the red horizontal lines give the final value.

respect to the training data, depends on different elements of the Gaussian process model such as the choice of the number and the initialisation of the inducing points, the choice of the covariance function and the initialisation of its (hyper-) parameters and the likelihood variance initialisation. While some of these aspects have been studied in this test case, further scrutiny of the impact of the different parameters could lead to an optimised ELBO value and therefore to a better description of the physics by the model.

V. Conclusions

This paper presents a multifidelity data fusion framework based on a variance-weighted combination of Gaussian process models. Rich exploitation of data describing large aircraft wing pressure distributions obtained from experimental and computational methods has been carried out. Computation and memory constraints due to the volume of the data sets have been addressed by applying stochastic variational inference for Gaussian processes. The multifidelity data

fusion framework has been applied on two- and four-dimensional input spaces using up to four information sources of different fidelity. The viability of the multifidelity data fusion approach for large data sets arising from high-fidelity experimental and computational aerodynamics is illustrated. Over a defined design space, the model provides results with a quantified uncertainty from uncertainty-aware disparate data sources. With standard Gaussian process regression, the model uncertainty is decreased where more data are available (while observing the practical limit of approximately 10,000 data points) while with SVGP regression this depends on the number and location of so-called inducing points. It has been demonstrated that there is a trade-off between accuracy and training time.

The use of the fidelity function in this case proves its importance but does not show its full potential to balance the influence of each intermediate surrogate model on the input space. In order to improve the data fusion framework, future work will look into a detailed definition of the fidelity function such as taking into account variations in the reliability of an information source under different flow conditions. Enabling a multifidelity surrogate model with quantified uncertainty to be obtained, it is possible to apply the framework to address the practical question of sensor placement. There is a question in choosing the position of pressure taps on a wing in order to calibrate the more advanced optical pressure-sensitive-paint technique. Bayesian optimisation techniques based on the presented multifidelity surrogate can be applied to determine the most relevant locations for pressure-tap sensor placement by evaluating different information sources with a defined computational cost and with a desired accuracy.

Acknowledgements

The first author is grateful for the financial support by the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training in Distributed Algorithms (grant number 2447391 as part of EP/S023445/1) in partnership with Aircraft Research Association Ltd. We thank the German Aerospace Center for access to the TAU flow solver and the University of Liverpool for computing time on the high-performance computing system. We also acknowledge the technical support provided by the UK Research and Innovation (UKRI) Science and Technology Facilities Council (STFC) Hartree centre. The simulation data that support the findings of this study are available from the authors upon reasonable request.

References

[1] Kraft, E., "After 40 Years Why Hasn't the Computer Replaced the Wind Tunnel?" *ITEA J*, Vol. 31, 2010.

[2] Kraft, E. M., "Transforming Ground and Flight Testing through Digital Engineering," *AIAA Scitech 2020 Forum*, American Institute of Aeronautics and Astronautics, Orlando, FL, 2020. https://doi.org/10.2514/6.2020-1840.

[3] Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., and Newman, P. A., "Approximation and Model Management in Aerodynamic Optimization with Variable-Fidelity Models," *Journal of Aircraft*, Vol. 38, No. 6, 2001, pp. 1093–1101. https://doi.org/10.2514/2.2877.

[4] Hutchison, M., Mason, W., Grossman, B., and Haftka, R., "Aerodynamic optimization of an HSCT configuration using variable-complexity modeling," *31st Aerospace Sciences Meeting*, American Institute of Aeronautics and Astronautics, Reno, NV, 1993. https://doi.org/10.2514/6.1993-101.

[5] Tang, C., Gee, K., and Lawrence, S., "Generation of Aerodynamic Data using a Design Of Experiment and Data Fusion Approach," *43rd AIAA Aerospace Sciences Meeting and Exhibit*, American Institute of Aeronautics and Astronautics, Reno, NV, 2005. https://doi.org/10.2514/6.2005-1137.

[6] Krige, D., "A statistical approach to some basic mine valuation problems on the Witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, Vol. 52, No. 6, 1951, pp. 119–139. https://doi.org/10.10520/AJA0038223X_4792.

[7] Rasmussen, C., and Williams, C., *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2006.

[8] Timme, S., and Badcock, K. J., "Transonic Aeroelastic Instability Searches Using Sampling and Aerodynamic Model Hierarchy," *AIAA Journal*, Vol. 49, No. 6, 2011, pp. 1191–1201. https://doi.org/10.2514/1.J050509.

[9] Han, Z.-H., and Görtz, S., "Hierarchical Kriging Model for Variable-Fidelity Surrogate Modeling," *AIAA Journal*, Vol. 50, No. 9, 2012, pp. 1885–1896. https://doi.org/10.2514/1.J051354.

[10] Myers, D. E., "Matrix formulation of co-kriging," *Journal of the International Association for Mathematical Geology*, Vol. 14, No. 3, 1982, pp. 249–257. https://doi.org/10.1007/BF01032887.

[11] Kuya, Y., Takeda, K., Zhang, X., and Forrester, A. I. J., "Multifidelity Surrogate Modeling of Experimental and Computational Aerodynamic Data Sets," *AIAA Journal*, Vol. 49, No. 2, 2011, pp. 289–298. https://doi.org/10.2514/1.J050384.

[12] Forrester, A. I., Sóbester, A., and Keane, A. J., "Multi-fidelity optimization via surrogate modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 463, No. 2088, 2007, pp. 3251–3269. https://doi.org/10.1098/rspa.2007.1900.

[13] Keane, A. J., "Wing Optimization Using Design of Experiment, Response Surface, and Data Fusion Methods," *Journal of Aircraft*, Vol. 40, No. 4, 2003, pp. 741–750. https://doi.org/10.2514/2.3153.

[14] Zheng, J., Shao, X., Gao, L., Jiang, P., and Li, Z., "A hybrid variable-fidelity global approximation modelling method combining tuned radial basis function base and kriging correction," *Journal of Engineering Design*, Vol. 24, No. 8, 2013, pp. 604–622. https://doi.org/10.1080/09544828.2013.788135.

[15] Qian, P. Z. G., and Wu, C. F. J., "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments," *Technometrics*, Vol. 50, No. 2, 2008, pp. 192–204. https://doi.org/10.1198/004017008000000082.

[16] Ghoreishi, S. F., and Allaire, D. L., "A Fusion-Based Multi-Information Source Optimization Approach using Knowledge Gradient Policies," *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, American Institute of Aeronautics and Astronautics, Kissimmee, FL, 2018. https://doi.org/10.2514/6.2018-1159.

[17] Renganathan, S. A., Harada, K., and Mavris, D. N., "Aerodynamic Data Fusion Toward the Digital Twin Paradigm," *AIAA Journal*, Vol. 58, No. 9, 2020, pp. 3902–3918. https://doi.org/10.2514/1.J059203.

[18] Wen, X., Li, Z., Peng, D., Zhou, W., and Liu, Y., "Missing data recovery using data fusion of incomplete complementary data sets: A particle image velocimetry application," *Physics of Fluids*, Vol. 31, No. 2, 2019, p. 025105. https://doi.org/10.1063/1.5079896.

[19] Gunes, H., Sirisup, S., and Karniadakis, G. E., "Gappy data: To Krig or not to Krig?" *Journal of Computational Physics*, Vol. 212, No. 1, 2006, pp. 358–382. https://doi.org/10.1016/j.jcp.2005.06.023.

[20] Willcox, K., "Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition," *Computers & Fluids*, Vol. 35, No. 2, 2006, pp. 208–226. https://doi.org/10.1016/j.compfluid.2004.11.006.

[21] Franz, T., and Held, M., "Data Fusion of CFD Solutions and Experimental Aerodynamic Data," *ODAS 2017*, 2017. URL https://elib.dlr.de/114707/.

[22] Mifsud, M., Vendl, A., Hansen, L.-U., and Görtz, S., "Fusing wind-tunnel measurements and CFD data using constrained gappy proper orthogonal decomposition," *Aerospace Science and Technology*, Vol. 86, 2019, pp. 312–326. https://doi.org/10.1016/j.ast.2018.12.036.

[23] Bertram, A., Bekemeyer, P., and Held, M., "Fusing Distributed Aerodynamic Data Using Bayesian Gappy Proper Orthogonal Decomposition," *AIAA AVIATION 2021 FORUM*, American Institute of Aeronautics and Astronautics, VIRTUAL EVENT, 2021. https://doi.org/10.2514/6.2021-2602.

[24] Lam, R., Allaire, D. L., and Willcox, K. E., "Multifidelity Optimization using Statistical Surrogate Modeling for Non-Hierarchical Information Sources," *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, American Institute of Aeronautics and Astronautics, Kissimmee, FL, 2015. https://doi.org/10.2514/6.2015-0143.

[25] Feldstein, A., Lazzara, D., Princen, N., and Willcox, K., "Multifidelity Data Fusion: Application to Blended-Wing-Body Multidisciplinary Analysis Under Uncertainty," *AIAA Journal*, Vol. 58, No. 2, 2020, pp. 889–906. https://doi.org/10.2514/1.J058388.

[26] Hensman, J., Fusi, N., and Lawrence, N. D., "Gaussian Processes for Big Data," *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, AUAI Press, Arlington, VA, 2013, p. 282–290.

[27] Masini, L., Timme, S., and Peace, A. J., "Analysis of a civil aircraft wing transonic shock buffet experiment," *Journal of Fluid Mechanics*, Vol. 884, 2020, p. A1. https://doi.org/10.1017/jfm.2019.906.

[28] Liu, K., Li, Y., Hu, X., Lucu, M., and Widanage, W. D., "Gaussian Process Regression With Automatic Relevance Determination Kernel for Calendar Aging Prediction of Lithium-Ion Batteries," *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 6, 2020, pp. 3767–3777. https://doi.org/10.1109/TII.2019.2941747.

[29] Liu, H., Ong, Y.-S., Shen, X., and Cai, J., "When Gaussian Process Meets Big Data: A Review of Scalable GPs," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 31, No. 11, 2020, pp. 4405–4423. https://doi.org/10.1109/TNNLS.2019.2957109.

[30] Hensman, J., Matthews, A., and Ghahramani, Z., "Scalable variational Gaussian process classification," *Artificial Intelligence and Statistics*, PMLR, 2015, pp. 351–360.

[31] Titsias, M., "Variational Learning of Inducing Variables in Sparse Gaussian Processes," *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 5, edited by D. van Dyk and M. Welling, PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, FL, 2009, pp. 567–574.

[32] Kingma, D. P., and Ba, J., "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, May 7-9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun, 2015. URL http://arxiv.org/abs/1412.6980.

[33] Winkler, R. L., "Combining Probability Distributions from Dependent Information Sources," *Management Science*, Vol. 27, No. 4, 1981, pp. 479–488. https://doi.org/10.1287/mnsc.27.4.479.

[34] Sartor, F., and Timme, S., "Delayed Detached–Eddy Simulation of Shock Buffet on Half Wing–Body Configuration," *AIAA Journal*, Vol. 55, No. 4, 2017, pp. 1230–1240. https://doi.org/10.2514/1.J055186.

[35] Masini, L., Timme, S., and Peace, A., "Scale-Resolving Simulations of a Civil Aircraft Wing Transonic Shock-Buffet Experiment," *AIAA Journal*, Vol. 58, No. 10, 2020, pp. 4322–4338. https://doi.org/10.2514/1.J059219.

[36] Lawson, S., Greenwell, D., and Quinn, M. K., "Characterisation of Buffet on a Civil Aircraft Wing," *54th AIAA Aerospace Sciences Meeting*, American Institute of Aeronautics and Astronautics, San Diego, CA, 2016. https://doi.org/10.2514/6.2016-1309.

[37] Martineau, D., Stokes, S., Munday, S., Jackson, A., Gribben, B., and Verhoeven, N., "Anisotropic Hybrid Mesh Generation for Industrial RANS Applications," *44th AIAA Aerospace Sciences Meeting and Exhibit*, American Institute of Aeronautics and Astronautics, Reno, NV, 2006. https://doi.org/10.2514/6.2006-534.

[38] Schwamborn, D., Gerhold, T., and Heinrich, R., "The DLR TAU-Code: Recent applications in research and industry," *ECCOMAS CFD 2006 CONFERENCE*, 2006.

[39] Allmaras, S. R., Forrester, J. T., and Spalart, P. R., "Modifications and Clarifications for the Implementation of the Spalart-Allmaras Turbulence Model," *Seventh International Conference on Computational Fluid Dynamics*, ICCFD7, Big Island, HI, 2012. URL https://www.iccfd.org/iccfd7/assets/pdf/papers/ICCFD7-1902_paper.pdf.

[40] Maskew, B., "A Computer Program for Calculating Nonlinear Aerodynamic Characteristics of Arbitrary Configurations," *Program VSAERO Theory Document*, NASA Contractor Report 4023, NASA, Ames Research Center, 1987.

[41] Matthews, A. G. d. G., van der Wilk, M., Nickson, T., Fujii, K., Boukouvalas, A., León-Villagrá, P., Ghahramani, Z., and Hensman, J., "GPflow: A Gaussian process library using TensorFlow," *Journal of Machine Learning Research*, Vol. 18, No. 40, 2017, pp. 1–6. URL http://jmlr.org/papers/v18/16-537.html.

[42] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," , 2015. Software available from https://www.tensorflow.org/.