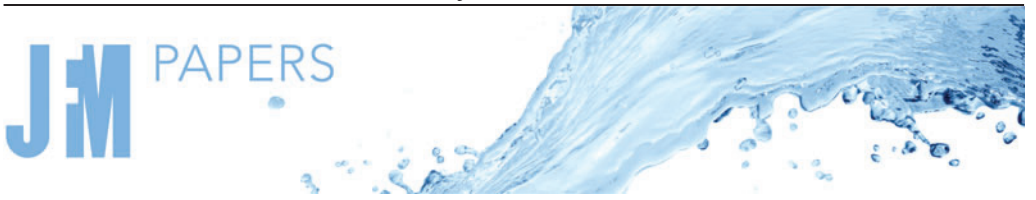
J. Fluid Mech. (2023), *vol*. 964, A13, doi:10.1017/jfm.2023.327



Efficient prediction of turbulent flow quantities using a Bayesian hierarchical multifidelity model

S. Rezaeiravesh[1,2,3,†], T. Mukha[2,3] and P. Schlatter[2,3,4]

[1]Department of Fluids and Environment, The University of Manchester, Manchester M13 9PL, UK

[2]SimEx/FLOW, Engineering Mechanics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

[3]Swedish e-Science Research Centre (SeRC), SE-100 44 Stockholm, Sweden

[4]Institute of Fluid Mechanics (LSTM), Friedrich–Alexander Universität Erlangen–Nürnberg (FAU), DE-91058 Erlangen, Germany

(Received 1 November 2022; revised 13 March 2023; accepted 11 April 2023)

Multifidelity models (MFMs) can be used to construct predictive models for flow quantities of interest (QoIs) over the space of uncertain/design parameters, with the purpose of uncertainty quantification, data fusion and optimization. For numerical simulation of turbulence, there is a hierarchy of methodologies ranked by accuracy and cost, where each methodology may have several numerical/modelling parameters that control the predictive accuracy and robustness of its resulting outputs. Compatible with these specifications, the present hierarchical MFM strategy allows for simultaneous calibration of the fidelity-specific parameters in a Bayesian framework as developed by Goh *et al.* (*Technometrics*, vol. 55, no. 4, 2013, pp. 501–512). The purpose of the MFM is to provide an improved prediction, mainly interpolation over the range covered by training data, by combining lower- and higher-fidelity data in an optimal way for any number of fidelity levels; even providing confidence intervals for the resulting QoI. The capabilities of the MFM are first demonstrated on an illustrative toy problem, and it is then applied to three realistic cases relevant to engineering turbulent flows. The latter include the prediction of friction at different Reynolds numbers in turbulent channel flow, the prediction of aerodynamic coefficients for a range of angles of attack of a standard airfoil and the uncertainty propagation and sensitivity analysis of the separation bubble in the turbulent flow over periodic hills subject to geometrical uncertainties. In all cases, based on only a few high-fidelity data samples, the MFM leads to accurate predictions of the QoIs. The result of the uncertainty quantification and sensitivity analyses are also found to be accurate compared with the ground truth in each case.

† Email address for correspondence: saleh.rezaeiravesh@manchester.ac.uk

Key words: computational methods, machine learning, turbulence simulation

1. Introduction

In science and engineering, different computational models can be derived to make realizations of the quantities of interest (QoIs) of a process or an event happening in reality. High-fidelity (HF) models can result in highly accurate and robust realizations, but running them is typically computationally expensive. In contrast, different low-fidelity (LF) models with lower computational cost can be developed for the same process which, however, potentially lead to lower accuracy QoIs due to partial or completely missing physics captured by the model. On the other hand, in different applications arising in uncertainty quantification (UQ), data fusion, and optimization, of numerous realizations of the QoIs are required, associated with the samples taken from the space of the inputs/parameters in order to make reliable estimations (these non-intrusive problems are referred to as the outer-loop problems). In this regard, multifidelity models (MFMs) can be constructed by combining realizations of the HF and LF models such that a balance between the overall computational cost and predictive accuracy is achieved. The goal is to provide, by combining HF and LF models, an estimate of the QoI that is better than any of the models alone.

In the recent years, different types of MFMs have been applied to a wide range of problems, see e.g. the recent review by Peherstorfer, Willcox & Gunzburger (2018). The use of the MFMs in studies of turbulent flows can be greatly advantageous, considering the wide range of engineering applications relying on these flows and also the high cost generally involved in the HF computations (such as scale-resolving simulations) and experiments of the turbulent flows. There is a distinguishable hierarchy in the fidelity of the computational models utilized for simulation of turbulence, (see e.g. Sagaut, Deck & Terracol 2013). Let us consider the wall-bounded turbulent flows where a turbulent boundary layer forms at the wall boundaries. Direct numerical simulation (DNS) can provide the highest-fidelity results for a given turbulent flow, however, it can become prohibitively expensive at high Reynolds numbers which are relevant to practical applications. The computational cost can be reduced by employing large eddy simulation (LES) which aims at directly resolving the scales larger than a defined size and modelling the unresolved effects. At the lowest cost and fidelity level, Reynolds-averaged Navier–Stokes (RANS) simulations can be performed which avoid directly resolving any flow fluctuations by resorting to a statistical description of turbulence. Between RANS and wall-resolving LES, other approaches such as hybrid RANS–LES and wall-modelled LES can be considered (see Sagaut *et al.* 2013; Larsson *et al.* 2016). Although this clear hierarchy is extremely beneficial when constructing MFMs, as will be thoroughly discussed and demonstrated in the present paper, there is a challenge to be dealt with: the realizations of different turbulence simulation approaches are, in general, sensitive to various modelling and numerical parameters as well as inputs. At lower fidelities like RANS, modelling effects are dominant while as moving towards LES and DNS, numerical factors become more relevant, including grid resolution and discretization properties. Hereafter, these fidelity-specific controlling parameters are referred to as tuning or calibration parameters.

Combining training data from different turbulence simulation approaches, MFMs are constructed over the space of design and uncertain parameters/inputs. An appropriate approach to construct MFMs for turbulent flow problems should systematically allow for

simultaneous calibration of the tuning parameters. An appropriate methodology which is employed in the present study is the hierarchical multifidelity predictive model developed in Higdon *et al.* (2004) and Goh *et al.* (2013) in which the calibration parameters of the involved fidelities are estimated using the data of the higher-fidelity models. This MFM, which is hereafter referred to as HC-MFM, can also incorporate the observational uncertainties. The HC-MFM can be seen as an extension of the model by Higdon *et al.* (2004) which was employed to combine experimental (field) and simulation data. A fundamental component of this class of MFMs is the Bayesian calibration of the computer models, as described in the landmark paper by Kennedy & O'Hagan (2001). At each level of the MFM, the Gaussian process regression (GPR) (Rasmussen & Williams 2005) is employed to construct surrogates for the simulators.

The application of the HC-MFM in the field of computational fluid dynamics (CFD) and turbulent flows is novel, and we make specific adaptations suitable for turbulence simulations. In this regard, the present paper aims at assessing the useful potential of the HC-MFM by applying it to three examples relevant to engineering wall-bounded turbulent flows. To highlight the contributions of the present work, the existing studies in the literature devoted to the development and application of the MFMs to CFD and turbulent flows are briefly reviewed here by classifying them according to their underlying MFM strategy. (i) A model was originally introduced by Kennedy & O'Hagan (2000) where a QoI at each fidelity is expressed as a first-order autoregressive model of the same QoI at the immediately lower fidelity. Co-kriging using GPR to construct surrogates is classified in this category, see e.g. Fatou Gomez (2018); Voet *et al.* (2021) for applications to turbulence simulations. To enhance the computational efficiency of the co-kriging for several fidelity levels, recursive algorithms have been proposed and applied to CFD problems, see Gratiet & Garnier (2014); Perdikaris *et al.* (2015). (ii) A class of MFMs has been developed based on non-intrusive polynomial chaos expansion (PCE) and stochastic collocation methods (Ng & Eldred 2012; Palar, Tsuchiya & Parks 2016), where an additive or a multiplicative term is considered to correct the LF model's predictions against the HF model. (iii) The multi-level multifidelity Monte Carlo (MLMF-MC) models (Fairbanks *et al.* 2017; Geraci, Eldred & Iaccarino 2017) are appropriate for the UQ forward problems. These models are developed by combining multilevel (Giles 2008) and control-variate (Pasupathy *et al.* 2012) MC methods to improve the rate of convergence of the stochastic moments of the QoIs compared with the the standard MC method. Jofre *et al.* (2018) applied MLMF-MC models to an irradiated particle-laden turbulent flow. The HF model was considered to be DNS and the two LF models were based on a surrogate particle approach and lower resolutions for flow and particles. (iv) Other models including the hierarchical kriging model based on GPR where the predictions of a LF model are taken as the trend in the HF kriging, see Han & Görtz (2012).

Recently, Voet *et al.* (2021) compared inverse weighted distance-, PCE- and co-kriging-based MFMs using the data of RANS and DNS for the turbulent flow over a periodic hill, and concluded that the co-kriging model outperforms the others in terms of accuracy. This is the first (and to our knowledge only) study where MFMs have been applied to engineering-relevant RANS and DNS data for the purpose of uncertainty propagation. Voet *et al.* (2021) also found that the performance of the co-kriging can deteriorate when there is no significant correlation between the RANS and DNS data and at the same time there is a significant deviation between them. Motivated by this deficiency, we adapt and use the HC-MFM where the discrepancy between the data (and not their correlation) over the space of design parameters is learned using independent Gaussian processes. The model is absolutely generative and can be extended to an arbitrary

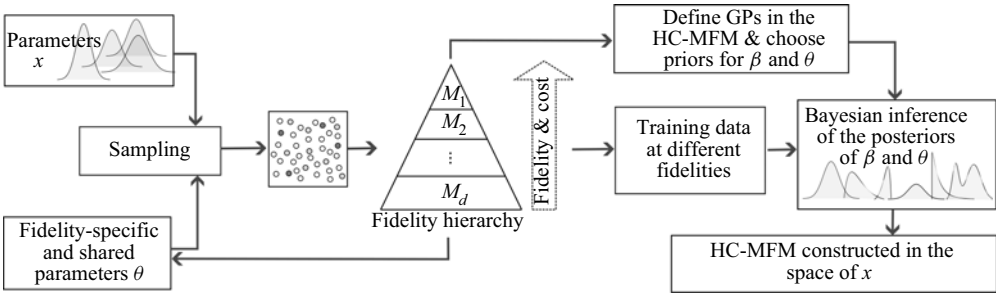


Figure 1. Schematic representation of the machinery for constructing the HC-MFM over the space of design/uncertain parameters $\boldsymbol{x}$. To make realizations of the QoI of the problem in hand, there is a hierarchy of fidelities where each fidelity may have its own parameters and also some parameters shared with others (together called $\boldsymbol{\theta}$). For joint samples of $\boldsymbol{x}$ and $\boldsymbol{\theta}$, training data for the QoI are generated (or are available) in a way that more samples are taken for LF models which are less costly to evaluate. The priors for $\boldsymbol{\theta}$ and GPs' hyperparameters $\boldsymbol{\beta}$ are defined after the HC-MFM is formulated as, for instance, (2.6). Given the training data, an MCMC sampling method is used to infer the posteriors of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$.

number of fidelity levels. Besides the systematic calibration of the fidelity-specific parameters during its training stage, the HC-MFM is also capable of handling uncertain data, as for instance happens when QoIs are turbulence statistics computed over a finite time-averaging interval (recently, a framework was proposed by Rezaeiravesh, Vinuesa & Schlatter (2022) to combine these observational uncertainties with parametric ones). Relying on these characteristics, the HC-MFM is suitable for application to the data of various turbulence simulation methodologies to address different types of the outer-loop problems. In contrast to all the previous studies (at least in CFD), we adopt a Bayesian inference to construct the HC-MFM, a feature which results in more accurate models as well as estimating confidence intervals for the predictions.

The rest of the paper is organized as follows. In § 2, various elements of the HC-MFM approach are introduced and explained. Section 3 is devoted to the application of the HC-MFM to an illustrative example, turbulent channel flow, polars for an airfoil and analysis of the geometrical uncertainties in the turbulent flow over a periodic hill. The summary of the paper along with the conclusions is presented in § 4.

2. Method

In this section, the hierarchical MFM with calibration (HC-MFM) developed by Goh *et al.* (2013), which forms the basis for the present study is reviewed. We will proceed by sequentially going through the aspects of GPR, model calibration and eventually the HC-MFM formulation. The workflow of the HC-MFM represented in figure 1 may help connect the following technical details.

2.1. *Gaussian process regression*

In general, the Gaussian processes (GPs) provide a way to systematically build a representation of the QoI as a function of the various inputs to the model. Eventually, regression can be performed by evaluating the GP at new inputs not seen by the model before. Let $\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^{p_x}$ represent the controllable inputs and parameters, adopting the notation from Kennedy & O'Hagan (2001). As a convention, all boldface letters are hereafter considered to be a vector or a matrix. The design and uncertain parameters appearing in optimization and UQ analyses, respectively, can also be classified as $\boldsymbol{x}$.

A GP $\hat{f}(\boldsymbol{x})$ can be employed to map the inputs $\boldsymbol{x}$ to a QoI or an output $y \subset \mathbb{R}$ of the computer codes (simulators) or field data. For a finite set of training samples $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ with corresponding observations $\{y_1, y_2, \ldots, y_n\}$, the collection of $\{\hat{f}(\boldsymbol{x}_1), \hat{f}(\boldsymbol{x}_2), \ldots, \hat{f}(\boldsymbol{x}_n)\}$ will have a joint Gaussian (multivariate normal) distribution (Rasmussen & Williams 2005). The GP $\hat{f}(\boldsymbol{x})$ is written as

$$\hat{f}(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')), \tag{2.1}$$

which is fully described by its mean $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ defined as

$$m(\boldsymbol{x}) = \mathbb{E}[\hat{f}(\boldsymbol{x})], \tag{2.2}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(\hat{f}(\boldsymbol{x}) - m(\boldsymbol{x}))(\hat{f}(\boldsymbol{x}') - m(\boldsymbol{x}'))]. \tag{2.3}$$

In general, the GPs can be used in the case of having observation noise $\boldsymbol{\varepsilon}$ in the y data. Using an additive error model, we have

$$y(\boldsymbol{x}) = \hat{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \tag{2.4}$$

where the noises are assumed to be independent and have Gaussian distribution $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2)$.

In the GPR, given a set of training data $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$ the posterior and posterior predictive distributions of $\hat{f}(\cdot)$ and y, respectively, at test inputs $\boldsymbol{x}^* \in \mathbb{X}$, can be inferred in a Bayesian framework (Rasmussen & Williams 2005). To this end, first a prior distribution for $\hat{f}(\boldsymbol{x})$, see (2.1), is assumed through specifying functions for the mean and covariance in (2.2) and (2.3), where there are unknown hyperparameters $\boldsymbol{\beta}$ in the functions. Using the training data, the posterior distribution of $\boldsymbol{\beta}$ is learned. As a main advantage of the GPR, the predictions at test samples will be accompanied by an estimate of uncertainty, see (A1) and (A2) in Appendix A.

2.2. *Model calibration*

As pointed out in § 1, the outputs of computational models (simulators) at a given $\boldsymbol{x}$ may depend on different tuning or calibration parameters, $\boldsymbol{t} \in \mathbb{T} \subset \mathbb{R}^{p_t}$. Given a set of observations, these parameters can be calibrated through conducting a UQ inverse problem which can be expressed in a Bayesian framework (Kennedy & O'Hagan 2001). The calibrated model can then not only be employed for prediction, but also for fusion of the field and simulation data, see Higdon *et al.* (2004). Consider n_1 data samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n_1}$ are observed for a physical process $\zeta(\boldsymbol{x})$. To statistically model the observations, a simulator $\hat{f}(\boldsymbol{x}, \boldsymbol{\theta})$ can be employed in which the $\boldsymbol{\theta}$ are the true or optimal values of $\boldsymbol{t}$ and are to be estimated from the training data. However, in general, it is possible that even the calibrated simulator $\hat{f}(\boldsymbol{x}, \boldsymbol{\theta})$ produces observations which systematically deviate from reality. To remove such a bias, a model discrepancy term $\hat{\delta}(\boldsymbol{x})$ can be added to the simulator (see Kennedy & O'Hagan 2001; Higdon *et al.* 2004). In many practical applications, particularly in CFD and turbulent flow simulations where the computational cost can be excessively high, the restrictions of the computational budget only allows for a limited number of simulations. In any realization, the adopted values for the tuning parameters $\boldsymbol{t}$ are not necessarily optimal and hence potentially lead to the outputs which are systematically different from the QoIs in reality. For the described calibration problem,

the Kennedy & O'Hagan (2001) model reads as

$$\left.\begin{array}{ll} y_i = \hat{f}(\boldsymbol{x}_i, \boldsymbol{\theta}) + \hat{\delta}(\boldsymbol{x}_i) + \varepsilon_i & i = 1, 2, \ldots, n_1 \\ y_i = \hat{f}(\boldsymbol{x}_i, \boldsymbol{t}_i) & i = 1 + n_1, 2 + n_1, \ldots, n_2 + n_1 \end{array}\right\}, \tag{2.5}$$

where $\hat{\cdot}$ specifies a GP and n_2 is the number of simulated data. Note that the samples $\{\boldsymbol{x}_i\}_{i=1+n_1}^{n_2+n_1}$ are not necessarily the same as $\{\boldsymbol{x}_i\}_{i=1}^{n_1}$ at which the observations are made. The index i should be seen as a global index which implies that a different model is used for each of the two subranges of i. Given the $n_1 + n_2$ data, the posterior distribution for the calibration parameters $\boldsymbol{\theta}$ along with that of the hyperparameters in the GPs, $\boldsymbol{\beta}$, is estimated. Further details are provided in the section below.

2.3. *The hierarchical MFM with automatic calibration (HC-MFM)*

Goh *et al.* (2013) extended the model (2.5) to an arbitrary number of fidelity levels which together form a modelling hierarchy for a physical process. As a main feature of the resulting MFM, each fidelity can, in general, have its own calibration parameters and also share some calibration parameters with other fidelities. The basics of the MFM comprising three fidelity levels are explained below, noting that adapting the formulation to any number of fidelities with different combinations of parameters is straightforward. We assume that the fidelity of the models decreases from M_1 to M_3, and in practice due to the budget limitations, the number of training data decreases with increasing the model fidelity. The HC-MFM for three fidelities reads as

$$\left.\begin{array}{ll} y_{M_1}(\boldsymbol{x}_i) = \hat{f}(\boldsymbol{x}_i, \boldsymbol{\theta}_3, \boldsymbol{\theta}_s) + \hat{g}(\boldsymbol{x}_i, \boldsymbol{\theta}_2, \boldsymbol{\theta}_s) + \hat{\delta}(\boldsymbol{x}_i) + \varepsilon_{1_i} & i = 1, 2, \ldots, n_1 \\ y_{M_2}(\boldsymbol{x}_i) = \hat{f}(\boldsymbol{x}_i, \boldsymbol{\theta}_3, \boldsymbol{t}_{s_i}) + \hat{g}(x_i, \boldsymbol{t}_{2_i}, \boldsymbol{t}_{s_i}) + \varepsilon_{2_i} & i = 1 + n_1, 2 + n_1, \ldots, n_2 + n_1 \\ y_{M_3}(\boldsymbol{x}_i) = \hat{f}(\boldsymbol{x}_i, \boldsymbol{t}_{3_i}, \boldsymbol{t}_{s_i}) + \varepsilon_{3_i} & i = 1 + n_1 + n_2, \ldots, n_3 + n_1 + n_2 \end{array}\right\}, \tag{2.6}$$

where subscript s denotes the parameters which are shared between the models, whereas $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$ are the calibration parameters specific to fidelities M_2 and M_3, respectively. The noises are assumed to have Gaussian distributions with zero mean. At each fidelity level, the associated simulator is created by adding a model discrepancy term to the simulator describing the immediately lower fidelity. Concatenating all training data, an augmented vector $\boldsymbol{Y}$ of size $n_1 + n_2 + n_3$ is obtained, for which the covariance matrix can be written in terms of the covariances of $\hat{f}(\cdot)$, $\hat{g}(\cdot)$, $\hat{\delta}(\cdot)$ and the observational noise:

$$\begin{aligned} \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_f + \begin{bmatrix} \boldsymbol{\Sigma}_g & \boldsymbol{0}_{(n_1+n_2)\times n_3} \\ \boldsymbol{0}_{n_3\times(n_1+n_2)} & \boldsymbol{0}_{n_3\times n_3} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Sigma}_\delta & \boldsymbol{0}_{n_1\times(n_2+n_3)} \\ \boldsymbol{0}_{(n_2+n_3)\times n_1} & \boldsymbol{0}_{(n_2+n_3)\times(n_2+n_3)} \end{bmatrix} \\ + \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon_1} & \boldsymbol{0}_{n_1\times n_2} & \boldsymbol{0}_{n_1\times n_3} \\ \boldsymbol{0}_{n_2\times n_1} & \boldsymbol{\Sigma}_{\varepsilon_2} & \boldsymbol{0}_{n_2\times n_3} \\ \boldsymbol{0}_{n_3\times n_1} & \boldsymbol{0}_{n_3\times n_2} & \boldsymbol{\Sigma}_{\varepsilon_3} \end{bmatrix}. \end{aligned} \tag{2.7}$$

Appropriate kernel functions should be chosen to express the structure of the covariances. Using samples i and j of the inputs and parameters, the associated element in the

covariance matrix $\boldsymbol{\Sigma}_f$ will be obtained from

$$\boldsymbol{\Sigma}_{f_{ij}} = \mathrm{cov}(\boldsymbol{x}_i, \boldsymbol{t}_{3_i}, \boldsymbol{t}_{s_i}, \boldsymbol{x}_j, \boldsymbol{t}_{3_j}, \boldsymbol{t}_{s_j}) = \lambda_f^2 k(\bar{d}_{f_{ij}}), \tag{2.8}$$

where λ_f is a hyperparameter and $\bar{d}_{f_{ij}}$ is the scaled Euclidean distance between two samples i and j over the space of $(\boldsymbol{x}, \boldsymbol{t}_3, \boldsymbol{t}_s)$

$$\bar{d}^2_{f_{ij}} = \bar{d}^2(\boldsymbol{x}_i, \boldsymbol{x}_j) + \bar{d}^2(\boldsymbol{t}_{3_i}, \boldsymbol{t}_{3_j}) + \bar{d}^2(\boldsymbol{t}_{s_i}, \boldsymbol{t}_{s_j}) \tag{2.9}$$

$$= \sum_{l=1}^{p_{\boldsymbol{x}}} \frac{(x_{l_i} - x_{l_j})^2}{\ell^2_{f_{x_l}}} + \sum_{l=1}^{p_{\boldsymbol{t}_3}} \frac{(t_{3_{l_i}} - t_{3_{l_j}})^2}{\ell^2_{f_{t_{3l}}}} + \sum_{l=1}^{p_{\boldsymbol{t}_s}} \frac{(t_{s_{l_i}} - t_{s_{l_j}})^2}{\ell^2_{f_{t_{sl}}}}. \tag{2.10}$$

Here, $p_{\boldsymbol{x}}, p_{\boldsymbol{t}_3}$ and $p_{\boldsymbol{t}_s}$ specify the dimensions of $\boldsymbol{x}$, $\boldsymbol{t}_3$ and $\boldsymbol{t}_s$, respectively. Correspondingly, the length scale over the lth dimension of each of these spaces is represented by $\ell_{f_{x_l}}$, $\ell_{f_{t_{3l}}}$ and $\ell_{f_{t_{sl}}}$, respectively. These length scales are among the hyperparameters $\boldsymbol{\beta}$ to be learned when constructing the HC-MFM. There are various options for modelling the covariance kernel function $k(\cdot)$, see e.g. Rasmussen & Williams (2005) and Gramacy (2020), among which the exponentiated quadratic and Matern-5/2 (Matern 1986) functions are used in the examples in § 3. These two functions respectively read as

$$k(\bar{d}_{f_{ij}}) = \exp(-0.5\bar{d}^2_{f_{ij}}) \tag{2.11}$$

and

$$k(\bar{d}_{f_{ij}}) = [1 + \sqrt{5}\bar{d}_{f_{ij}} + \tfrac{5}{3}\bar{d}^2_{f_{ij}}] \exp(-\sqrt{5}\bar{d}_{f_{ij}}). \tag{2.12}$$

Similar expressions can be derived for $\boldsymbol{\Sigma}_{g_{ij}} = k_g(\boldsymbol{x}_i, \boldsymbol{t}_{2_i}, \boldsymbol{t}_{s_i}, \boldsymbol{x}_j, \boldsymbol{t}_{2_j}, \boldsymbol{t}_{s_j})$ and $\boldsymbol{\Sigma}_{\delta_{ij}} = k_\delta(\boldsymbol{x}_i, \boldsymbol{x}_j)$ appearing in (2.7). This leads to introducing new hyperparameters associated with the GPs. Note that, given how the training vector $\boldsymbol{Y}$ and associated inputs are assembled, correct combinations of training data for the inputs and parameters will be used in the kernels. The unknown parameters to estimate include calibration parameters in different models, $\boldsymbol{\theta}$, and hyperparameters $\boldsymbol{\beta}$ appearing in the GPs. Following the Bayes rule, the posterior distribution of these parameters given the training data $\boldsymbol{Y}$ can be inferred from (Kennedy & O'Hagan 2001; Higdon *et al.* 2004; Goh *et al.* 2013)

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{Y}) \propto \pi(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\beta})\pi_0(\boldsymbol{\theta})\pi_0(\boldsymbol{\beta}), \tag{2.13}$$

where $\pi(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\beta})$ specifies the likelihood function and $\pi_0(\cdot)$ represents a prior distribution. Note that all priors are assumed to be independent. For all GPs in § 3, the prior distribution for λ appearing in the covariance matrices such as (2.8) is taken to be half-Cauchy whereas the length scales ℓ in (2.11) and (2.12) are assumed to have gamma distributions. The exact definition of the priors will be provided later for each case in § 3, and table 2 in Appendix B summarizes the formulation of the standard distributions used as priors. The standard deviations of the noises are assumed to be the same for which a half-Cauchy prior is adopted. For the calibration parameters $\boldsymbol{\theta}$, Gaussian or uniform priors are considered. In some cases, we may consider a constant mean function for the GPs, where a Gaussian distribution is used as the prior. Due to this, the predictions of a trained MFM when it is used to extrapolate in $\boldsymbol{x}$ (outside of the range of training samples) should be used with caution. To avoid potential inaccuracies, in general, more elaborate mean functions can be used when constructing the HC-MFMs (this, however, is not the subject of the present study). Further details about the choice of the kernels and priors for the

parameters/hyperparameters as well as the use of the HC-MFM for extrapolation can be found in Appendix A.

Given the training data $\boldsymbol{Y}$, a Markov chain Monte Carlo (MCMC) technique can be used to draw samples from the posterior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, and hence construct a HC-MFM. In the present study, the described HC-MFM (2.6) has been implemented in Python using the PyMC3 (Salvatier, Wiecki & Fonnesbeck 2016) package with the no-U-turn sampler (NUTS) MCMC sampling approach (Hoffman & Gelman 2014). As it will be shown in § 3.6, the MCMC sampling method may lead to more accurate results compared with the point estimators.

After being constructed, an HC-MFM can be used for predicting the QoI y for any new sample taken from the space of inputs $\boldsymbol{x}$. The accuracy of the predicted QoIs will be assessed by measuring their deviation from the validation data of the highest fidelity M_1. As detailed in Goh *et al.* (2013), the joint distribution of the training $\boldsymbol{Y}$ and new y^* (associated with a test sample $\boldsymbol{x}^*$) conditioned on $\boldsymbol{\theta}, \boldsymbol{\beta}$ will have a multivariate normal distribution with a covariance matrix of the same structure as $\boldsymbol{\Sigma}$ in (2.7). For any joint sample drawn from the posterior distribution of $\pi(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{Y})$, a sample prediction for y^* is made. Repeating this procedure for a large number of times, valid estimations for the posterior of the predictions y^* can be achieved. Therefore, estimating the confidence in the predictions is straightforward. Note that at this stage, various UQ analyses can be performed using the HC-MFM as a surrogate of the physical process over $\boldsymbol{x}$.

3. Results and discussion

Four examples are considered to which the HC-MFM described in the previous section is applied. The first example in § 3.1 is used to validate the implementation of the MFM, and the next three examples are relevant to fundamental and engineering analysis of wall-bounded turbulent flows.

3.1. *An illustrative example*

Consider the following analytical model taken from Forrester, Sóbester & Keane (2007) to generate HF and LF samples of the QoI y for input $x \in [0, 1]$:

$$\left.\begin{aligned} y_H(x) &= (\theta x - 2)^2 \sin(2\theta x - 4) \\ y_L(x) &= y_H(x) + B(x - 0.5) - C \end{aligned}\right\}. \tag{3.1}$$

In Forrester *et al.* (2007), θ is taken to be fixed and equal to 6, but here it is treated as an uncertain calibration parameter that is to be estimated during the construction of the MFM. Note that the notations of the general model (2.5) can be adopted for (3.1). For simulator $\hat{f}(x, t)$ and model discrepancy $\hat{\delta}(x)$ the covariance matrix in (2.8) is used with the exponentiated quadratic kernel (2.11). The following prior distributions are considered: $\lambda_f, \lambda_\delta \sim \mathcal{HC}(\alpha = 5), \ell_{f_x}, \ell_{f_t}, \ell_{\delta_x} \sim \Gamma(\alpha = 1, \beta = 5), \varepsilon \sim \mathcal{N}(0, \sigma)$ with $\sigma \sim \mathcal{HC}(\alpha = 5)$ and $\theta \sim \mathcal{U}[5.8, 6.2]$. Here, $\mathcal{HC}$, Γ, $\mathcal{N}$ and $\mathcal{U}$ denote the half-Cauchy, gamma, Gaussian, and uniform distributions, respectively, see table 2. The HF training samples are taken at $x = \{0, 0.4, 0.6, 1\}$, therefore, $n_H = 4$ is fixed. To investigate the effect of n_L, three sets of LF samples of size 10, 15 and 20 are considered which are generated by Latin hypercube sampling from the admissible space $[0, 1] \times [5.8, 6.2]$ corresponding to x and t (uncalibrated instance of θ), respectively.

Using the data, the HC-MFM (2.6) for problem (3.1) is constructed. The first row in figure 2 shows the predicted y with the associated 95 % confidence interval (CI) along with the training data and reference true data generated with $\theta = 6$. For all n_L, the predicted y is

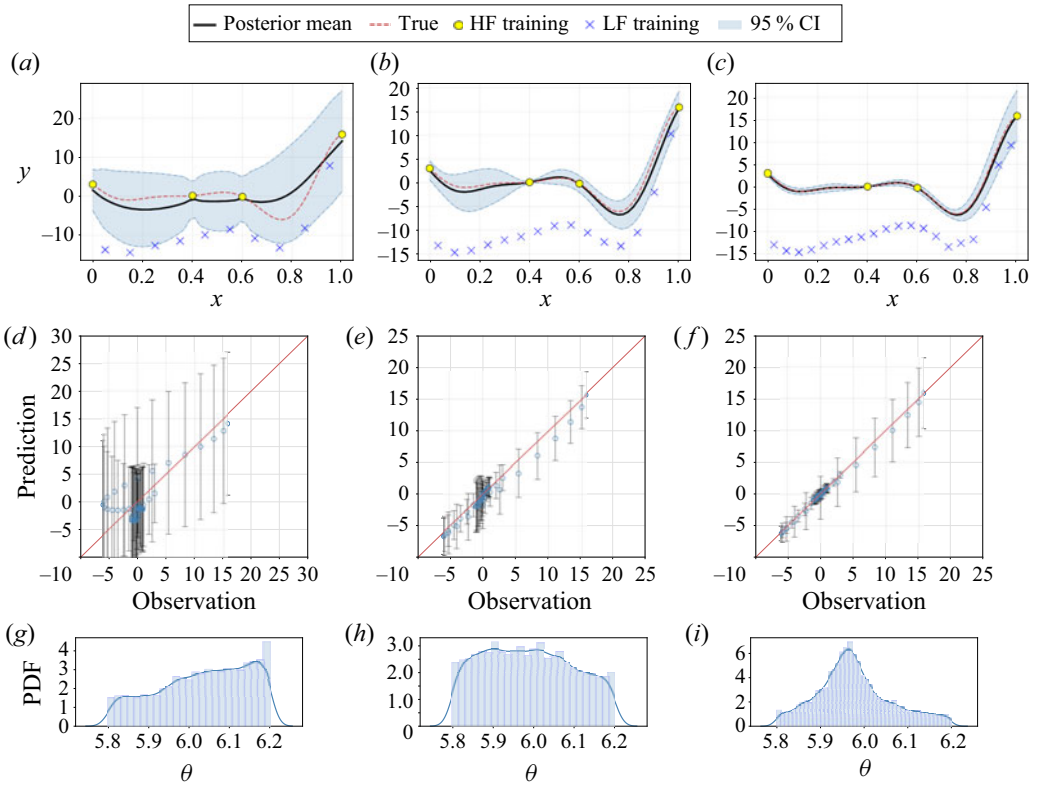


Figure 2. (*a–c*) Predicted QoI y by HC-MFM (2.6) along with the training and true data, (*d–f*) predicted y vs true observations at 50 test samples of $x \in [0, 1]$ with error bars representing 95 % CI, (*g–i*) posterior probability density function (PDF) of θ based on 10^4 MCMC samples. The y_H and y_L training data are generated from (3.1) using $B = C = 10$. The training data include 4 HF samples combined with (left column) 10, (middle column) 15 and (right column) 20 LF samples. The true data are generated by (3.1) using $\theta = 6$.

closer to HF data than the LF data, however, for $n_L = 15$ and 20, the agreement between the mean of the predicted y and the true data is significantly improved. A better validation can be made via the plots in the second row of figure 2, where the predicted y and true values of y_H at 50 uniformly spaced test samples for $x \in [0, 1]$ are plotted. Clearly, increasing the number of the LF samples while keeping $n_H = 4$ fixed improves the predictions and reduces the uncertainty. In the third row of figure 2, the posterior densities of θ are presented. In all cases, a uniform (non-informative) prior distribution over [5.8, 6.2] was considered for θ. Only for $n_L = 20$, the resulting posterior density of θ is high near the true value 6. Therefore, it is confirmed that, as explained by Goh *et al.* (2013), the main capability of the HC-MFM (2.6) is in making accurate predictions for y and only if a sufficient number of training data are available, accurate distributions for the calibration parameters are also obtained. This is shown here by fixing n_H and increasing n_L, which is favourable in practice. It is also noteworthy that if θ was known and hence treated as a fixed parameter, then even with $n_L = 10$ very accurate predictions for y could be already achieved (not shown here).

It also should be noted that the mean of the posterior distribution of σ, the noise standard deviation, is found to be negligible, as expected. This is in fact the case for all other examples in this section.

3.2. *Turbulent channel flow*

Turbulent channel flow is one of the most canonical wall-bounded turbulent flows. The flow develops between two parallel flat walls which are apart by the distance 2δ, and the flow is periodic in the streamwise and spanwise directions. Channel flow is fully defined by a Reynolds number, for instance the bulk Reynolds number $Re_b = U_b\delta/\nu$, where U_b and ν specify the streamwise bulk velocity and kinematic viscosity, respectively. Among different possible QoIs, here we only focus on the friction velocity $\langle u_\tau\rangle$, as a function of Reynolds number. This quantity is defined as $\sqrt{\langle\tau_w\rangle/\rho}$, where τ_w and ρ are the magnitude of the wall-shear stress and fluid density, respectively, and $\langle\cdot\rangle$ represents averaging over time and the periodic directions. Three fidelity levels are considered: DNS (M_1), wall-resolved LES (WRLES) (M_2) and a reduced-order algebraic model (M_3), where the fidelity reduces from the former to the latter. We use the DNS data of Iwamoto, Suzuki & Kasagi (2002), Lee & Moser (2015) and Yamamoto & Tsuji (2018).

The WRLES of channel flow have been performed at different Reynolds numbers without any explicit subgrid-scale model using OpenFOAM (Weller *et al.* 1998), which is an open-source finite-volume flow solver. Linear interpolation is used for the evaluation of face fluxes, and a second-order backward-differencing scheme is used for time integration. The pressure-implicit with splitting of operators (PISO) algorithm is used for pressure–velocity coupling. For further details on the simulation set-up, see Rezaeiravesh & Liefvendahl (2018). The results in that paper show that for a fixed resolution in the wall-normal direction, variation of the grid resolutions in the wall-parallel directions could significantly impact the accuracy of the flow QoIs. Therefore, in the context of the HC-MFM, the calibration parameters for WRLES are taken to be Δx^+ and Δz^+, which are the cell dimensions in the streamwise and spanwise directions, respectively, expressed in wall units ($\Delta x^+ = \Delta x u_\tau^\circ/\nu$ where u_τ° is the reference u_τ from DNS).

At the lowest fidelity, the following reduced-order algebraic model is considered which is derived by averaging the streamwise momentum equation for the channel flow in the periodic directions and time:

$$\langle u_\tau\rangle^2/U_b^2 = \frac{1}{Re_b}\frac{\mathrm{d}}{\mathrm{d}\eta}\left((1+\zeta(\eta))\frac{\mathrm{d}\langle u\rangle/U_b}{\mathrm{d}\eta}\right), \tag{3.2}$$

where η is the distance from the wall normalized by the channel half-height δ, and $\zeta(\eta)$ is the normalized eddy viscosity ν_t, i.e. $\zeta(\eta) = \nu_t(\eta)/\nu$. Reynolds & Tiederman (1967) proposed the following closed form for $\zeta(\eta)$:

$$\zeta(\eta) = \frac{1}{2}\left[1 + \frac{\kappa^2 Re_\tau^2}{9}(1-(\eta-1)^2)^2(1+2(\eta-1)^2)^2\left(1-\exp\left(\frac{-\eta Re_\tau}{A^+}\right)\right)^2\right]^{1/2} - \frac{1}{2}, \tag{3.3}$$

where $Re_\tau = \langle u_\tau\rangle\delta/\nu$ is the friction-based Reynolds number, and κ and A^+ are two modelling parameters. At any Re_b (and given value of κ and A^+), (3.2) is integrated over $\eta \in [0, 1]$ and is iteratively solved using (3.3) to estimate $\langle u_\tau\rangle$. Expressing the channel flow example in the terminology of MFM (2.6), $\langle u_\tau\rangle/U_b$ is the QoI y, $x = Re_b$, $\boldsymbol{t}_3 = (\kappa, A^+)$ and $\boldsymbol{t}_2 = (\Delta x^+, \Delta z^+)$. The training data set consists of the following databases. For DNS, $\langle u_\tau\rangle$ is taken from Iwamoto *et al.* (2002), Lee & Moser (2015) and Yamamoto & Tsuji (2018) at $Re_b = 5020$, 6962, 10 000, 20 000, 125 000 and 200 400. In total, 16 WRLES $\langle u_\tau\rangle$ samples are obtained from a design of experiment based on the prior distributions $\Delta x^+ \sim \mathcal{U}[15, 85]$ and $\Delta z^+ \sim \mathcal{U}[9.5, 22]$ at $Re_b = 5020$, 6962, 10 000, and 20 000. Here, we do not consider the observational uncertainty in $\langle u_\tau\rangle$ which could, for instance, be

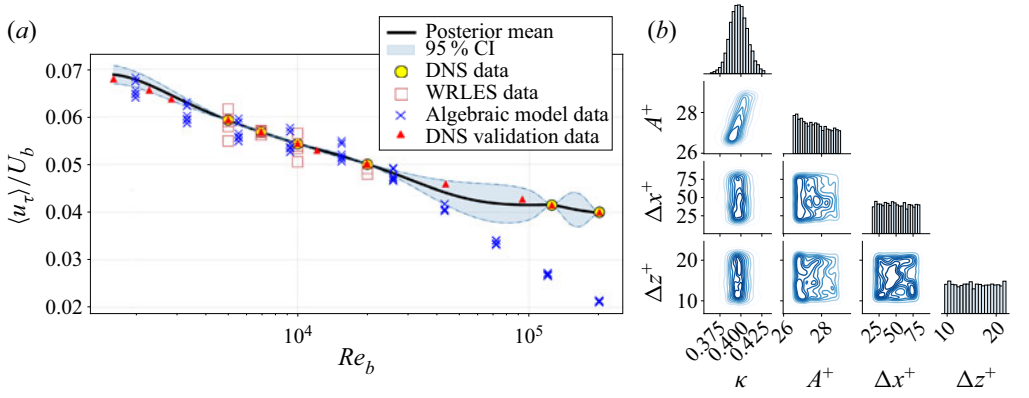


Figure 3. (*a*) Mean prediction of $\langle u_\tau \rangle / U_b$ and associated 95 % CI along with the training data and validation data from DNS of Iwamoto *et al.* (2002), Lee & Moser (2015) and Yamamoto & Tsuji (2018), (*b*) diagonal, posterior density of parameters κ, A^+, Δx^+ and Δz^+; off diagonal, contour lines of the joint posterior densities of these parameters. The value of the contour lines increases from the lightest to darkest colour.

due to finite time averaging in DNS and WRLES, but in general the HC-MFM could take such information into account. The reduced-order model (3.2) which is computationally cheap is run for 10 values of Re_b in range [2000, 200 200]. For each Re_b, 9 joint samples of (κ, A^+) are generated assuming $\kappa \sim \mathcal{U}[0.36, 0.43]$ and $A^+ \sim \mathcal{U}[26.5, 29]$ (note that κ is the von Kármán coefficient). For all the GPs in the MFM (2.6), the exponentiated quadratic covariance function (2.11) is used. The prior distribution of the hyperparameters are set as the following: $\lambda_f, \lambda_g \sim \mathcal{HC}(\alpha = 5)$, $\lambda_\delta \sim \mathcal{HC}(\alpha = 3)$, $\ell_{f_x}, \ell_{f_{t_3}}, \ell_{g_x}, \ell_{g_{t_2}}, \ell_{\delta_x} \sim \Gamma(\alpha = 1, \beta = 5)$ and the noise standard deviation $\sigma \sim \mathcal{HC}(\alpha = 5)$ (assumed to be the same for all fidelities).

Using the described training data in the HC-MFM (2.6) and running the MCMC chain for 7000 samples, after an initial 2000 samples discarded due to burn in, the model is constructed. This means extra MCMC samples are generated from which a sufficiently large number of initial samples is discarded to avoid any bias introduced by the initialization. According to figure 3(*a*), the predicted mean of $\langle u_\tau \rangle / U_b$ follows the trend of the DNS data. This approximately holds even at high Reynolds numbers, where there is a large systematic error in the algebraic model and no WRLES data are available. As expected, in this range due to scarcity of the DNS data, the uncertainty in the predictions is high. The plot in figure 3(*b*) shows the joint posterior distributions of the calibration parameters κ, A^+, Δx^+ and Δz^+ along with the histogram of each parameter. As mentioned above, the prior distributions of all of these parameters were assumed to be uniform and mutually independent. However, the resulting posterior densities for κ and A^+ are not uniform and the samples of these parameters are correlated. More interestingly, the peak of the posterior density of κ is close to the value of 0.4 that is assumed to be universal across various flows and Reynolds numbers within the turbulence community. On the other hand, the distribution of A^+ does not show a clear maximum, which may indicate that the range of admissible values in the prior could be extended. Note that indeed the value of $A^+ = 25$ is associated with the van-Driest wall damping commonly used for eddy-viscosity models. In contrast, the posteriors of Δx^+ and Δz^+ are found to be still close to the prior uniform distributions and no correlation between their samples is observed. This may be at least partially be due to the fact that the number of the WRLES data is limited as they are obtained only at 4 Reynolds numbers. Nevertheless, over this

range of the Reynolds number the posterior prediction of the QoI $\langle u_\tau \rangle$ is very accurate and has the lowest uncertainty, which seems to indicate that the significant dependence of the wall-shear stress on the WRLES resolution did not lead to a bias in the prediction by the HC-MFM. This may in fact be an important aspect for building future wall models.

3.3. *Polars for the NACA0015 airfoil*

In this section, the HC-MFM (2.6) is applied to a set of data for the lift and drag coefficients, C_L and C_D, respectively, of a wing with a NACA0015 airfoil profile at Reynolds number 1.6×10^6. The angle of attack (AoA) between the wing and the ambient flow is taken to be the design parameter x.

The data consist of the following sources with respective fidelities in brackets: wind-tunnel experiments by Bertagnolio (2008) (M_1), detached-eddy simulations (DES) ($M2$) and two-dimensional RANS (M_3) both by Gilling, Sørensen & Davidson (2009). The simulations of Gilling *et al.* (2009) were performed with a finite-volume code using a fourth-order central-difference scheme and a second-order accurate dual time-stepping algorithm to integrate the momentum equations. The PISO algorithm was used to enforce pressure–velocity coupling. In the study, the authors investigated the sensitivity of the DES results to the resolved turbulence intensity (TI) of the fluctuations imposed at the inlet boundary. The sensitivity was found to be particularly significant near the stall angle. Therefore, when constructing an MFM, the calibration parameter t_2 in fidelity M_2 is taken to be the TI.

For the covariance of the Gaussian processes $\hat{f}(x)$ and $\hat{g}(x, t_2)$ in HC-MFM (2.6), the exponentiated quadratic and Matern-5/2 kernel functions (2.11) and (2.12) are used, respectively. The following prior distributions are assumed for the hyperparameters: $\lambda_f, \lambda_g \sim \mathcal{HC}(\alpha = 5)$, $\ell_{f_x} \sim \Gamma(\alpha = 1, \beta = 5)$, $\ell_{g_x}, \ell_{g_{t_2}} \sim \Gamma(\alpha = 1, \beta = 3)$ and the noise standard deviation $\sigma \sim \mathcal{HC}(\alpha = 5)$ (assumed to be the same for all fidelities). To make the model capable of capturing large-scale separation, the stall AoA, x_{stall}, is included as a new calibration parameter in the MFM. Our suggestion is to consider a piecewise kernel function for the covariance of $\hat{\delta}(x)$ where x_{stall} is the merging point. If the kernel functions for the AoAs smaller and larger than x_{stall} are denoted by $k_{\delta_1}(\cdot)$ and $k_{\delta_2}(\cdot)$, respectively, then the covariance function for $\hat{\delta}(x)$ may be constructed as

$$\Sigma_{\delta_{ij}} = \lambda_{\delta_1}^2 \varphi(x_i) k_{\delta_1}(\bar{d}_{\delta_{ij}}) \varphi(x_j) + \lambda_{\delta_2}^2 \varphi(x_i) k_{\delta_2}(\bar{d}_{\delta_{ij}}) \varphi(x_j), \tag{3.4}$$

where $\bar{d}_{\delta_{ij}}$ is defined similar to those in (2.9) and (2.10) but only in x, and $\varphi(x)$ is a function to smoothly merge the two covariance functions. In particular, we use the logistic function:

$$\varphi(x) = [1 + \exp(-\alpha_{stall}(x - x_{stall}))]^{-1}, \tag{3.5}$$

where α_{stall} is a new hyperparameter. The kernel functions $k_{\delta_1}(\cdot)$ and $k_{\delta_2}(\cdot)$ are both modelled by the Matern-5/2 function (2.12). As the prior distributions, we assume $\lambda_{\delta_1} \sim \mathcal{HC}(\alpha = 3)$, $\lambda_{\delta_2} \sim \mathcal{HC}(\alpha = 5)$, $\ell_{\delta_1} \sim \Gamma(\alpha = 1, \beta = 5)$, $\ell_{\delta_2} \sim \Gamma(\alpha = 1, \beta = 0.5)$, $\alpha_{stall} \sim \mathcal{HC}(\alpha = 2)$ and $x_{stall} \sim \mathcal{N}(14, 0.2)$. The prior for the TI is also considered to be Gaussian: $TI \sim \mathcal{N}(0.5, 0.15)$. These Gaussian priors are selected based on our prior knowledge about the approximate stall AoA and the discussion about the influence of TI in Gilling *et al.* (2009).

The admissible range of $x = AoA$ is assumed to be $[0^\circ, 20^\circ]$ over which the experimental and RANS data (Bertagnolio 2008; Gilling *et al.* 2009) are available. The training HF data are taken to be a subset of size 7 of the experimental data of Bertagnolio (2008).

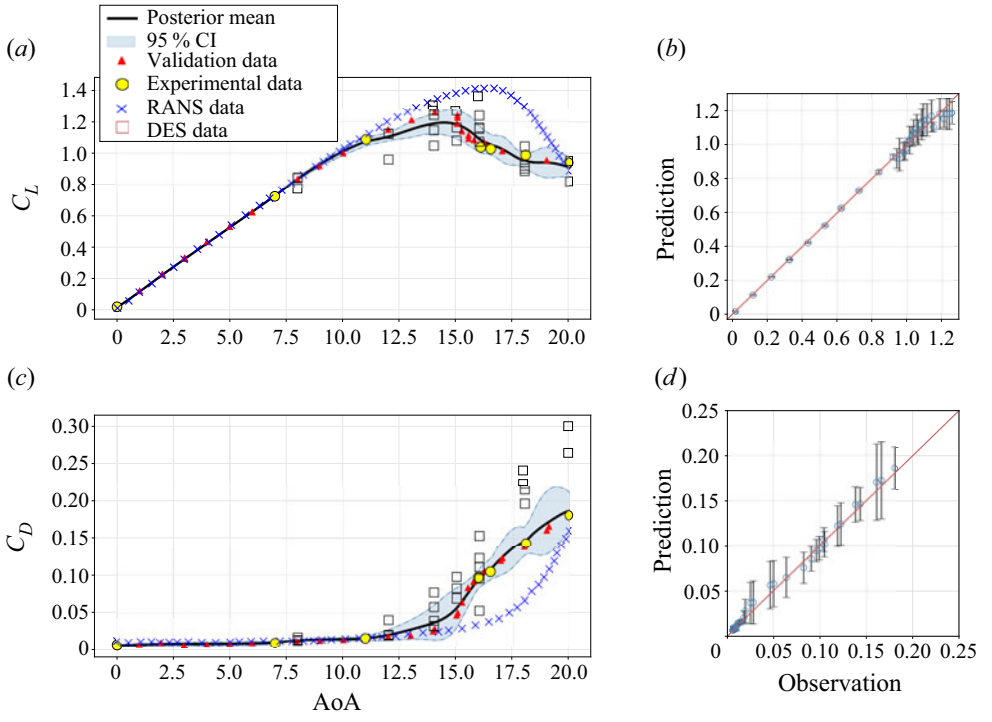


Figure 4. (*a*) Lift coefficient C_L and (*c*) drag coefficient C_D plotted against the AoA: the HC-MFM (2.6) is trained by the experimental data of Bertagnolio (2008) (yellow circles), as well as the DES (squares) and RANS (crosses) data by Gilling *et al.* (2009). The DES were performed with the resolved turbulence intensities $TI = 0\,\%, 0.1\,\%, 0.5\,\%, 1\,\%$ and $2\,\%$ at the inlet. The validation data (red triangles) are also taken from the experiments of Bertagnolio (2008). The mean prediction by the HC-MFM (2.6) is represented by the solid line along with associated 95 % CI (shaded area). Scatter plots of (*b*) C_L, (*d*) C_D predictions by the HC-MFM (vertical axis) against the validation data (horizontal axis). The red straight line is diagonal and provided to evaluate the accuracy of the HC-MFM predictions: if the hollow markers which represent the mean posterior prediction by the HC-MFM at the AoAs where validation data are available for, are close to the diagonal line, then they are more accurate. Each mean prediction has an error bar which represents the associated 95% CI.

The rest of the experimental data are used to validate the predictions of the MFM. For the purpose of examining the capability of the MFM in a more challenging situation, the training HF samples are explicitly selected to exclude the range of AoAs where the stall happens. The DES data of Gilling *et al.* (2009) are available at 7 $AoA \in [8^\circ, 19^\circ]$ and 5 different values of TI. Employing these training data in the HC-MFM (2.6) and drawing 10^4 MCMC samples after excluding an extra 5000 initial samples for burn in, the predictions for C_L and C_D shown in figure 4(*a*,*c*) are obtained. The expected value of the predictions has a trend similar to that of the experimental validation data of Bertagnolio (2008) and is not diverging towards either the physically invalid RANS data or scattered DES data at $AoA \gtrsim 10^\circ$. A more elaborate comparison is made through scatter plot of the MFM predictions against the validation data in figure 4(*b*,*d*). For both C_L and C_D, the agreement between the predicted mean values with the validation data at lower AoAs (before stall) is excellent and for most of the higher AoAs, even near and after the stall, is very good. Due to the scarcity of the HF training data and also systematic error in the RANS and DES data, the error bars at the predicted values can be relatively large, as more evident in the case of C_D in figure 4(*d*).

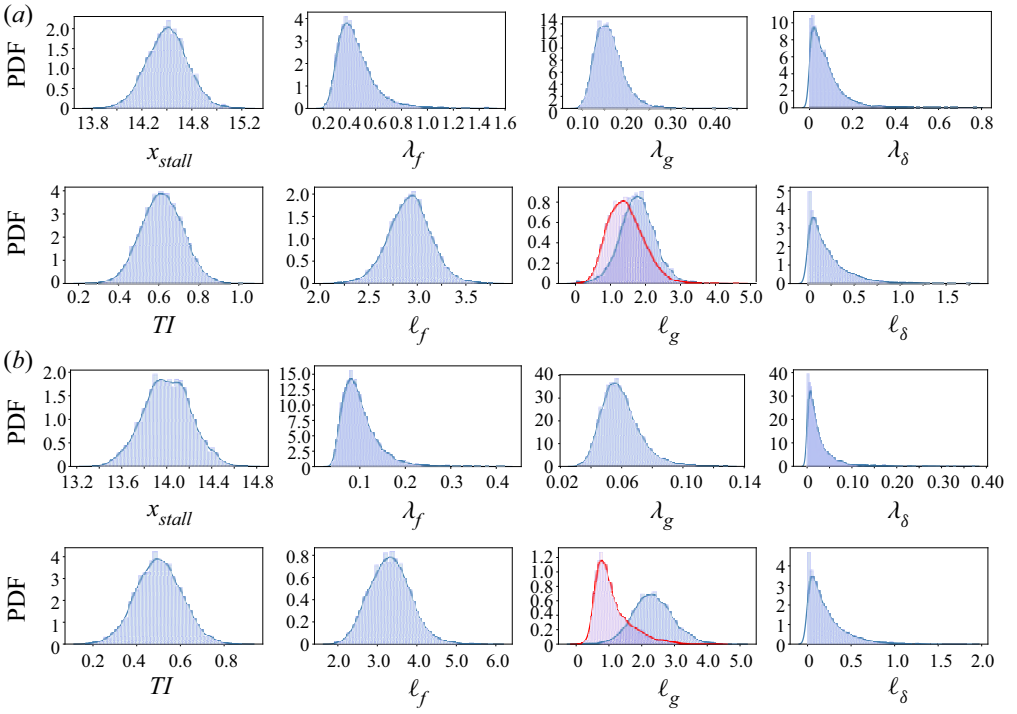


Figure 5. Posterior PDFs of the calibration parameters and hyperparameters of the GPs appearing in the HC-MFM (2.6) for (*a*) C_L and (*b*) C_D. Associated training data and predictions are shown in figure 4. In the plots of ℓ_g, the blue and red histograms correspond to the AoA and TI, respectively. Note that the values of TI are in percentage.

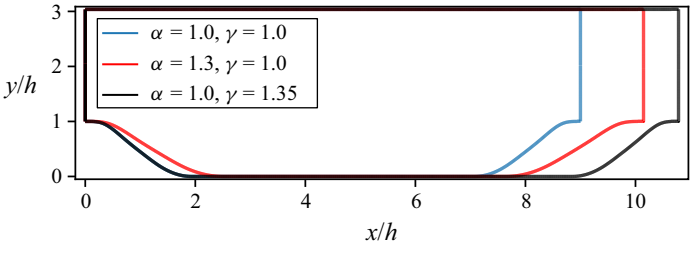


Figure 6. The geometry of the periodic hill simulations, illustrating the effect of parameters α and γ using three sets of values for them.

Figure 5 shows the posterior densities of different parameters appearing in the MFM constructed for C_L and C_D. As expected, the distribution of the kernels' hyperparameters varies between the two QoIs. But more importantly, the posterior distributions of x_{stall} and calibrated TI are also dependent on the QoI. This clearly shows the suitability of the present class of MFMs in which calibration of the parameters of different fidelities is performed as a part of constructing the MFM for a given QoI. The alternative strategy, which is common in practice (e.g. for co-kriging models without calibration), is to calibrate the LF models against the HF data of one of the QoIs and then run the calibrated LF model to make realizations of all QoIs. However, given the present results, this strategy seems clearly less efficient and leads to inferior quality of prediction.

As a general goal, an MFM constructs a surrogate for the QoIs in the space of the design/controlled parameters aiming for the surrogate outputs to be as close as possible to the HF data. In this regard, the MFMs facilitate applying different types of sample-based UQ techniques and optimization (see Smith 2013). In connection with the present example, consider a UQ forward problem to estimate the stochastic moments of C_L and C_D due to the variation of the AoA. For instance, assume $AoA \sim \mathcal{N}(15, 0.1)$ degrees. This results in the following estimations for the expectation and variance of C_L and C_D with the associated 95 % CIs: $\mathbb{E}_x[C_L] = 1.1775 \pm 0.1037$, $\mathbb{V}_x[C_L] = 2.6675 \times 10^{-5} \pm 4.5096 \times 10^{-5}$, $\mathbb{E}_x[C_D] = 5.6891 \times 10^{-2} \pm 2.5722 \times 10^{-2}$ and $\mathbb{V}_x[C_D] = 1.0618 \times 10^{-5} \pm 9.5994 \times 10^{-6}$. Note that, without the HC-MFM, and only based on the data of RANS or/and DES, such estimations would be at best inaccurate, but in general impossible to make.

3.4. *Effect of geometrical uncertainties in the periodic hill flow*

In this last flow case, we consider the turbulent flow over periodic hills with geometrical uncertainties at Reynolds number 5600, see the sketch in figure 6. The outline in blue corresponds to the configuration studied in several prior works (hereafter, baseline geometry), for example by Fröhlich *et al.* (2005) and more recently by Gao, Cheng & Samtaney (2020). The latter reference can be consulted for a good overview of other previous efforts. The shape of the hill is defined by six segments of third-order polynomials, see e.g. Xiao *et al.* (2020) for the exact definition. In that work, a parameterization of the geometry was introduced by scaling the length of the hill. Particularly, the authors performed a series of DNS for $L_x/h = 3.858\alpha + 5.142$, where L_x is the length of the geometry, h is the height of the hill and α is a parameter. The value $\alpha = 1$ corresponds to the baseline geometry. The corresponding DNS data set for several values of α has been made public on Github, which was extended in 2021 with additional data introducing a new parameter γ:

$$L_x/h = 3.858\alpha + 5.142\gamma. \tag{3.6}$$

The effect of α and γ on the geometry is illustrated in figure 6 with red and black curves, respectively. The purpose of the present example is to demonstrate how the HC-MFM can be used to economically assess the effect of uncertain parameters α and γ on the flow QoIs. To that end we combine the DNS data discussed above from Xiao *et al.* (2020) with data from RANS simulations performed by us using ANSYS Fluent (2019). The data from the latter are made publicly available (archive available via the following doi:10.6084/m9.figshare.21440418).

For the DNS we use the data for $\alpha \in \{0.5, 1.0, 1.5\}$ and $\gamma \in \{0.4166, 1.0, 1.5834\}$. The same values are also used for the RANS, complemented by two additional samples for both α and γ for which DNS results are not available. These are selected based on the Gauss–Legendre quadrature rule and are equal to $\alpha = \{0.702, 1.297\}$ and $\gamma = \{0.653, 1.347\}$. Therefore, there is a total of 5×5 samples over the space of α–γ corresponding to which RANS simulations are performed. For the uncertainty propagation and sensitivity analysis, see below, we assume α and γ to be independent, and $\alpha \sim \mathcal{U}[0.448, 1.552]$ and $\gamma \sim \mathcal{U}[0.356, 1.644]$.

The standard k–ω turbulence model is used in the RANS simulations, as defined in ANSYS Fluent (2019) based on the work of Wilcox (2006). The available low-Reynolds-number correction to the model was not used. The model depends on a number of parameters which are listed in ANSYS Fluent (2019, p. 61). It can be shown (see

Wilcox 2006, p. 134) that the parameters α_∞, β^*_∞, σ and β_i are coupled to the von Karmán coefficient κ through the following equation:

$$\alpha_\infty = \beta_i/\beta^*_\infty - \sigma\kappa^2/\sqrt{\beta^*_\infty}. \tag{3.7}$$

To illustrate the automatic calibration capability of the HC-MFM, we assume κ to be uncertain and perform simulations for five sample values of $\kappa \in \{0.348, 0.367, 0.4, 0.433, 0.452\}$, which follow the Gauss–Legendre quadrature rule. To prescribe the desired value of κ, we set $\alpha_\infty = 0.52$, $\sigma = 0.5$, $\beta_i = 0.0708$ (as suggested by Wilcox 2006, p. 135), and manipulate β^*_∞ according to (3.7). Note that the considered training samples include the standard choice $\beta^*_\infty = 0.09$, corresponding to $\kappa = 0.4$. Using five samples for each of α, γ and κ and using a tensor-product rule, a total of 125 RANS simulations were performed for this study.

For RANS simulations, quadrilateral cells were used to discretize the computational domains, with the total grid size ranging from $\approx 150 \times 10^3$ to $\approx 500 \times 10^3$ cells, depending on the domain size as defined by α and γ. Since the selected turbulence model requires accurate resolution of the boundary layer, the selection of the mesh size was guided by the discretization in the corresponding DNS case (Xiao *et al.* 2020). Specifically, we adopted the same number of cells in the streamwise and wall-normal directions as in the DNS, and applied size grading in the wall-normal direction to ensure low values of y^+. This means that in the streamwise direction the mesh may be unnecessarily fine for RANS, but since these simulations are two-dimensional and steady state, they are still negligibly cheap compared with the DNS.

Second-order numerical schemes were used to discretize the RANS equations in ANSYS Fluent (2019). Specifically, for the convective fluxes, a second-order upwind scheme was used, combined with linear interpolation for the diffusive fluxes. The coupled solver was used for pressure–velocity coupling, which did an excellent job at converging the simulations.

3.4.1. *Creating ground truth and verifying the model*

Using all nine DNS data sets available from Xiao *et al.* (2020), the response surface of the QoIs at the space of α–γ and associated PDF of the QoIs can be estimated. These will be used as the ground truth or reference to evaluate the performance of the HC-MFM in the following analyses. The interpolation from the DNS samples to an arbitrary mesh covering the whole admissible range of α and γ can be done using polynomial-based methods such as PCE (used here) or Lagrange interpolation, as well as GPR. Based on the data available from Xiao *et al.* (2020) and the performed RANS simulations, different QoIs can be considered. Hereafter, to demonstrate the power of the method, we take the normalized height of the separation bubble, H_{bubble}/h at the streamwise location $x/h = 2.5$ as the QoI. Alternatively other locations x/h as well as different flow quantities could be considered. The response surface of the QoI and associated PDF are illustrated in figure 7. Based on the pattern of the isolines, we can observe that the parameter α exhibits a stronger influence on the QoI than γ. This can be quantitatively confirmed via the values of the total Sobol indices (Sobol 2001) as reported in table 1. The resulting PDF has one peak showing the most probable observed value of H_{bubble}/h at $x/h = 2.5$ and a plateau approximately over $H_{bubble}/h = [0.46, 0.53]$.

The reference posterior distribution of κ as the RANS calibration parameter can now be inferred. To this end, the HC-MFM described in § 3.4.2 is constructed using nine DNS data sets of Xiao *et al.* (2020) and 125 RANS simulations. The prior distribution of κ is taken to be uniform over the range of [0.3, 0.5]. This non-informative prior distribution removes

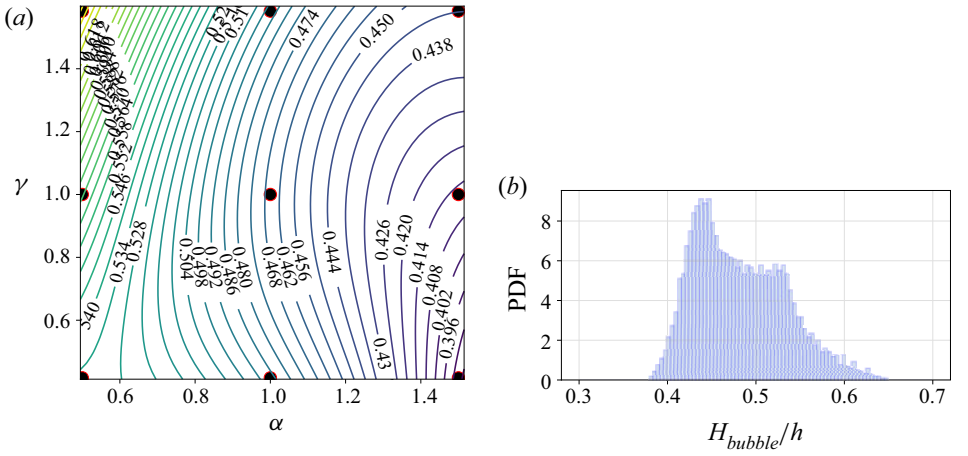


Figure 7. (*a*) Isolines of the response surface and (*b*) PDF of H_{bubble}/h at $x/h = 2.5$ due to the variation of α and γ using all of the nine DNS data of Xiao *et al.* (2020) (represented by the symbols in the left plot). These plots are considered as ground truth or reference for evaluating the performance of the MFM.

		Moments due to α, γ		Total Sobol Indices of R with respect to		
	Data set	$\mathbb{E}[R]$	$\mathbb{S}[R]$	α	γ	κ
	Reference	0.48104	0.05027	0.94551	0.08094	—
	LF	0.51999	0.01666	0.44275	0.02689	0.60623
Case-A	HF	0.50047	0.05045	0.81448	0.19144	—
	Multifidelity (MF)	0.49209	0.05238	0.88133	0.12680	—
Case-B	HF	0.46981	0.05198	0.82327	0.18237	—
	Multifidelity (MF)	0.48437	0.05215	0.91355	0.10097	—

Table 1. Estimated mean, standard deviation and total Sobol indices of the QoI $R = H_{bubble}/h$ at $x/h = 2.5$ due to the uncertainty in α and γ. For the LF (RANS) data the uncertainty and sensitivity with respect to parameter κ is also included. For the case-A and case-B data sets used for multifidelity modelling, see figure 9.

any bias towards any particular value in the distribution of κ. Through a Bayesian inference via an MCMC method, the sample posterior distribution of κ shown in figure 8(*a*) is obtained. Note that this calibration is in fact a pure UQ inverse problem, see e.g. Smith (2013), where all the RANS data are utilized to construct a surrogate for κ, and the DNS data are used as training data to infer the distribution of κ. The estimated mean and standard deviation of the posterior distribution of κ are 0.47087 and 0.05387, respectively. As compared with the standard value 0.4 being used in the literature, the estimated mean is somewhat larger. However, from a physical point of view, one would not expect an accurate value of κ for this type of flow due to the separated nature and the relatively low Reynolds number.

Another advantage of using all the available RANS and DNS data in the HC-MFM is that the implementation (algorithm and coding) of the HC-MFM can be verified. As shown in figure 8(*b*), the prediction of the HC-MFM constructed by combining all DNS and RANS data sets completely agree with the predictions of the single-fidelity model based on only the DNS data. Note that the predicted marginal PDF of the QoI in the HC-MFM is the same as the reference PDF in figure 7.

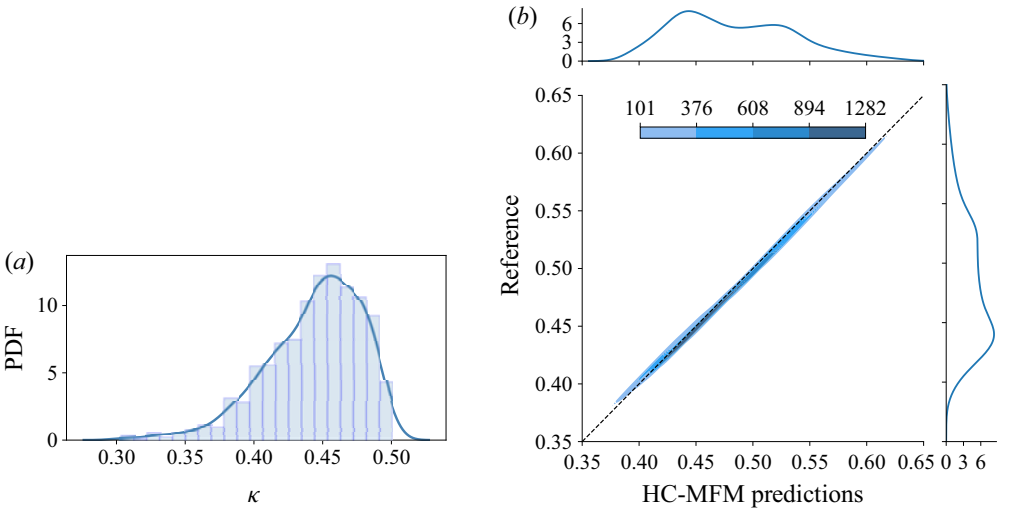


Figure 8. (*a*) The sample posterior PDF of κ and (*b*) sample joint PDF of H_{bubble}/h at $x/h = 2.5$ obtained from the HC-MFM using all nine DNS data sets of Xiao *et al.* (2020) along with the 125 RANS simulations performed in the present study. The RANS simulations are performed using five samples of κ equal to 0.348, 0.367, 0.4, 0.433 and 0.452. The marginal PDFs on the top and right axes are found using the kernel density estimation method.

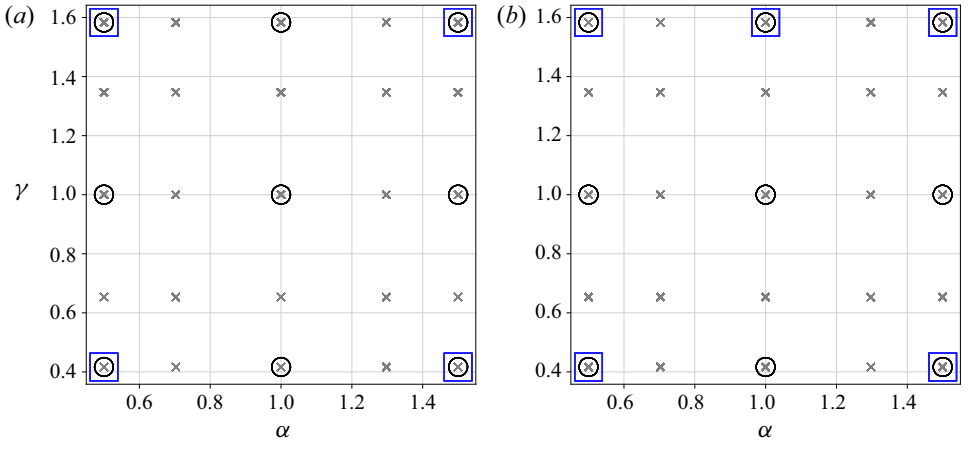


Figure 9. Schematic representation of the samples from α and γ corresponding to the LF, $\times$, HF, $\square$ and all available DNS data from Xiao *et al.* (2020), $\circ$. In the text, (*a*) and (*b*) are referred to as case-A and case-B, respectively. Note that, for both cases, there are five samples for κ associated with each of the LF samples represented here.

3.4.2. *Application of the HC-MFM*

Adopting the general notation of § 2, the HC-MFM for the present example can be written as

$$\left.\begin{aligned} y_{M_1}(\boldsymbol{x}_i) &= \hat{f}(\boldsymbol{x}_i, \theta_{2_i}) + \hat{\delta}(\boldsymbol{x}_i) + \varepsilon_{1_i} \quad i = 1, 2, \ldots, n_1 \\ y_{M_2}(\boldsymbol{x}_i) &= \hat{f}(\boldsymbol{x}_i, t_{2_i}) + \varepsilon_{2_i} \quad\quad\quad\quad i = 1 + n_1, 2 + n_1, \ldots, n_2 + n_1 \end{aligned}\right\}, \tag{3.8}$$

where M_1 and M_2 denote DNS and RANS, respectively, the design parameters are $\boldsymbol{x}_i = (\alpha_i, \gamma_i)$ and, t_2 and θ_2 refer to the simulated and calibrated instances of κ, respectively. The

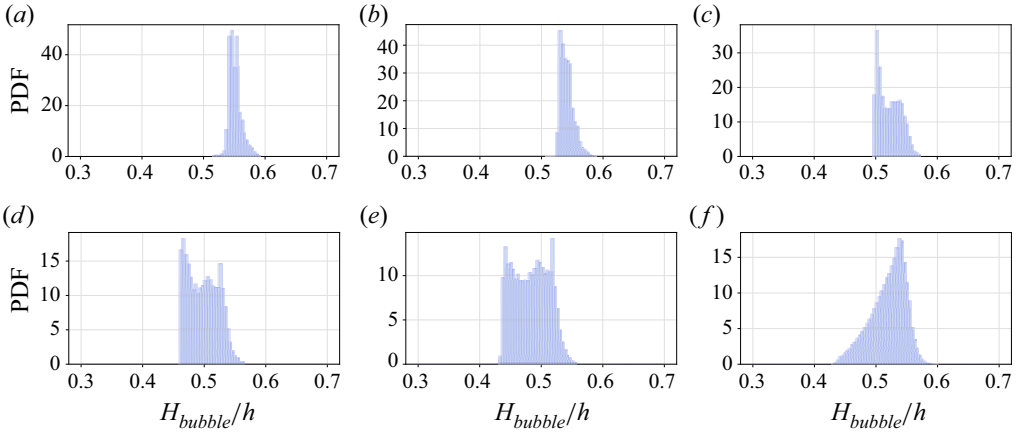


Figure 10. (*a–e*) The PDF of H_{bubble}/h at $x/h = 2.5$ due to the variation of α and γ using the RANS data simulated with κ equal to 0.348, 0.367, 0.400, 0.433, 0.452, respectively. Note that 5×5 samples are taken from the α–γ space at each of these constant-κ simulations. The PDF in (*f*) is obtained using all the $5 \times 5 \times 5$ samples from α, γ and κ.

kernel of $\hat{f}(\boldsymbol{x}, t_2)$ is taken to be the exponentiated quadratic function (2.11), while for $\hat{\delta}(\boldsymbol{x})$, the Matern-5/2 function (2.12) is employed. For the hyperparameters, the following prior distributions are considered: $\kappa \sim \mathcal{U}[0.3, 0.5]$, $\lambda_f \sim \mathcal{HC}(\alpha = 5)$, $\ell_{f_x}, \ell_{f_{t_2}} \sim \Gamma(\alpha = 1, \beta = 5)$, $\lambda_\delta \sim \mathcal{HC}(\alpha = 1)$ and $\ell_{\delta_x} \sim \Gamma(\alpha = 1, \beta = 1)$. In this example, the uncertainty in the DNS and RANS data is neglected. Despite this, when implementing the model in PyMC3 (Salvatier *et al.* 2016), the prior of the Gaussian noise standard deviation is set to be $\sigma \sim \mathcal{HC}(\alpha = 5)$ (same for both fidelities). But, as expected, the mean and standard deviation of the posterior distribution of σ are obtained to be approximately zero.

The hyperparameters will be inferred from the combined set of n_1 DNS and n_2 RANS data. In the analyses to follow, we use all the RANS simulations as LF data, therefore $n_2 = 125$. Two subsets of the DNS data of Xiao *et al.* (2020) with sizes $n_1 = 4$ and 5 are taken to be the HF data. Combining these two HF data sets with the LF data, case-A and case-B data sets are obtained for multifidelity modelling. Figure 9 represents the samples of these two cases in the space of α–γ parameters.

Before constructing the HC-MFM, it is important to look at the LF data. In figure 10, the PDF of the QoI due to the variation of α and γ for different training samples of κ is represented. The two expected yet important observations are that the PDF of the QoI is significantly influenced by the value of κ used in the RANS simulations, and the fact that the PDFs are much different from the ground truth PDF shown in figure 7. The larger influence of κ compared with α and γ is also reflected in the associated Sobol indices (Sobol 2001), as reported in table 1. Note that the PDF of the QoI considering the simultaneous variation of α, γ, and κ is shown in figure 10(*f*).

The response surface and PDF of the QoI obtained from only the HF data of case-A and case-B are illustrated in figure 11. For case-A with $n_1 = 4$ HF data, the PDF of the QoI has a plateau which makes it clearly different from the reference PDF in figure 7. By adding only one more DNS data point and obtaining case-B, the response surface becomes more similar to the reference, however, the associated PDF is still single mode. The improved predictions through the application of the HC-MFM are shown in figure 12. Compared with the HF-data in figure 11, the PDFs of the QoI clearly exhibit a second

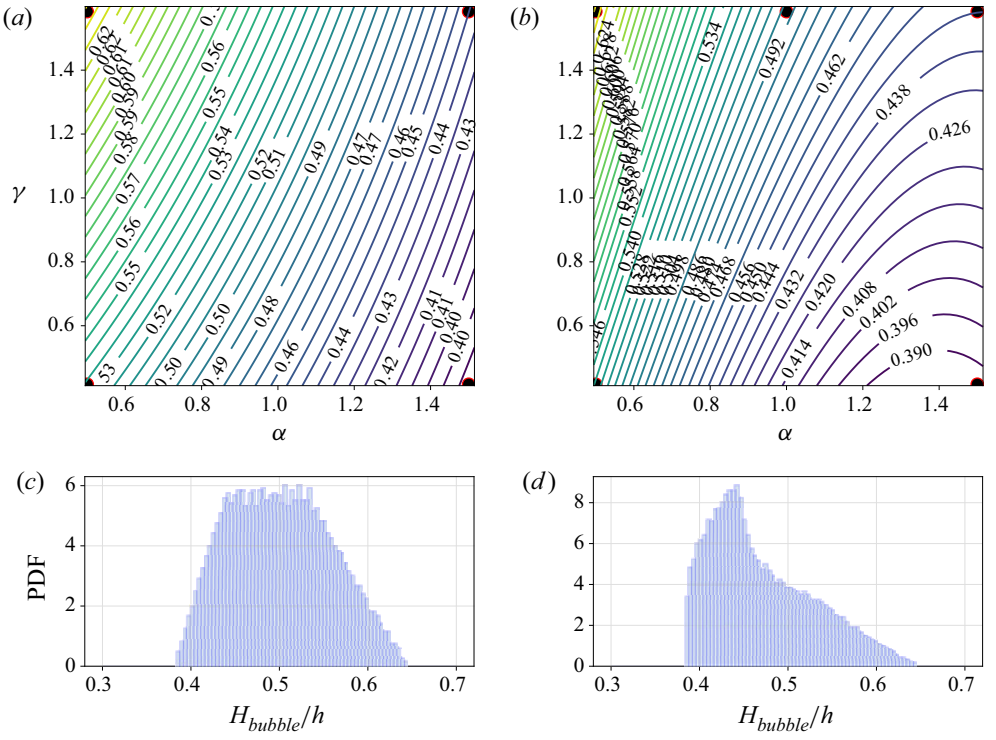


Figure 11. (*a*,*b*) Isolines of the response surface and (*c*,*d*) PDF of H_{bubble}/h at $x/h = 2.5$ due to the variation of α and γ using the HF data of (*a*,*c*) case-A and (*b*,*d*) case-B. The data are taken from the DNS of Xiao *et al.* (2020) and are specified by dots in (*a*,*b*).

peak for the values between 0.5 and 0.6. This peak has been introduced by adding the LF data, see figure 10(*f*), and this is, in fact, the task for the MFM to adjust the involved hyperparameters such that the fusion of the data at the two fidelities leads to a PDF similar to the reference. Clearly, for case-B with only five DNS data samples included, the PDF and response surface of the QoI are very close to the ground truth in figure 7 (nine DNS). This can also be confirmed by plotting the associated HC-MFM predictions against the reference data at all test points in the α–γ plane, see figure 13. The joint PDF of these two sets of data for both considered cases is narrow, specifically for case-B, and hence implies a low point-to-point deviation of the predictions from the reference values. It is also interesting to look at the posterior distribution of the RANS calibration parameter κ. It is not surprising that the resulting PDFs from the multifidelity data sets case-A and case-B are different in spite of having the same uniform prior distribution. Two important observations here are the following: first, in contrast to the previous examples in the present study, even with a small number of HF-data, i.e. case-A, a significantly informative PDF for the LF calibration parameter is obtained. Second, for case-B, the posterior distribution of κ is very similar to the reference case where nine DNS data sets are used, see figure 8. Thus, depending on the case, the HC-MFM is capable of calibrating the model parameters in a fairly accurate way at the same time of constructing an accurate predictive model. This somewhat challenges the conclusion which could be drawn from the previous examples in the present study and also by Goh *et al.* (2013), where the priority of the HC-MFM is found to make accurate predictions for the QoI rather than providing accurate calibration of the fidelity-specific parameters.

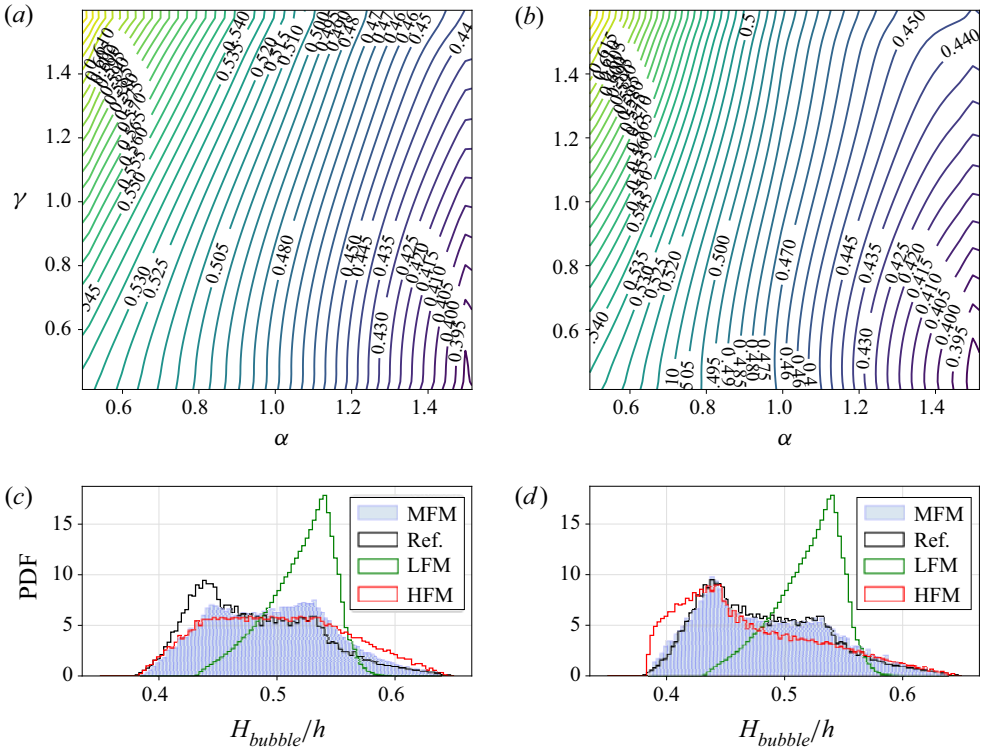


Figure 12. (*a*,*b*) Isolines of the response surface and (*c*,*d*) PDF of H_{bubble}/h at $x/h = 2.5$ due to the variation of α and γ obtained from the HC-MFM with the data of (*a*,*c*) case-A and (*b*,*d*) case-B. In (*c*,*d*), the PDF resulting from the HC-MFM is compared with the PDFs of the ground truth (see figure 7), LF data (LFM, figure 10) and HF data (HFM, figure 11).

To conclude the periodic hill example, we can quantify stochastic moments of the QoI as well as the Sobol sensitivity indices (Sobol 2001) due to the uncertainty in α and γ. These UQ measures are integral quantities over the admissible range of the parameters and can be computed using the reference data (all available DNS), LF, HF and MF data sets. Note that for the LF data, the uncertainty in κ is also taken into account. Noting the parameters α, γ and κ are uniformly distributed and independent from each other, the results summarized in table 1 are obtained using the generalized polynomial chaos expansion (Xiu & Karniadakis 2002) for the reference and LF data sets, and the Monte Carlo method for the multifidelity cases. All the UQ analyses have been performed using UQit (Rezaeiravesh, Vinuesa & Schlatter 2021). In general, for all cases but the pure LF data, the prediction of the mean and standard deviation of the QoI are close to the reference. For the total Sobol indices, the closest estimates to the reference values is obtained from the HC-MFM applied to case-B, and on the second rank, case-A. Noting the improvement of the Sobol indices accuracy in each of the multifidelity cases compared with the estimates from the associated HF data, the effectiveness of the HC-MFM is once again confirmed. This is an important outcome considering the forward UQ problems and global sensitivity analyses are of most relevance in CFD applications.

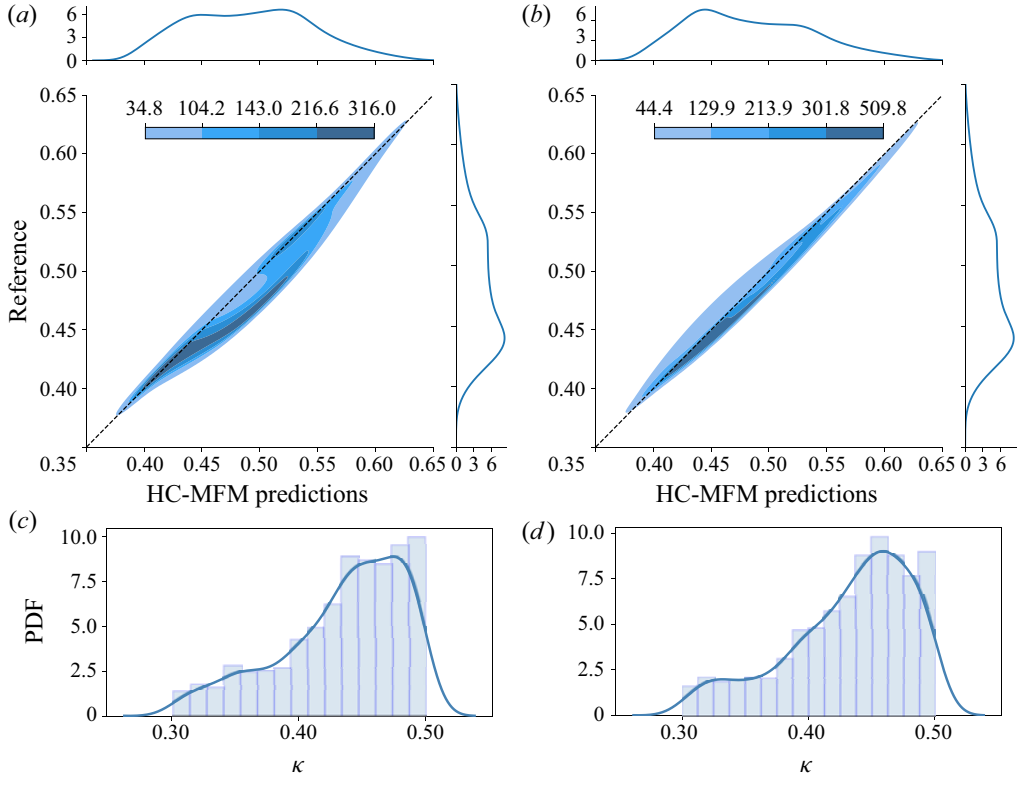


Figure 13. (*a*,*b*) Joint and marginal PDFs of H_{bubble}/h at $x/h = 2.5$, and (*c*,*d*) associated sample posterior distribution of κ for (*a*,*c*) case-A and (*b*,*d*) case-B data sets. In (*a*,*b*), the contours belong to the joint PDF with associated values specified in the colour bar.

3.5. *Keeping the RANS parameter fixed*

In all the examples in the present study, the fidelity-specific calibration parameters are involved in the multifidelity modelling and posterior distributions for them are learned during the construction of the MFM. But, the methodology behind the HC-MFM described in § 2 is general and flexible to be directly applicable to the cases where the fidelity-specific parameters are kept fixed. To demonstrate this, let us apply the HC-MFM to the example of the periodic hill and use the RANS data at constant values of κ, the RANS modelling parameter in (3.7). We use the case-B data sets shown in figure 9 which means having 5 and 25 samples for the DNS and RANS simulations, respectively, in the α–γ space. The validation of the PDF of the QoI, H_{bubble}/h at $x/h = 2.5$, of the HC-MFM for two values of κ is represented in figure 14. Adopting $\kappa = 0.433$ ($\beta^*_\infty = 0.084$) shows a clear improvement in the predictions compared with the standard value $\kappa = 0.4$ ($\beta^*_\infty = 0.09$). This observation is consistent with the posterior PDF of κ in figure 13, where the mode of the distribution is higher than 0.4. From this test, not only the validity of the HC-MFM for fixed values of the fidelity-specific parameters is confirmed, but also the fact that such accuracy-controlling parameters should be actively part of the data generation and hence the construction of HC-MFM is emphasized.

Another important point is that the good predictive accuracy in figure 14 is obtained despite the poor correlation between the DNS and RANS data. This is because of accurate construction of the model discrepancy term in (3.8) and also accurate estimation of

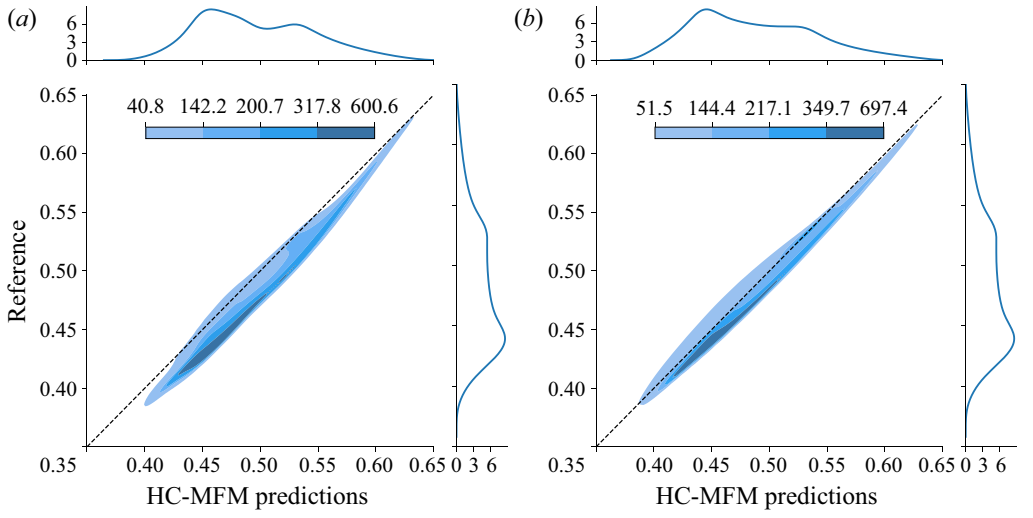


Figure 14. Joint and marginal PDFs of H_{bubble}/h at $x/h = 2.5$ for case-B data sets using fixed values of κ equal to (a) 0.4 and (b) 0.433. Note that the PDF of the QoI due to the variation of α and γ corresponding to these κ values is plotted in figure 10(c,d), respectively.

various hyperparameters. It is also noteworthy that the present example is comparable to what is performed by Voet *et al.* (2021) using 7 DNS and 30 RANS data sets using different multifidelity modelling approaches while fixing the value of the coefficients in the RANS closure model. However, a direct comparison between the two studies is not possible since, in the study by Voet *et al.* (2021), the plots of the MFM predictions vs reference values of the QoI as in figure 14 were not provided.

3.6. *Impact of replacing the MCMC by a point-estimator*

In all the examples presented in this study, the HC-MFM is constructed using the MCMC method to draw samples from the posterior distribution of the hyperparameters and calibration parameters, i.e. $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, respectively, in the Bayes formula (2.13). Similarly, to predict the sample distribution of the QoI, direct samples from these posterior distributions are used in the HC-MFM. As an alternative to these sample-based methods within the Bayesian framework, point estimators such as maximum *a posteriori* probability (MAP) and maximum likelihood estimators (MLE) can be adopted. The point-estimated values are considered to be the representatives of the corresponding distribution. Note that the use of the uniform priors in (2.13) makes the MAP estimations identical to the MLE. Our investigations showed that using point estimators instead of the MCMC method, could deteriorate the accuracy of the HC-MFM predictions, independent from how the LF and HF data are combined.

For instance, according to figure 15, the PDF of the QoIs predicted by the HC-MFM using the MAP estimator is significantly worse than what is given by the MCMC method as shown in figure 13. This is an important message of the present study noting that all previous multifidelity studies in the literature relevant to the fluid flows have been based on using the point estimators, see e.g. Voet *et al.* (2021) where the MLE is adopted.

The reason for this observation is that, to obtain the best fit for the GP-based models, here the HC-MFM, for a given set of data, the global optimum of the parameters appearing in the model should be found. This, in general, is not an easy task considering the

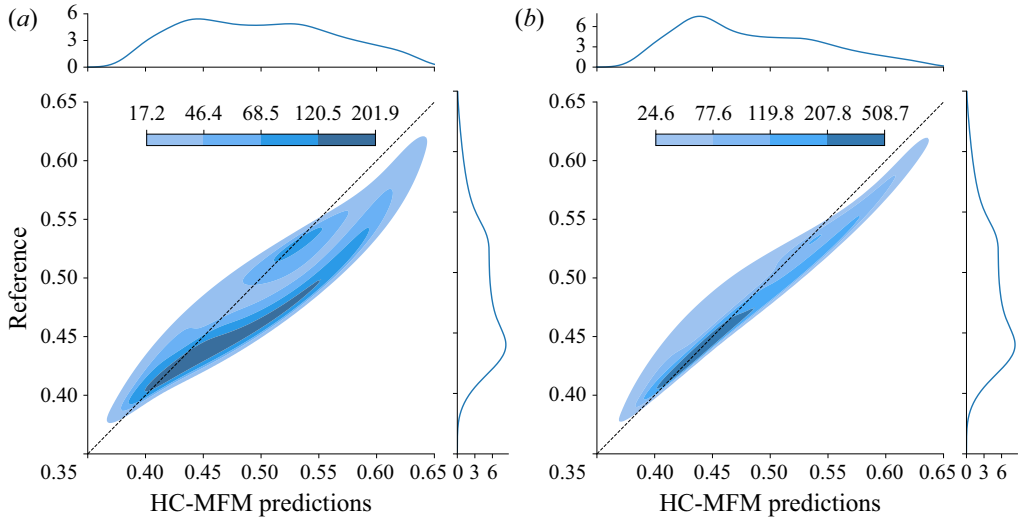


Figure 15. Joint and marginal PDFs of H_{bubble}/h at $x/h = 2.5$ for (a) case-A and (b) case-B data sets. Here, a MAP estimator is used to construct the HC-MFM, in contrast to figure 13 and the rest of the examples in the present study which are obtained using an MCMC method.

optimization problem can be non-convex with many local optima (Rasmussen & Williams 2005; Gramacy 2020). Depending on the problem, the point estimators such as MAP and MLE may or may not be successful in finding the global optimum. In contrast, a thorough exploration of the admissible space of the parameters via MCMC samples can rectify the issue. More concrete examples and discussion in this regard can be pursued in an extension of the present study.

4. Summary and conclusions

The Bayesian hierarchical MFM with automatic calibration (HC-MFM) developed by Goh *et al.* (2013) is adapted to several examples relevant to wall-bounded turbulent flows. The HC-MFM is general, accurate, applicable to an arbitrary number of fidelity levels and well suited to simulations of turbulent flows since as a part of the MFM construction the fidelity-specific parameters can be automatically calibrated using training data of higher fidelities. This is an important feature noting that in all approaches for simulation of turbulence, different numerical and modelling uncertain parameters can influence the accuracy of the QoIs. Because we used Gaussian processes, the predictions made by the HC-MFM are accompanied by CIs. Moreover, it is possible to incorporate the observational uncertainties in the data at all fidelity levels, and hence perform various UQ analyses for combinations of different types of uncertainties, see Rezaeiravesh *et al.* (2022).

Based on the examples, the following main conclusions can be made. (i) For a fixed number of HF training data, the HC-MFM prioritizes the prediction of QoIs so that they become as close as possible to the HF validation data, while the posterior distributions of the calibration parameters are found to be accurate only if sufficiently many LF training data are provided. A similar conclusion was drawn by Goh *et al.* (2013) by systematically increasing the amount of both HF and LF data. For the periodic hill subject to geometrical uncertainties, § 3.4.2, fixing the number of the LF data and considering two sets of HF data, the posterior distribution of the RANS (LF) parameter was found to be close to

what would be obtained by using all the available HF data. This, again, confirms the importance of providing sufficient LF training data through well exploring the space of the design and calibration parameters. (ii) When there are more than one QoI, the posterior distribution of the calibration parameters may depend on the QoI, see § 3.3. Therefore, the calibration parameters are more numerical than physical and hence, predictions by the HC-MFM for a QoI can be more accurate than the case of *a priori* calibrating the LF models against HF data of another QoI (an example is calibrating a RANS closure model by the HF data of the lift coefficient, and then using the calibrated model in a simulation aiming for the drag coefficient with optimal accuracy). (iii) As show in § 3.6, the method for estimating the hyperparameters and parameters in the HC-MFM can significantly affect the resulting predictive accuracy. In fact, the MCMC sampling method is shown to result in more accurate predictions compared with a point estimator like MAP. This important point is usually overlooked in most of, if not all, the previous studies regarding the multifidelity modelling in CFD. (iv) If the fidelity-specific calibration parameters are kept fixed, the HC-MFM is still applicable without any need to modifying its general formulation. Obviously, the predictive accuracy of the model will depend on the validity of the value chosen for such parameters when generating the training data. The success of the HC-MFM relies on the accurate modelling of the discrepancy terms between different fidelities, and also the use of the MCMC methods to find optimal values for underlying hyperparameters through exploration of the parameter space.

A user of the HC-MFM should be aware of the fact that the model has no intrinsic knowledge of the physics and what parameter/hyperparameter values are reasonable. Therefore, an assessment of the resulting predictions by the model based on relevant physical constraints is necessary. Furthermore, the choice of kernel functions and priors for the kernel hyperparameters may affect the predictions of the MFM, as discussed in Appendix A. However, the benefit of the Bayesian approach is that any prior knowledge can already be put to use during the set-up of the model, i.e. via the selection of appropriate prior distributions for the parameters. These factors, together with the selection of the training samples, may impact the robustness of the HC-MFM. Therefore, the method should be assessed in further applications in the field of turbulence, with varying complexity and prior physical knowledge.

The present study may be extended in several directions. For instance, in addition to the scalar QoIs, spatio–temporal fields can be considered in the HC-MFM, making it possible to predict full flow fields by combining different fidelities. Such an approach may be particularly interesting as an alternative to the more black-box machine-learning tools when it comes to super-resolution and related methods. Another potential extension is in the combination of the HC-MFM with a Bayesian optimization for CFD applications and turbulent flow problems (see Morita *et al.* 2022). In this case, the surrogate for the optimizer is based on the MFM, and is thus potentially cheaper to evaluate during the optimization process. In particular, applications in flow control for turbulence, where the main computational time lies in the evaluation of the objective function (i.e. the CFD solver), may greatly benefit from a well-calibrated MFM. Another development can be towards integrating adaptive sampling methods with the HC-MFM, where the decision about a new sample is made through maximizing the predictive accuracy (or minimizing uncertainty) of the model and minimizing the overall computational cost (by choosing a suitable fidelity to sample from).

Acknowledgements. The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC), partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

Funding. This work has been supported by the EXCELLERAT project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 823691. Additional funding was provided by the Knut and Alice Wallenberg Foundation (KAW).

Declaration of interests. The authors report no conflict of interest.

Data availability statement. The data from periodic hill RANS simulations are made publicly available on Figshare via the following doi:10.6084/m9.figshare.21440418.

Author ORCIDs.
S. Rezaeiravesh https://orcid.org/0000-0002-9610-9910;
T. Mukha https://orcid.org/0000-0002-2195-8408;
P. Schlatter https://orcid.org/0000-0001-9627-5903.

Author contributions. S.R.: conceptualization, methodology, software, formal analysis, writing – original draft. T.M.: designing and performing RANS simulations of the periodic hill case, writing – original draft, writing – review and editing. P.S.: conceptualization, investigation, funding acquisition, writing – review and editing.

Appendix A. Remarks on the techniques used

The purpose of this appendix is to provide a brief overview on some aspects of the Gaussian processes, HC-MFM and Bayesian inference to the extent relevant to the context of the present paper. For further details, a reader is referred to the relevant resources including those cited here.

Further notes on the GPR: according to Rasmussen & Williams (2005), a GP is defined as 'a collection of random variables, any finite number of which have a joint Gaussian distribution'. Therefore, $\hat{f}(\boldsymbol{x})$ in (2.1) is a multivariate Gaussian distribution for any joint sample taken from $\boldsymbol{x}$. The multivariate Gaussian distribution identity remains unchanged, either we define a prior for $\hat{f}(\boldsymbol{x})$ (before seeing the data), or after the posterior of $\hat{f}(\boldsymbol{x})$ is inferred from data. When the prior of $\hat{f}(\boldsymbol{x})$ is defined, it means we define a structure for $m(\boldsymbol{x})$ and $k(\boldsymbol{x}, \boldsymbol{x}')$. Within such structures, there are various hyperparameters for which we assume prior distributions. For the training data comprising observations $\boldsymbol{Y} = \{y_i\}_{i=1}^{n}$ at samples $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{n}$, the posterior predictive distribution of $\hat{f}(\boldsymbol{x})$ at the test samples $\boldsymbol{X}^* = \{\boldsymbol{x}_i^*\}_{i=1}^{n^*}$ is multivariate Gaussian with the mean and variance given by the following expressions, respectively (Rasmussen & Williams 2005):

$$m(\boldsymbol{Y}^*|\boldsymbol{X}^*, \boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X})(\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{K}_N)^{-1}\boldsymbol{Y}^T, \tag{A1}$$

$$v(\boldsymbol{Y}^*|\boldsymbol{X}^*, \boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X}^*) - \boldsymbol{K}(\boldsymbol{X}^*, \boldsymbol{X})(\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{K}_N)^{-1}\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}^*). \tag{A2}$$

Here, without loss of generality, $m(\boldsymbol{x})$ is taken to be zero, $\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}')$ is a $n \times n'$ matrix where $[\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}')]_{ij} = k(\boldsymbol{x}^{(i)}, \boldsymbol{x}'^{(j)})$ and $\boldsymbol{K}_N$ represents the covariance matrix of the noise. Clearly, in the absence of $m(\boldsymbol{x})$, it is the kernel function and the posterior of its hyperparameters which determine the mean and variance of the posterior predictive distribution of $\hat{f}(\boldsymbol{X}^*)$.

Choice of the kernels: in the absence of $m(\boldsymbol{x})$, the most influential factor to determine the accuracy of the GP predictions with respect to a given data set is the choice of the kernel function. This is evident from the (A1) and (A2), and the fact that $k(\boldsymbol{x}, \boldsymbol{x}')$ specifies the degree of similarity or covariance between $\hat{f}(\boldsymbol{x})$ and $\hat{f}(\boldsymbol{x}')$. In each problem, the distribution of the available data in the space of the inputs, or our insights into how the predictions should eventually be in the space of the parameters, can act as directives to select or design kernel functions. For instance, seeing/expecting periodicity in the training data

over the space of the inputs, necessitates using periodic kernels. Another example is the case studied in § 3.3, where we knew there should have been a sudden change in the C_L and C_D curves due to the stall, and because of that we broke the kernel function into two parts introducing the new parameter x_{stall} and let the HC-MFM learn it along with other parameters/hyperparameters.

If we assume that the function $\hat{f}(\boldsymbol{x})$ should be smooth (infinitely differentiable), then the quadratic exponentiated kernel can be chosen. To relax the smoothness assumption, the family of the Matern kernels can be considered which have a parameter ν controlling the degree of smoothness of the learned function, see e.g. Rasmussen & Williams (2005) and Matern (1986) (if $\nu \to \infty$, then the exponentiated quadratic is recovered). In the examples considered in the manuscript, as the baseline, we have considered the exponentiated quadratic kernel. For the airfoil and periodic hill examples, we observed (through experimentation) that considering the Matern-5/2 for the model discrepancy GPs $\hat{\delta}(\boldsymbol{x})$ and the additive kernels in the airfoil example (see (3.4)) results in slightly more accurate predictions by the HC-MFM compared with the case of using the exponentiated quadratic kernel. Although, we should emphasize that the degree of improvement was not really significant. In the end, it is recalled that there are studies like Duvenaud *et al.* (2013); Lloyd *et al.* (2014) where algorithms for automatic selection of the most suitable kernel functions for a given problem have been proposed.

Choice of the prior for the (hyper)parameters: any function selected for the $m(\boldsymbol{x})$ and $k(\boldsymbol{x}, \boldsymbol{x}')$ in the GPs in (2.1) and (2.6) relies on a set of hyperparameters, denoted by $\boldsymbol{\beta}$ in the text. In addition to these, we have $\boldsymbol{\theta}$, the fidelity-specific parameters in the HC-MFM. As a general rule, the prior for any parameter should be chosen such that it only allows for its associated admissible values. Moreover, for any choice for the priors, we should examine the validity of the resulting posteriors (Gelman *et al.* 2013). Given these, the particular choice made for the priors of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be justified as follows.

The uniform priors are non-informative, therefore, they are chosen in the examples in § 3, when the purpose was making the construction of an accurate HC-MFM more challenging, i.e. with providing no prior information about the distribution of the fidelity-specific parameters such as θ in the toy-problem example (§ 3.1), Δx^+, Δz^+, κ and A^+ in the channel flow example (§ 3.2) and κ in the periodic hill example (§ 3.4). However, we still define the admissible range of a parameter via specifying the bounds of the associated uniform distribution. In general, when there are few training data available, the choice of the uniform priors may have a significant impact on the posteriors, see Gelman *et al.* (2013). Indeed, we clearly observed this, for instance, in the toy-problem example (§ 3.1), where by increasing the number of LF samples, the posterior of θ changed significantly and became farther from a uniform distribution. Only on one occasion, in the airfoil example (§ 3.3), did we use the Gaussian prior for x_{stall} and TI as we had a good prior knowledge (from aerodynamics) about the range of the AoAs over which the stall would happen, and the order of the turbulence intensity in the experiments by Bertagnolio (2008) and the discussion in Gilling *et al.* (2009). For the noise standard deviation as well as the scaling factor of the kernels (such as λ in (2.8)), we adopt the half-Cauchy distribution as it only admits non-negative values and has a good performance for the hierarchical models (Gelman 2006; Gelman *et al.* 2013). Another type of prior considered in the manuscript is the gamma distribution for the length scale of the GPs' covariance kernels. This was based on the literature, see e.g. Rasmussen & Williams (2005) and the fact that the length scales needed to be non-negative.

Extrapolation by HC-MFM: in the examples considered in § 3, the prediction made by the HC-MFM for a QoI in the space of the design parameters $\boldsymbol{x}$ was within the range

Distribution	PDF	Support in x
Uniform ($\mathcal{U}$)	$\rho(x;\alpha,\beta)=(\beta-\alpha)^{-1}$	$[\alpha,\beta]$
Gaussian ($\mathcal{N}$)	$\rho(x;\alpha,\beta)=(2\pi\beta^2)^{-1/2}\exp(-\frac{1}{2}(\frac{x-\alpha}{\beta})^2)$	$(-\infty,\infty)$
Half-Cauchy ($\mathcal{HC}$)	$\rho(x;\alpha)=2[\pi\alpha(1+(\frac{x}{\alpha})^2)]^{-1}$	$[0,\infty)$
Gamma (Γ)	$\rho(x;\alpha,\beta)=\beta^{\alpha}x^{\alpha-1}\exp(-\beta x)/\Gamma(\alpha)$	$(0,\infty)$

Table 2. The PDF and associated support of the standard distributions used in § 3.

covered by the training data at various fidelities. This was an intentional choice, given the range of availability of HF data for the validation of the HC-MFM's predictions and the fact that in §§ 3.3 and 3.4, propagation of uncertainty from $\boldsymbol{x}$ and sensitivity analysis were targeted. A valid question is about the performance of the HC-MFM in an extrapolation mode, i.e. making prediction for the QoI in $\boldsymbol{x}$ beyond the range covered by the training data. Note that, for the data at different fidelities, the covered range in $\boldsymbol{x}$ may be different, which makes the interpretation of extrapolation slightly non-trivial. In any case, the HC-MFM relies on the GPs and therefore its performance when used for extrapolation can be case dependent. This means that if accurate predictions by the model are intended, the choice of the mean and covariance kernels in the GPs should be made in the way that allows for such purpose (Rasmussen & Williams 2005; Gramacy 2020). The particular optimal choice is, however, problem-dependent and can be the subject of an extension of the present study.

Appendix B. The PDF of a set of standard distributions

For the prior distribution of the parameters and hyperparameters of the MFMs in § 3, a set of standard distributions was used. table 2 summarizes the PDF and associated support of such distributions.

References

ANSYS Fluent 2019 Theory guide. Release 2019R3.

Bertagnolio, F. 2008 NACA0015 measurements in LM wind tunnel and turbulence generated noise. Denmark. *Forskningscenter Risoe*. 1657(EN). Danmarks Tekniske Universitet.

Duvenaud, D., Lloyd, J., Grosse, R., Tenenbaum, J. & Zoubin, G. 2013 Structure discovery in nonparametric regression through compositional kernel search. In *Proceedings of the 30th International Conference on Machine Learning* (ed. S. Dasgupta & D. McAllester), Proceedings of Machine Learning Research, vol. 28, pp. 1166–1174. PMLR.

Fairbanks, H.R., Doostan, A., Ketelsen, C. & Iaccarino, G. 2017 A low-rank control variate for multilevel Monte Carlo simulation of high-dimensional uncertain systems. *J. Comput. Phys.* **341**, 121–139.

Fatou Gomez, J. 2018 Multi-fidelity co-Kriging optimization using hybrid injected RANS and LES. PhD thesis, Delft University of Technology.

Forrester, A.I.J., Sóbester, A. & Keane, A.J. 2007 Multi-fidelity optimization via surrogate modelling. *Proc. R. Soc.* A **463** (2088), 3251–3269.

Fröhlich, J., Mellen, C.P., Rodi, W., Temmerman, L. & Leschziner, M.A. 2005 Highly resolved large-eddy simulation of separated flow in a channel with streamwise periodic constrictions. *J. Fluid Mech.* **526**, 19–66.

Gao, W., Cheng, W. & Samtaney, R. 2020 Large-eddy simulations of turbulent flow in a channel with streamwise periodic constrictions. *J. Fluid Mech.* **900**, A43.

Gelman, A. 2006 Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** (3), 515–534.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. 2013 *Bayesian Data Analysis*, 3rd ed. Taylor & Francis.

GERACI, G., ELDRED, M.S. & IACCARINO, G. 2017 A multifidelity multilevel Monte Carlo method for uncertainty propagation in aerospace applications. In *19th AIAA Non-Deterministic Approaches Conference*. Available at: https://arc.aiaa.org/doi/pdf/10.2514/6.2017-1951.

GILES, M.B. 2008 Multilevel Monte Carlo path simulation. *Oper. Res.* **56** (3), 607–617.

GILLING, L., SØRENSEN, N. & DAVIDSON, L. 2009 Detached eddy simulations of an airfoil in turbulent inflow. In *47th AIAA Aerospace Sciences Meeting*, vol. 24 (AIAA 2009-270), pp. 1–13. American Institute of Aeronautics and Astronautics.

GOH, J., BINGHAM, D., HOLLOWAY, J.P., GROSSKOPF, M.J., KURANZ, C.C. & RUTTER, E. 2013 Prediction and computer model calibration using outputs from multifidelity simulators. *Technometrics* **55** (4), 501–512.

GRAMACY, R.B. 2020 *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Chapman Hall/CRC.

GRATIET, L.L. & GARNIER, J. 2014 Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *Intl J. Uncertain. Quantif.* **4** (5), 365–386.

HAN, Z.-H. & GÖRTZ, S. 2012 Hierarchical kriging model for variable-fidelity surrogate modeling. *AIAA J.* **50** (9), 1885–1896.

HIGDON, D., KENNEDY, M., CAVENDISH, J.C., CAFEO, J.A. & RYNE, R.D. 2004 Combining field data and computer simulations for calibration and prediction. *SIAM J. Sci. Comput.* **26** (2), 448–466.

HOFFMAN, M.D. & GELMAN, A. 2014 The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Machine Learning Res.* **15** (47), 1593–1623.

IWAMOTO, K., SUZUKI, Y. & KASAGI, N. 2002 Reynolds number effect on wall turbulence: toward effective feedback control. *Intl J. Heat Fluid Flow* **23** (5), 678–689.

JOFRE, L., GERACI, G., FAIRBANKS, H.R., DOOSTAN, A. & IACCARINO, G. 2018 Exploiting multi-fidelity strategies to quantify uncertainty in irradiated particle-laden turbulent flow. In *Proceedings of the 7th European Conference on Computational Fluid Dynamics*, pp. 1–12. European Community on Computional Methods in Applied Sciences (ECCOMAS).

KENNEDY, M.C. & O'HAGAN, A. 2000 Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87** (1), 1–13.

KENNEDY, M.C. & O'HAGAN, A. 2001 Bayesian calibration of computer models. *J. R. Stat. Soc.: Ser.* B *(Stat. Methodol.)* **63** (3), 425–464.

LARSSON, J., KAWAI, S, BODART, J. & BERMEJO-MORENO, I. 2016 Large eddy simulation with modeled wall-stress: recent progress and future directions. *Bull. JSME* **3** (1), 15-00418.

LEE, M. & MOSER, R.D. 2015 Direct numerical simulation of turbulent channel flow up to $Re_\tau \approx 5200$. *J. Fluid Mech.* **774**, 395–415.

LLOYD, J.R., DUVENAUD, D., GROSSE, R., TENENBAUM, J.B. & GHAHRAMANI, Z. 2014 Automatic construction and natural-language description of nonparametric regression models. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, p. 1242–1250. AAAI.

MATERN, B. 1986 *Spatial Variation*. Springer.

MORITA, Y., REZAEIRAVESH, S., TABATABAEI, N., VINUESA, R., FUKAGATA, K. & SCHLATTER, P. 2022 Applying Bayesian optimization with Gaussian process regression to computational fluid dynamics problems. *J. Comput. Phys.* **449**, 110788.

NG, L.W.-T. & ELDRED, M. 2012 Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference* (AIAA 2012-1852), pp. 1–17. American Institute of Aeronautics and Astronautics.

PALAR, P.S., TSUCHIYA, T. & PARKS, G.T. 2016 Multi-fidelity non-intrusive polynomial chaos based on regression. *Comput. Meth. Appl. Mech. Engng* **305**, 579–606.

PASUPATHY, R., SCHMEISER, B.W., TAAFFE, M.R. & WANG, J. 2012 Control-variate estimation using estimated control means. *IIE Trans.* **44** (5), 381–385.

PEHERSTORFER, B., WILLCOX, K. & GUNZBURGER, M. 2018 Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev.* **60** (3), 550–591.

PERDIKARIS, P., VENTURI, D., ROYSET, J.O. & KARNIADAKIS, G.E. 2015 Multi-fidelity modelling via recursive co-kriging and Gaussian-Markov random fields. *Proc. R. Soc.* A*: Math. Phys. Engng Sci.* **471** (2179), 20150018.

RASMUSSEN, C.E. & WILLIAMS, C.K.I. 2005 *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. MIT Press.

REYNOLDS, W.C. & TIEDERMAN, G.W.M.F. 1967 Stability of turbulent channel flow, with application to Malkus's theory. *J. Fluid Mech.* **27** (part 2), 253–272.

REZAEIRAVESH, S. & LIEFVENDAHL, M. 2018 Effect of grid resolution on large eddy simulation of wall-bounded turbulence. *Phys. Fluids* **30** (5), 055106.

REZAEIRAVESH, S., VINUESA, R. & SCHLATTER, P. 2021 UQit: a python package for uncertainty quantification (UQ) in computational fluid dynamics (CFD). *J. Open Source Softw.* **6** (60), 2871.

REZAEIRAVESH, S., VINUESA, R. & SCHLATTER, P. 2022 An uncertainty-quantification framework for assessing accuracy, sensitivity, and robustness in computational fluid dynamics. *J. Comput. Sci.* **62**, 101688.

SAGAUT, P., DECK, S. & TERRACOL, M. 2013 *Multiscale and Multiresolution Approaches in Turbulence*, 2nd edn. Imperial College Press.

SALVATIER, J., WIECKI, T.V. & FONNESBECK, C. 2016 Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55.

SMITH, R.C. 2013 *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics.

SOBOL, I.M. 2001 Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Maths Comput. Simul.* **55** (1), 271–280.

VOET, L.J.A., AHLFELD, R., GAYMANN, A., LAIZET, S. & MONTOMOLI, F. 2021 A hybrid approach combining DNS and RANS simulations to quantify uncertainties in turbulence modelling. *Appl. Math. Model.* **89**, 885–906.

WELLER, H., TABOR, G., JASAK, H. & FUREBY, C. 1998 A tensorial approach to computational continuum mechanics using object-oriented techniques. *Comput. Phys.* **12** (6), 620–631.

WILCOX, D.C. 2006 *Turbulence Modeling for CFD*. DCW Industries.

XIAO, H., WI, J.-L., LAIZET, S. & DUAN, L. 2020 Flows over periodic hills of parameterized geometries: a dataset for data-driven turbulence modeling from direct simulations. *Comput. Fluids* **200**, 104431.

XIU, D. & KARNIADAKIS, G.E. 2002 The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM. J. Sci. Comput.* **24** (2), 619–644.

YAMAMOTO, Y. & TSUJI, Y. 2018 Numerical evidence of logarithmic regions in channel flow at $Re_\tau = 8000$. *Phys. Rev. Fluids* **3**, 012602.