Bayesian Model-Scenario Averaged predictions of RANS uncertainties using MAP estimates.

Wouter N. Edeling[1]
Center for Turbulence Research, Stanford University, Stanford, CA 94305

Martin Schmelzer[2] and Richard P. Dwight[3]
Faculty of Aerospace Engineering, Delft University of Technology,

Kluyverweg 2, 2629 HS Delft, The Netherlands

Paola Cinnella[4]
Laboratoire DynFluid, Arts et Métiers ParisTech,

151 Boulevard de l'Hopital, 75013 Paris, France

Computational Fluid Dynamics analyses of high Reynolds-number flows mostly rely on the Reynolds-Averaged Navier-Stokes equations. The associated closure models are based on multiple simplifying assumptions and involve numerous empirical closure coefficients, calibrated on a set of simple reference flows. Predicting new flows using a single closure model with nominal values for the closure coefficients may lead to biased predictions. Bayesian Model-Scenario Averaging is a statistical technique providing an optimal way to combine the predictions of several competing models calibrated on various sets of data (scenarios). The method allows to obtain a stochastic estimate of a Quantity-of-Interest in an unmeasured prediction scenario by i) propagating posterior probability distributions of the parameters obtained for multiple calibration scenarios, and ii) by computing a weighted posterior predictive distribution. While step ii) has a negligible computational cost, step i) requires a large number of samples of the solver.

---

[1] Postdoc, Center for Turbulence Research, 488 Escondido Mall, 500A, Stanford, CA 94305-3035
[2] PhD candidate, Department of Aerodynamics, Kluyverweg 2, 2629 HS Delft, The Netherlands
[3] Professor, Department of Aerodynamics, Kluyverweg 2, 2629 HS Delft, The Netherlands
[4] Professor, Laboratoire DynFluid, 151 Boulevard de l'Hopital, 75013 Paris, France

To enable the application of the proposed approach to computationally expensive flow configurations, we use a modified formulation, where a maximum posterior probability approximation is used to drastically reduce the computational burden. The predictive capability of the proposed simplified approach is assessed for 2D separated and 3D compressible flows.

I. Introduction

Direct numerical computation of turbulent flow fields is computationally intractable for most applications. Several levels of approximation are possible, according to flow scales that are resolved/modelled. In all cases, scale separation introduces unclosed terms that need to be modelled. The choice of the appropriate modelling level remains essentially a matter of expert judgement. In particular, it will always depend on cost versus accuracy considerations. On the other hand, even once a given level has been selected, the simplifying assumptions introduce unclosed terms for which several possible sub-models may be designed. These may differ both in their mathematical structure and their closure parameters. The common practice in turbulence modelling is often to leave the choice of a specific model structure to expert judgement, while treating model constants as adjustable parameters that are calibrated in such a way as to reproduce simple, well-documented flows. Both of the preceding aspects, however, represent sources of uncertainty in the prediction of a new flow.

While more general modelling approaches like Large Eddy Simulation (LES) have made considerable progress in the last decade, turbulence models based on Reynolds Averaged Navier Stokes (RANS) will remain the work-horse tool for engineering design and optimization in the decade to come. An extremely large variety of RANS models have been proposed in the past, ranging from simple algebraic models to sophisticated Reynolds Stress models (see [1] for a review). However, in order to arrive at any given closure model, numerous assumptions in the mathematical derivation and subsequent calibration must be made. No universally accepted and valid model has been identified in CFD literature, meaning that the appropriateness of a given model structure is highly

uncertain when predicting a new configuration. Additionally, RANS models are developed and calibrated on simple flows using noisy experimental data, and applied to radically different flows with the same coefficients. Furthermore, since these calibrations are deterministic (and as experimental uncertainty is generally ignored), one obtains point estimates for the closure coefficients. This ultimately leads to predictions which are subject to unknown degrees of uncertainty, especially if the flow configuration of interest is far removed from the flow scenarios under which the model was calibrated.

The above-mentioned uncertainties in RANS models can be addressed by means of statistical tools. All such tools are meant to quantify the uncertainty in RANS closures, and should not be considered as a replacement for deterministic simulations. In the context of uncertainty quantification (UQ), uncertainties associated with the closure coefficients are referred to as parametric uncertainties. These can be effectively addressed by replacing standard deterministic calibration with statistical calibration, see e.g. [2–4]. Specifically, Bayesian calibration allows one to infer posterior probability distributions of the closure parameters given some available set of experimental data. However, it is important to stress that a posterior distribution might only display a predictive capability for the flow case for which it was calibrated. In [5], we performed multiple Bayesian calibrations for the coefficients of the well-known $k-\varepsilon$ model [6], using experimental velocity data from boundary-layers subject to a wide range of different pressure gradients. Even though the flow topology was the same for all calibration cases (i.e. boundary-layers), the resulting posterior distributions were significantly different. This variability of posterior distributions is a measure for the lack of predictive capability of a given model due to the simplifying assumptions in its mathematical structure. The model needs adjusted coefficients for each new case to counterbalance deficiencies inherent to neglecting part of the flow physics.

The second class of uncertainty, namely model-form uncertainty, arises through the assumptions that are made in the mathematical form of the closure model. Many different models exist, and their performance is known to be flow dependent, see e.g. [7, 8]. Previous attempts to tackle model-form uncertainty include calibrations of a modified eddy viscosity field [9], the introduction of statistical error terms in the calibration procedure [2, 10] or in the model transport equations [11], and the use

of a stochastic counterpart of the Reynolds stress tensor [12]. Even UQ techniques for the RANS equations with no modelling, where the transport terms are represented with direct numerical simulation data, have been developed [13, 14]. All these methods rely on detailed data from direct numerical simulations to infer about stochastic fields. Although promising, the above-mentioned methodologies suffer from the following drawbacks: i) DNS data are available only for relatively simple, low-Reynolds configurations; ii) inferring the stochastic fields involves costly optimization algorithms, e.g. to maximize the likelihood of observing the data; iii) extrapolating the stochastic correction terms to the prediction of unseen configurations is still an open problem, especially for realistic flow configurations.

Other means for representing model-form uncertainty include the use of a multi-model framework. A relevant example can be found in Poroseva et al. [15], where Dempster-Shafer evidence theory is used to obtain a measure of the total uncertainty in the mean velocity predictions around a RAE 2822 airfoil. Dempster' rule is used to fuse the predictions of two closure models, and using only sparse experimental data, the authors of [15] predict the velocity with quantified uncertainty at locations in the same flow where no data is available. Multi-model approaches have also found application in meteorology [16, 17], hydrology [18] and climate science [19], to name just a few.

We will focus on Bayesian Model-Scenario Averaging [20] (BMSA), which provides a coherent framework to address both parametric and model-form uncertainties, providing a robust predictive method (see for instance [21] for an application to groundwater modelling uncertainties). Like Bayesian Model Averaging (BMA), BMSA combines the predictions from multiple models, thereby providing a measure for (closure) model uncertainty [22]. In addition, BMSA can combine the posterior distributions from different calibration scenarios, which allows one to compute error estimates due to scenario variability. Unless this variability is negligible, one should not assume that obtained parameter estimates are extrapolative to other flow scenarios. In other words: combining multiple non-overlapping posteriors adds uncertainty for the assumption that a posterior distribution can be applied outside the flow scenario under which it was calibrated.

In this article, we rely solely on RANS models for prediction. Hence we cannot assume that the true model which generated the data is contained in our selected model set, although in principle

higher fidelity models could be included in the Bayesian average, at increased computational cost. It should therefore be noted that subject-matter expertise, applied to the selection of competing models (and scenarios), is still important for maintaining predictive accuracy. As noted by Draper[20], although multi-model approaches still introduce a bias in the prediction, due to the selection of a finite set of model structures, they play a useful role in reducing the bias with respect to predictions based on a single model structure.

In previous work [23], BMSA was used to construct a stochastic model for predicting boundary layer flows subject to an arbitrary pressure gradient, based on multiple RANS models calibrated using selected experimental boundary-layer data. Specifically, we applied BMSA using 5 closure models and 14 boundary-layer calibration scenarios (corresponding to various external pressure gradients) to predict unobserved boundary-layer flows subject to an arbitrary pressure gradient. Each individual calibration was performed using the Bayesian method developed by the authors of [2, 24]. Other notable means of stochastic parameter estimation can be found in [25, 26]. Both the calibration and the prediction phases relied on a fast boundary-layer code. When addressing more complex, computationally intensive, flow configurations, a clear downside of BMSA is the associated computational burden. The first cost is associated to the calibration phase, in which the Markov-Chain Monte Carlo method [27] is used to draw samples from the posterior distribution. Furthermore, the predictive phase of the BMSA procedure dictates the need to propagate $I \times K$ (I being the number of competing models and K the number of calibration scenarios) posterior distributions through a RANS code applied to the flow scenario of interest, which is prohibitively expensive for most practical flow configurations. A common option is to replace the expensive RANS code with a much cheaper surrogate model, such as for instance a polynomial approximation created with stochastic collocation or polynomial chaos methods [28]. Still, creating the surrogate model will require samples from the full RANS code, and surrogate modeling techniques are subject to the so-called curse of dimensionality. The number of required code samples therefore rises exponentially with the number of unknown parameters.

Our goal for this paper is to assess the predictive capability of the BMSA method for computationally expensive flow cases, while using the most drastic cost reduction technique possible

to speed up the prediction step. We assume that calibration cost is less critical, since it is done only once in a preliminary phase. Specifically, in this work we use the posterior distributions of [23], and propagate them through new flow configurations described by a full RANS solver instead of a boundary layer code. Note that posterior distributions resulting from additional calibration scenarios can be added as soon as they become available.

To drastically reduce the computational cost associated with the propagation step, we approximate each posterior as a Dirac delta distribution centered at the *Maximum A Posteriori* (MAP) estimate. Now, instead of having to propagate $I \times K$ posteriors, we are left with $I \times K$ deterministic evaluations. In order to evaluate the predictive capability of the proposed modified BMSA approach, we selected a number of flow problems which include physics that the boundary-layer trained BMSA model was not subjected to during the calibration phase.

This paper is organized as follows. First, we will give a very brief overview of the competing closure models considered in this paper. In Section III, we describe our reduced BMSA method using MAP estimates. Preliminary validations of the MAP-base BMSA approach are carried out in Section IV using the boundary layer flow cases of [23]. Section V shows the results obtained for the different prediction flow cases. Finally, we give our conclusion in Section VI.

II. RANS turbulence models

In the present work, we restrict our attention to three transport equation models largely employed in engineering applications. These are selected from the family of so-called linear eddy-viscosity models, which adopt a linear representation of the Reynolds stress tensor, based on the well-known Boussinesq assumption. We use the $k-\varepsilon$, $k-\omega$ and Spalart-Allmaras (SA) models.

As these are very well-known models, we do not give full details for the sake of conciseness. Rather, we refer to [1] for a complete description of their mathematical structure. Also, all values for the closure coefficients (both nominal and perturbed), can be found in Appendix A.

III. Bayesian predictive Methodology

We describe a statistical methodology based on Bayesian Model Scenario Averaging (BMSA), for making predictions of a chosen Quantity of Interest (QoI). BMSA provides a coherent framework

for combining predictions from multiple competing conceptual models and calibration data to attain a more realistic and reliable description of the predictive uncertainty.

A. BMSA formulation

Consider a particular flow of interest, including boundary-conditions, material parameters, and all other physical properties needed to define it unambiguously. This we term a *scenario*, denoted S. Assume that there is reference data $\boldsymbol{z}$ available for S, which may be measurements of any flow quantity or data from higher fidelity simulations (e.g. DNS). Let $m_{\mathrm{CFD}}(S; M, \boldsymbol{\theta})$ be a governing model, here represented by a CFD code, which takes as arguments the scenario S, a turbulence model M and its closure-coefficients $\boldsymbol{\theta}$. Note that the size of $\boldsymbol{\theta}$ and the meaning of its components will depend on M. Let the model return the full state of the fluid in all variables. Then we can define an operator $H_z(\cdot)$ that maps the state to the observed quantities $\boldsymbol{z}$ (this can be, e.g., a procedure for extracting the aerodynamic coefficients from the full solution for the flow around a body or a velocity profile at a given station). The model output is then given by $H_z \circ m_{\mathrm{CFD}}(S; M, \boldsymbol{\theta})$ and can be related to the observed data by means of a statistical model. One of such models, used in [5, 23] is represented by:

$$\boldsymbol{z} = \eta \; H_z \circ m_{\mathrm{CFD}}(S; M, \boldsymbol{\theta}) + \epsilon, \tag{1}$$

where ϵ, η are random-variables (RVs) representing measurement noise and model error respectively. The latter accounts for the fact that, even with the best possible parameters, the model does not predict the true value for $\boldsymbol{z}$, due to errors intrinsic to modelling assumptions. The multiplicative model-error term of (1) was first introduced by [29]. Additive models are also available, see e.g. [26]. We assume throughout that the CFD calculations have run on sufficiently fine grids, such that errors due to grid resolution are significantly less than the uncertainty due to modelling assumptions. We also assume that the calculations have well converged to the steady state solution. The experimental noise term ϵ is modelled using a zero-mean Gaussian distribution.

The observations available for a given S can be used to infer the model parameters $\boldsymbol{\theta}$ using a

Bayesian calibration approach, i.e. by applying Bayes' rule:

$$p\left(\boldsymbol{\theta}|\boldsymbol{z}, M, S\right) = \frac{p\left(\boldsymbol{z}|\boldsymbol{\theta}, M, S\right) p(\boldsymbol{\theta} \mid M, S)}{p(\boldsymbol{z} \mid M, S)}, \tag{2}$$

where $p\left(\boldsymbol{\theta}|\boldsymbol{z}, M, S\right)$ is the joint posterior probability distribution of the closure coefficients under the given turbulence model and the calibration data $\boldsymbol{z}$ observed for scenario S. Furthermore, $p(\boldsymbol{\theta} \mid M, S)$ is the joint prior probability distribution for the model coefficients, which summarizes available knowledge about $\boldsymbol{\theta}$ before observing any data. The likelihood function $p\left(\boldsymbol{z}|\boldsymbol{\theta}, M, S\right)$, which stems from the statistical model (1), describes the probability of observing the data for scenario S, given a model and a realization of the closure coefficients. Evaluating $p\left(\boldsymbol{z}|\boldsymbol{\theta}, M, S\right)$ for a given realization of $\boldsymbol{\theta}$ involves running m_{CFD} and is, as such, a costly operation. Finally $p(\boldsymbol{z}|M, S)$ is the evidence of the data, which normalizes the posterior distribution. Since $p(\boldsymbol{z}|M, S)$ does not depend upon $\boldsymbol{\theta}$, it does not have to be computed while calibrating $\boldsymbol{\theta}$.

Given the complexity of the likelihood function, the posterior $p(\boldsymbol{\theta}|\boldsymbol{z}, M, S)$ cannot be obtained analytically and is instead evaluated numerically. We employ the Markov-Chain Monte-Carlo method [27] to draw samples from (2). To reach convergence of the Markov-chain, we observed that $\mathcal{O}\left(10^4\right)$ code samples were required [5]. Ordinarily this would constitute an excessive strain on available computational resources in a CFD context. In previous work [23] our experimental data consisted of boundary-layer quantities, and m_{CFD} reduced to a fast boundary-layer code. As such, no real computational bottleneck exists during the calibration phase.

The calibration outcome is a discrete approximation of the joint posterior probability distribution of the parameters, which is then available for the prediction of new (unobserved) scenarios. Let Δ be a quantity of interest (QoI) to be predicted for a new scenario $\tilde{S}$, for which we do not have data. The QoI may be a scalar, vector, or functional quantity derived from the flow-state. Since m_{CFD} is a model for Δ, we write

$$\Delta \simeq H_{\Delta} \circ m_{\mathrm{CFD}}(\tilde{S}; M, \boldsymbol{\theta}), \tag{3}$$

where, analogous to H_z, H_{Δ} is an observation operator extracting Δ from the full flow-state, M is the closure model adopted for the prediction, $\tilde{S}$ is the prediction scenario. Equation (3) is a prediction

based on a single choice for M and the associated parameters, i.e. a deterministic prediction. A stochastic prediction accounting for parameter uncertainty is obtained by propagating the posterior distributions for $\boldsymbol{\theta}$ through (3). This however does not account for the fact that different distributions may be obtained by calibrating the model against different scenarios, nor for the fact that multiple alternative closure models M may be used.

To estimate the effect of these uncertainties, we build the full BMSA predictive distribution of Δ for a set of alternative conceptual models $\mathcal{M} = (M_1, M_2, ..., M_i, ..., M_I)$, under different calibration scenarios $\mathcal{S} = (S_1, S_2, ..., S_k, ..., S_K)$ for which we have data $\mathcal{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_k, ..., \boldsymbol{z}_K)$[20, 22]:

$$p(\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S}) = \sum_{k=1}^{K} \sum_{i=1}^{I} p(\Delta \mid \tilde{S}, \boldsymbol{z}_k, M_i, S_k)\, p(M_i \mid \boldsymbol{z}_k, S_k)\, p(S_k) \tag{4}$$

Equation (4) is an average of the $I \times K$ posterior predictive distributions (ppds) $p(\Delta \mid \tilde{S}, \boldsymbol{z}_k, M_i, S_k)$. These are weighted by the posterior model probabilities $p(M_i \mid \boldsymbol{z}_k, S_k)$ and by scenario probabilities $p(S_k)$. The ppds are obtained by marginalizing Δ over the posterior distributions of the parameters:

$$p(\Delta \mid \tilde{S}, \boldsymbol{z}_k, M_i, S_k) = \int p(\Delta \mid \tilde{S}, M_i, S_k, \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \boldsymbol{z}_k, M_i, S_k)\, d\boldsymbol{\theta} \tag{5}$$

In simpler terms, this amounts to propagating the $I \times K$ posterior distributions $p(\boldsymbol{\theta} \mid \boldsymbol{z}_k, M_i, S_k)$ through the new scenario $\tilde{S}$ by a suitable uncertainty quantification method.

The posterior model probabilities, on the other hand, reflect how well model M_i fits the data over the support of the prior. They can be determined from a separate application of Bayes' rule, i.e.

$$p\left(M_i \mid \boldsymbol{z}_k, S_k\right) = \frac{p(\boldsymbol{z}_k \mid M_i, S_k) p(M_i \mid S_k)}{\sum_{j=1}^{I} p(\boldsymbol{z}_k \mid M_j, S_k) p(M_j \mid S_k)}, \tag{6}$$

where $p(M_i \mid S_k)$ are the prior model probabilities (generally modelled as equi-probable). Note that $p(\boldsymbol{z}_k \mid M_i, S_k)$ is the evidence for model M_i and scenario S_k, previously introduced in (2). Thus, in case of a model ensemble, the denominator of (2) is not ignored and is computed via

$$p(\boldsymbol{z}_k \mid M_i, S_k) = \int p(\boldsymbol{z}_k \mid M_i, S_k, \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid M_i, S_k) \mathrm{d}\boldsymbol{\theta}. \tag{7}$$

The first two moments of $p(\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S})$ can be derived directly from (4) and read [20, 23]:

$$\mathbb{E}[\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S}] = \sum_{i=1}^{I} \sum_{k=1}^{K} \mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, M_i, S_k]\, p(M_i \mid \boldsymbol{z}_k, S_k)\, p(S_k) \tag{8}$$

$$\mathbb{V}[\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S}] = \sum_{i=1}^{I}\sum_{k=1}^{K} \mathbb{V}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, M_i, S_k]\, p(M_i \mid \boldsymbol{z}_k, S_k)\, p(S_k)$$
$$+ \sum_{i=1}^{I}\sum_{k=1}^{K} (\mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, M_i, S_k] - \mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, \mathcal{M}, S_k])^2\, p(M_i \mid \boldsymbol{z}_k, S_k)\, p(S_k)$$
$$+ \sum_{k=1}^{K} (\mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, \mathcal{M}, S_k] - \mathbb{E}[\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S}])^2\, p(S_k), \tag{9}$$

where $\mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, \mathcal{M}, S_k] := \sum_{i=1}^{I} \mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, M_i, S_k] p(M_i \mid \boldsymbol{z}_k, S_k)$ is the BMA prediction within scenario S_k. Equation (9) shows that the posterior predictive distribution variance for Δ consists of three terms: the first term is called the *within model, within scenario* variance, and accounts for predictive uncertainties due to the residual parameter variability after calibration of a model M_i under a scenario S_k. The second one is called the *between model, within scenario* variance. It is large if the models, using the posterior parameter distributions originating from the same calibration scenario, provide different predictions for the new configuration. It can therefore be considered as a measure for model error. The last term is the *between scenario* variance, which accounts for the fact that different calibration scenarios lead to different estimates of the parameters. Thus, when applying the BMSA model to a prediction scenario $\tilde{S}$ for which it was not calibrated, this uncertainty should be taken into account.

B. Implementation of BMSA for CFD problems

In this paper, we focus on the prediction step, and refer the reader to [5, 23] for details about the calibration step. All flow predictions presented in the following are based on the posterior parameter distributions previously computed for a class of flat plate boundary layer flows subject to various external pressure gradients [23]. These constitute a database of posterior probability distributions, which is available for propagation through new flow configurations. The interested reader may download the samples of the posterior $\boldsymbol{\theta}$ distributions used in the following applications from [30]. Since these parameter estimates where determined for wall-bounded flow configurations, we expect these may be used to predict other (more complex) wall bounded flows. The validity of this working assumption will be assessed in subsequent sections.

The calibration step also provides information about the posterior model probabilities (6). Indeed, for each model and calibration scenario, it is possible to compute the evidence by Monte-

Carlo integration of (7). For instance, for the cheap flat plate boundary layer scenarios, the evidence is computed by drawing a large number of samples, $\boldsymbol{\theta}_n,\ n = 1, \cdots, N$, from the prior $p(\boldsymbol{\theta} \mid M_i, S_k)$, and by approximating (7) as

$$p(\boldsymbol{z}_k \mid M_i, S_k) \approx \frac{1}{N} \sum_{n=1}^{N} p(\boldsymbol{z}_k \mid M_i, S_k, \boldsymbol{\theta}_n). \tag{10}$$

The posterior model weights can now be calculated through the simple algebraic relation (6), the results of which we tabulated in Appendix A. It should be noted that computing the evidence using samples from the prior is not always feasible. Convergence of the MC estimate (10) will become slow if the region where the likelihood has high probability is significantly different from the region of high prior probability content [31]. In this case more advanced sampling techniques will be required, see e.g. [29, 32].

The prediction step involves:

1. Propagating posterior parameter distributions for each competing model through the prediction scenario $\tilde{S}$;

2. Assigning a suitable probability mass function to the scenarios S_k;

3. Applying the BMSA relations (4), and/or (8),(9) to obtain the predictive posterior distribution and/or its moments for the QoI Δ.

Steps 2. and 3. only involve a negligible or small computational load. The computational bottleneck is represented by step 1., which requires running the CFD model on scenario $\tilde{S}$ for all models in $\mathcal{M}$, and for a large number of samples in each posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{z}_k, M_i, S_k)$, leading to $I \times K$ stochastic estimates of Δ. Since each CFD simulation is likely to be expensive itself, propagation by means of Monte Carlo sampling is unacceptably costly. The computational cost can be greatly alleviated by using surrogate models (see e.g. [33]), yet a significant number of CFD runs ($\mathcal{O}(100)$ or higher) is still required even when using advanced surrogate modelling techniques. For this reason, in Section III D we introduce an approximation that drastically reduces the computational cost associated to the prediction step. In the next section we complete the description of the BMSA by discussing the choice of the scenario probability mass function (step 2).

C. Weighting the scenarios

Whereas the posterior model probabilities $p\left(M_i \mid S_k, \boldsymbol{z}_k\right)$ are informed from the calibration data $\boldsymbol{z}_k$, the scenario probabilities $p\left(S_k\right)$ are of a predictive nature, and incorporate uncertainty for the fact that the MAP estimates are applied to a predictive scenario $\tilde{S}$ for which they were not calibrated. Ideally experimental data from $\tilde{S}$ would be available, which we could use to inform $p\left(S_k\right)$. However, since in many engineering applications of CFD this will not be the case, we assume we only have data coming from the K calibration scenarios S_k. Instead of informing on data, we specify $p(S_k)$ based on model agreement per S_k. Specifically, following to [23], we set:

$$p(S_k) := \frac{\xi_k^{-p}}{\sum_{j=1}^{K} \xi_j^{-p}}, \quad \xi_k = \sum_{i=1}^{I} \| \mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, M_i, S_k] - \mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, \mathcal{M}, S_k] \|_2 \quad \forall S_k \in \mathcal{S}. \tag{11}$$

Here, ξ_k measures model agreement by means of an L_2-normed variation around the BMA expectation $\mathbb{E}[\Delta \mid \tilde{S}, \boldsymbol{z}_k, \mathcal{M}, S_k]$. If all turbulence models show a high level of agreement regarding the value of Δ when using the posteriors calibrated on S_k, such a scenario is given a high probability and vice versa. The rationale is that if a given S_k is similar to the predictive $\tilde{S}$ at hand, the models are expected to give similar predictions given that the parameter distributions were calibrated on the same data $\boldsymbol{z}_k$. Additionally, $p \in \mathbb{N}_0$ is an integer which controls the degree to which variation is penalized. For $p = 0$ we obtain a uniform distribution and for higher values of p we effectively damp the results coming from those scenarios with a low level of model agreement ξ_k. In section IV we carry out numerical experiments to demonstrate that ξ_k indeed represents a good model for the true discrepancy between the BMSA prediction and the data and we discuss the role of the exponent p. Finally, note that (11) requires a minimum of two models, the use of one model will always yield a uniform distribution. With just a single model, one might still use the variability over S_k to inform the $p(S_k)$, or simply rely on expert opinion.

A valid criticism of (11) is that the 'best' S_k will not be favoured if all models in $\mathcal{M}$ for a given 'wrong' S_k produce predictions which are incorrect in a very similar manner. Due to the fact that we do not allow ourselves any reference data for the predictive flow scenario, we cannot fully exclude the possibility of such a situation. However, we can suggest means to decrease its risk through: i) selecting a larger set of diverse models of different fidelity, which are unlikely to agree with each other, ii) selecting a relevant set of calibration scenarios, since in previous research [23] we found

that (11) does favour S_k similar to $\tilde{S}$ if the flow topology is comparable, and iii) limit the value of p, which gives the user control over how much $p(S_k)$ is allowed to deviate from a uniform distribution.

D. Speeding up BMSA predictions: MAP approximation of the posterior parameter distributions

A way to drive the estimate of the individual predictive distributions $p(\Delta \mid \tilde{S}, \mathbf{z}_k, M_i, S_k)$ to an acceptable computational cost consists in approximating the marginal posterior probability distributions $p(\boldsymbol{\theta} \mid \boldsymbol{z}_k, M_i, S_k)$ with Dirac-δ functions at their *maximum a posteriori* (MAP) values:

$$\boldsymbol{\theta}_{i,k}^{\mathrm{MAP}} := \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\; p(\boldsymbol{\theta} \mid \boldsymbol{z}_k, S_k, M_i) \tag{12}$$

so that

$$p(\boldsymbol{\theta} \mid \boldsymbol{z}_k, S_k, M_i) \simeq \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_{i,k}^{\mathrm{MAP}}\right). \tag{13}$$

The consequence of this approximation is to neglect the effect of "within-model, within-scenario" variance. The ppd variance still includes the effect of multiple models and scenarios. Note that if perfectly plentiful data were available in the training scenarios (and the models were able to fit the data exactly for some values of the closure coefficients), then the $p(\boldsymbol{\theta} \mid S_k, M_i, z_k)$ would tend toward Dirac δ-functions. So one way to think of this approximation is as neglecting the effect of imperfect training. The MAP estimates (see Appendix A), were estimated from kernel-density estimates constructed with McMC samples of previous research [30], see also Figure 1. While this might only be an approximation to each individual MAP value, it is the spread of different MAP values that is important in our approach, see the discussion in Section IV A.

Substituting (13) into (5) leads to an approximation of the posterior predictive distribution $p(\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S}) \simeq \hat{p}(\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S})$ by

$$\hat{p}(\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S}) = \sum_{i=1}^{I} \sum_{k=1}^{K} \delta\left(\Delta - \Delta_{\mathrm{CFD}}\left(\tilde{S}; M_i, \boldsymbol{\theta}_{i,k}^{\mathrm{MAP}}\right)\right)\, p(M_i \mid S_k, \boldsymbol{z}_k)\, p(S_k) \tag{14}$$

where we define $\Delta_{\mathrm{CFD}}(\tilde{S}) := H_\Delta \circ m_{\mathrm{CFD}}(\tilde{S})$ and we used the fact that propagating a δ distribution requires running m_{CFD} only at $\boldsymbol{\theta} = \boldsymbol{\theta}_{i,k}^{\mathrm{MAP}}$, returning a δ distribution centered at $\Delta_{\mathrm{CFD}}(\tilde{S})$ for the output. The approximate ppd is then reduced to a weighted-sum of $I \times K$ δ-functions, one at each prediction of $\Delta_{\mathrm{CFD}}(\tilde{S})$ for each model, and each scenario's MAP-estimate of $\boldsymbol{\theta}$. The cost of

evaluating the ppd is then $I \times K$ runs of $\Delta_{\mathrm{CFD}}(\tilde{S})$. Finally, the mean and variance of (14), are easy to evaluate using the properties of δ-functions, and contain terms from our database ($p(M_i \mid S_k, \boldsymbol{z}_k)$ and $\boldsymbol{\theta}_{i,k}^{\mathrm{MAP}}$), a user-choice ($p(S_k)$) and the RANS output for Δ via m_{CFD}:

$$\mathbb{E}[\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S}] = \sum_{i=1}^{I} \sum_{k=1}^{K} \Delta_{\mathrm{CFD}} \left(\tilde{S}; M_i, \boldsymbol{\theta}_{i,k}^{\mathrm{MAP}} \right) p(M_i \mid S_k, \boldsymbol{z}_k)\, p(S_k) \tag{15}$$

$$\begin{aligned} \mathbb{V}[\Delta \mid \tilde{S}, \mathcal{Z}, \mathcal{M}, \mathcal{S}] = & \sum_{k=1}^{K} \sum_{i=1}^{I} \left[\Delta_{\mathrm{CFD}} \left(\tilde{S}; M_i, \boldsymbol{\theta}_{i,k}^{\mathrm{MAP}} \right) - \mathbb{E}[\Delta \mid \tilde{S}, z_k] \right]^2 p(M_i \mid S_k, z_k)\, p(S_k) \\ & + \sum_{k=1}^{K} \left(\mathbb{E}[\Delta \mid \tilde{S}, z_k] - \mathbb{E}[\Delta \mid \tilde{S}, \mathbf{z}] \right)^2 p(S_k). \end{aligned} \tag{16}$$

Note that compared to (9), the ppd variance for Δ now contains only two terms: the between model, within scenario variance, and the between scenario variance. The effect of the MAP approximation on the ppd variance is investigated in Section IV.

IV. Preliminary validation

In this section we use our numerical boundary-layer flow database [23] along with experimental data from [34] for the preliminary validation of two key ingredients of the present BMSA formulation, namely, the MAP approximation and the scenario weighting. The boundary-layer flows considered in the present calculations and the corresponding nomenclature are described in Appendix A.

A. Influence of the MAP approximation

The MAP approximation ignores the effect of imperfect training within each calibration scenario (i.e. it ignores the variance of each individual posterior), As an example, consider the results of Figure 1, where all 14 posterior distributions of the coefficient $C_{\varepsilon 2}$ (of the $k-\varepsilon$ model, see Appendix A) and the corresponding MAP estimates are plotted. The variance of these marginals is not considered small, due to the fact that even at the best model parameters, there is still model inadequacy which will prevent the posteriors from approaching delta distributions [24, 26, 29]. Moreover, not all parameters will be equally sensitive to the reference data. We previously observed [5, 23] that parameters with low Sobol indices will yield posteriors that do not deviate much from our (uniform) prior. As such, the MAP approximation might seem to neglect a large amount of uncertainty.

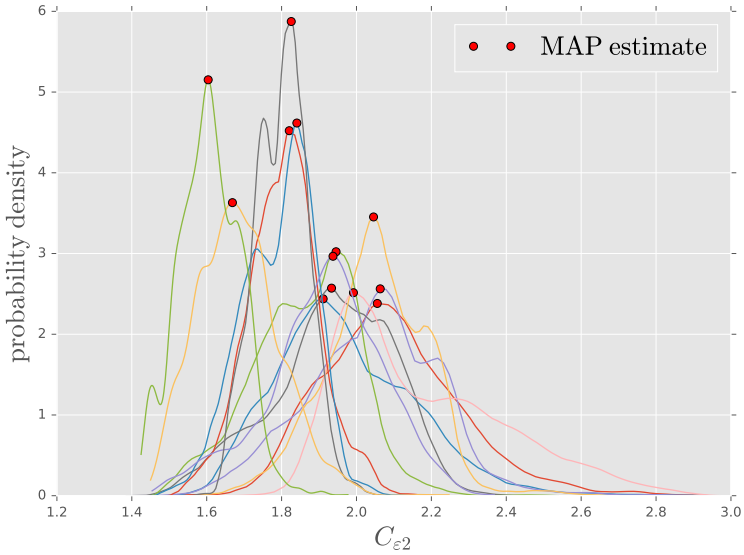
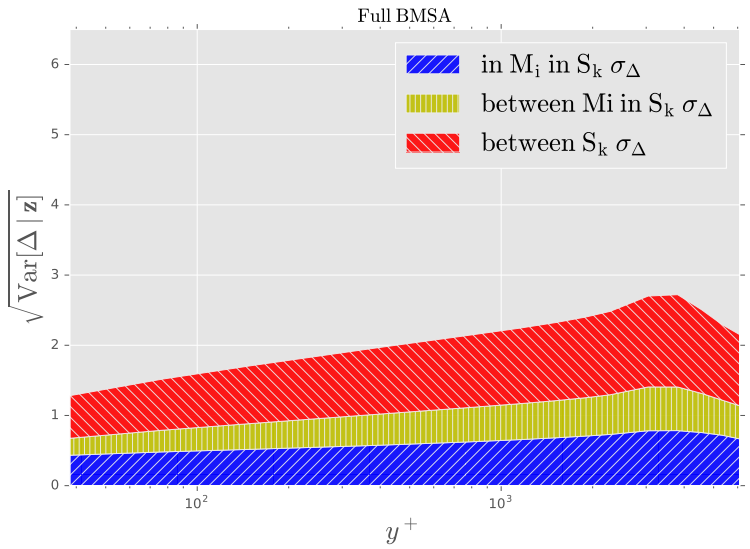


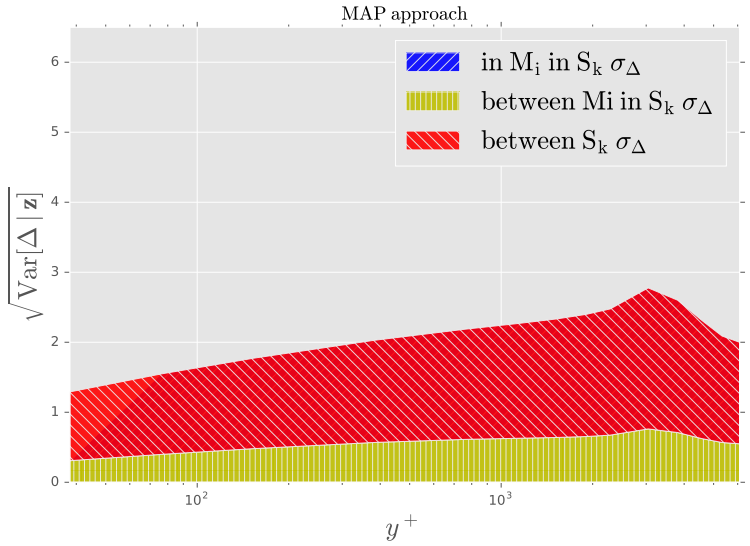
Fig. 1 All marginal posterior distributions of $C_{\varepsilon 2}$ (obtained from kernel-density estimates using the McMC samples of [23]), and the corresponding MAP estimates.

However, since we consider 14 different calibration scenarios, and since the modes of the posteriors significantly disagree on the most likely value of $\boldsymbol{\theta}$, the MAP approximations still span a large section of the $\boldsymbol{\theta}$ domain, see again Figure 1 for an example. Hence, the MAP approximation is of the least influence when this span is sufficient to yield an envelope on the QoI that is similar in size to the full posterior ppd variance, obtained by propagating all posterior $\boldsymbol{\theta}$ distributions. It should be noted however that this will only hold for lower-order statistical moments, all information contained in the tails of the posteriors will be lost in the MAP approximation.

We can evaluate the effect of the MAP approximation on the boundary-layer flows of the calibration phase, by computing the posterior standard-deviation of the full BMSA method, and the current proposed approach of Section III D. Figure 2 shows the predictive results for an unseen boundary layer, where Δ is a u^+ profile, defined as $u^+ := \langle U \rangle / u_\tau$. Here, $\langle U \rangle$ is the mean steamwise velocity and u_τ is the friction velocity. Precisely, we used posteriors from the remaining 13 boundary layer scenarios in Table 6 (Appendix A) to predict case 1400. Note that the total variability in Δ is fairly similar in both cases. Due to the lack of a within-model, within-scenario component in the MAP approach, a larger part of the standard-deviation in Δ is lumped into the between-scenario component. We repeated this procedure for a number of unseen boundary layers, and obtained



(a) Full BMSA.



(b) MAP approach.

Fig. 2 The breakdown of the posterior standard-deviation of $\Delta = u^+$, plotted versus y^+ for both the full BMSA method (computed with (9)) and the simplified MAP approach (computed with (16)). Both figures represent the the same predictive flow case, and were generated with uniform $p(S_k)$.

similar results. Thus, for those cases considered where we could compute the full BMSA ppd standard-deviation, we did not observe a significant underestimation of the posterior predictive standard-deviation when the MAP approach was used.

B. Influence of the scenario weighting

Equation (11) raises the question: is ξ_k a good measure for the true model error contained in each S_k? Let us first define the true model error as:

$$\tilde{\xi}_k := \sum_{i=1}^{I} \left\| H_z \circ m_{\text{CFD}} \left(\tilde{S}; M_i, \boldsymbol{\theta}_{i,k}^{\text{MAP}} \right) - \boldsymbol{z}_k \right\|_2 ,$$

where in our case the $\boldsymbol{z}_k$ are experimental u^+ profiles. We can then evaluate its proximity to our error-model ξ_k by performing a leave-one-out validation analysis on the K calibration scenarios. Note that a given MAP estimate $\boldsymbol{\theta}_{i,k}^{MAP}$ has not been calibrated on the data of the other $K-1$ calibration scenarios $S_j,\ j \in \{1, \cdots, K\} \setminus \{k\}$. The scenarios are roughly ordered from 'easy' zero-pressure gradient scenarios at $k = 1$, to the most difficult, highly adverse scenarios near $k = K$, see Appendix A for a more complete description. By using each $\boldsymbol{\theta}_{i,k}^{MAP}$ to compute u^+ in these remaining $K-1$ scenarios we can evaluate both ξ_k and $\tilde{\xi}_k$ due to the availability of data. The results are shown in Figure 3, where the dotted line corresponds to perfect agreement $\xi_k = \tilde{\xi}_k$. Each subplot shows the results of a single MAP estimate, applied to all other scenarios. Note that in general the results lie close to the $\xi_k = \tilde{\xi}_k$ line. However, MAP estimates calibrated under 'easy' zero or favorable pressure gradient scenarios can yield discrepancies between ξ_k and $\tilde{\xi}_k$. More specifically, we observe such discrepancies when using a MAP estimate calibrated under an easy scenario to predict a QoI under a difficult scenario, see e.g. the results of applying $\boldsymbol{\theta}_{i,1}^{MAP}$ to predict the u^+ profile of S_{14}. On the other hand, MAP estimates that were calibrated under the most difficult strongly-adverse scenarios perform well across all S_k, see the results for $\boldsymbol{\theta}_{i,13}^{MAP}$ and $\boldsymbol{\theta}_{i,14}^{MAP}$.

The parameter p of (11) is user specified. To estimate its value we use a similar leave-one-out analysis and the experimental data from the calibration scenarios. In this case we fix a scenario S_k, and use the remaining $\boldsymbol{\theta}_{i,j}^{MAP},\ j \in \{1, \cdots, K\} \setminus \{k\}$ to predict the QoI for a given value of p. Let us define a relative error in the BMSA prediction as

$$\epsilon_k^{rel} := \frac{\|\mathbb{E}\left[\Delta|\boldsymbol{z}\right] - \boldsymbol{z}_k\|_2}{\|\boldsymbol{z}_k\|} \times 100\%. \tag{17}$$

For all S_k we can now examine the behavior of (17) as a function of p. The results are depicted in Figure 4. For most S_k the relative error is either minimized or reaches a plateau for $p \in \{1, 2, 3\}$. Based on these results we will set $p = 2$ in predictive cases where no experimental data is available.

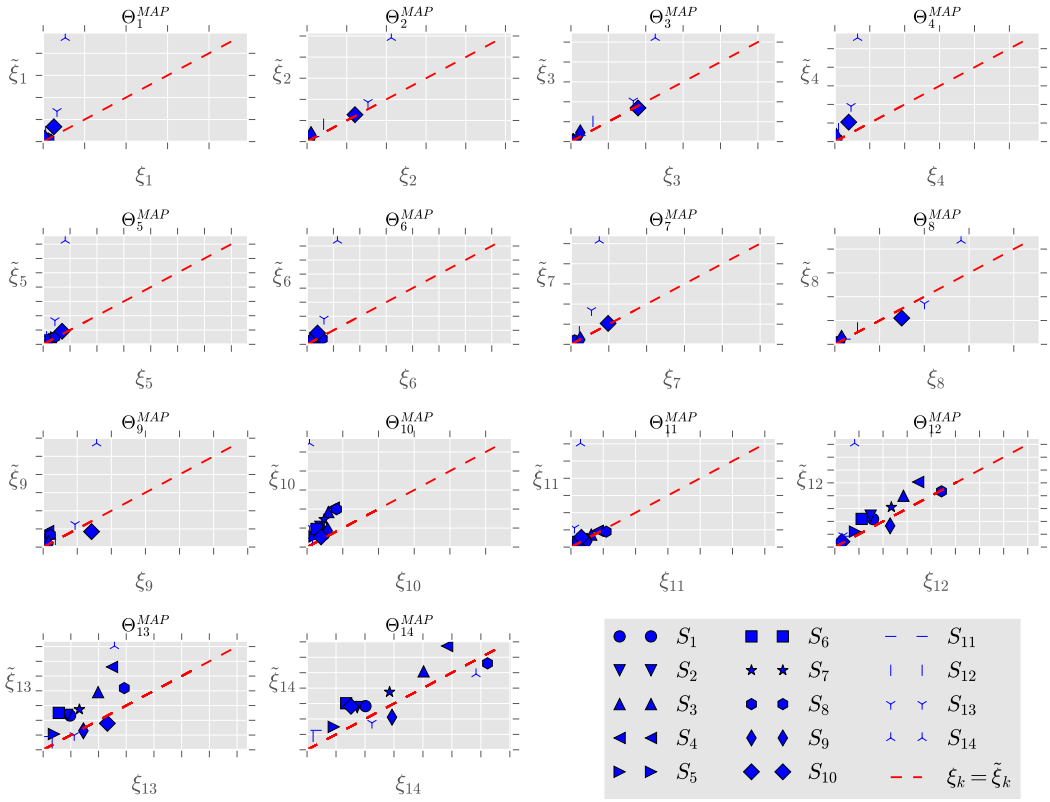


Fig. 3 The results of the cross validation study. In each of the k subplots ($k = 1, \cdots, 14$), the MAP estimates are fixed to those obtained from calibration scenario S_k. The MAP estimates are then used to compute both the true error $\tilde{\xi}_k$ and the modelled error ξ_k (11) when applied to prediction scenarios S_j, $j \in \{1, \cdots, K\} \setminus \{k\}$. The true errors $\tilde{\xi}_k$ were computed using the experimental u^+ data from [34].

V. Assessment of BMSA for selected flow cases

In this section, the MAP-based BMSA predictive methodology is assessed against selected flow cases, namely separated incompressible flow, and 3D transonic flow past a wing. The considered applications are rather different from our calibration scenarios, although they still involve boundary layers subject to variable pressure gradients, and exhibit features beyond the limits of applicability of linear RANS models, due to the strong non-equilibrium effects. As such, they represent severe tests for the proposed approach.

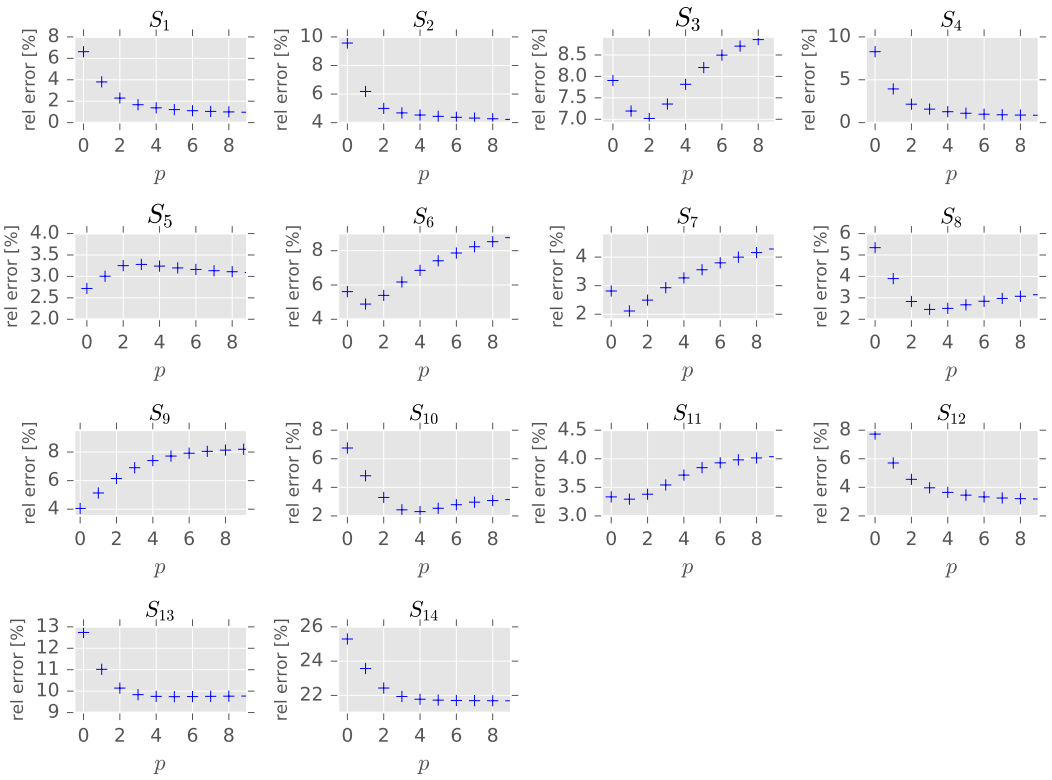


Fig. 4 The relative error in the BMSA prediction of u^+ for all K scenarios S_k.

A. Flow over periodic hills at $Re_H = 5600$

The flow over a series of periodically arranged hills deals both with flow separation on the curved surface of the hills and reattachment between the hills. The test case is known to be especially challenging for linear eddy-viscosity models, which are not able to predict the mean effect of the unsteady fluctuation of the separation and reattachment points correctly [35].

The periodicity of the hill-geometry is mimicked with periodic boundary conditions at the inlet and outlet and no-slip conditions at the walls. A volume forcing is applied to each cell, which maintains a bulk velocity of $U_b = 1.0$ between the hill's crest and the upper surface. The functional form of the lower surface is defined according to the ERCOFTAC test case description[**?**]. The considered Reynolds-number $Re_H = 5600$ based on the hill height H and bulk velocity between the hill's crest and the upper surface is maintained by a volume forcing. The CFD simulations are performed using OpenFOAM's `simpleFoam`-solver [36]. We use validation data of the mean-flow field from Breuer et al. for comparison [37]. Three turbulence models were used, Spalart-Allmaras,

Launder-Sharma $k-\epsilon$ and Wilcox (2006) $k-\omega$, with $y^+ \leq 1.0$. A mesh convergence study was conducted for each model using baseline coefficients resulting in meshes with sizes of 100×110 (Spalart-Allmaras and Launder-Sharma $k-\epsilon$) and of 150×140 (Wilcox (2006) $k-\omega$) cells in $x \times y$ direction respectively.

The baseline simulations for the three models show the expected behavior as reported in the literature (for a larger Re-number [35]): while both Spalart-Allmaras and the Wilcox (2006) $k-\omega$ over-predict the size of the recirculation zone, characterised by a zero velocity component close to the wall, the Launder-Sharma $k-\epsilon$ under-predicts this flow feature. For S_{14} the simulation using Wilcox (2006) $k-\omega$ didn't converge so that this scenario was excluded from the set, leading to only 13 scenarios being used for the BMSA predictions.

The BMSA expectation $\mathbb{E}[\Delta|\tilde{\mathrm{S}}, \mathbf{z}]$ of the stream-wise velocity, as shown in Figure 5, over-predicts the recirculation zone and gives similar velocity profiles for $x = 3.0$ and 4.0 close to the lower surface. Especially, for $x = 0.0$ the expectation doesn't capture the local maximum of the velocity close to $y/H = 1.0$, i.e. at the hills crest. Throughout the different locations the expectation shows differences compared to the DNS for the upper part of the velocity profile, which might be due to a compensation-effect caused by the fixed mass-flow.

The 99% confidence interval captures the DNS data for $x \geq 3$, but not the region close to the upper wall for $x \geq 5$ and the local velocity maxima in the area of the free-shear layer in the leeward region of $x = 0$ to 2.0. Interestingly for $x \leq 3.0$ when the range of the confidence interval shrinks locally in y/H-direction also the DNS is still inside. However, this pattern is not the same for every y/H-position, e.g. at $x = 2.0$ the expectation and the DNS match for $0.5 \leq y/H \leq 1$, but the confidence interval is large. Interestingly, in the reattachment region between the two hills, where the case is similar to a flat plate, the confidence interval is able to capture the DNS data consistently.

In general, the application of BMSA to this test case demonstrates the aforementioned fact that the used linear eddy-viscosity models suffer from restrictions, which inhibit the reproduction of effects due to e.g. streamline curvature as well as Reynolds-stress anisotropy. These are essential characteristics of this challenging test case [35]. Adding second-order closures to $\mathcal{M}$ might remedy this deficiency, at increased computational cost. However, the BMSA confidence interval captures

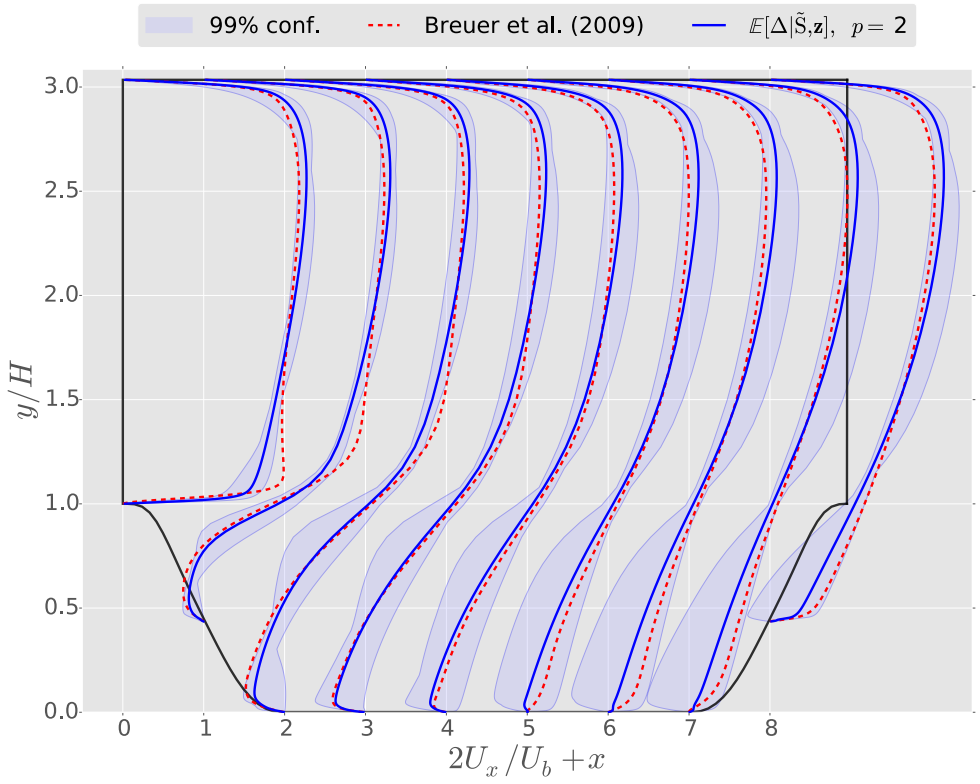


Fig. 5 BMSA for the periodic hill flow.

most of the modelling error showing robustness of the method even for this challenging test case.

B. Transonic flow over the ONERA M6 swept wing

The second application of our boundary-layer trained BMSA is the transonic flow over the ONERA M6 wing. This is a swept wing based on the ONERA D airfoil with no twist. The M6 wing is known to produce complex physics at transonic Mach numbers. See for instance the pressure distribution depicted in Figure 6. Due to the configuration of local supersonic regions [38], a λ-shock is formed where the double shock merges into a single one between $y/b = 0.8$ and $y/b = 0.9$.

A well-known CFD validation case (see e.g. [38, 40, 41]) involves the ONERA M6 wing at Mach 0.835, a Reynolds number (based on the free-stream conditions and the mean aerodynamic chord) of 11.72×10^6, and with an angle of attack of 3.06 degrees and 0.0 degrees side slip. The pressure and temperature at the farfield are given by $p_\infty = 315980\ [Pa]$ and $T_\infty = 255.5\ [K]$. The CFD calculations were performed using Ansys Fluent, using the pressure-based solver. More details on

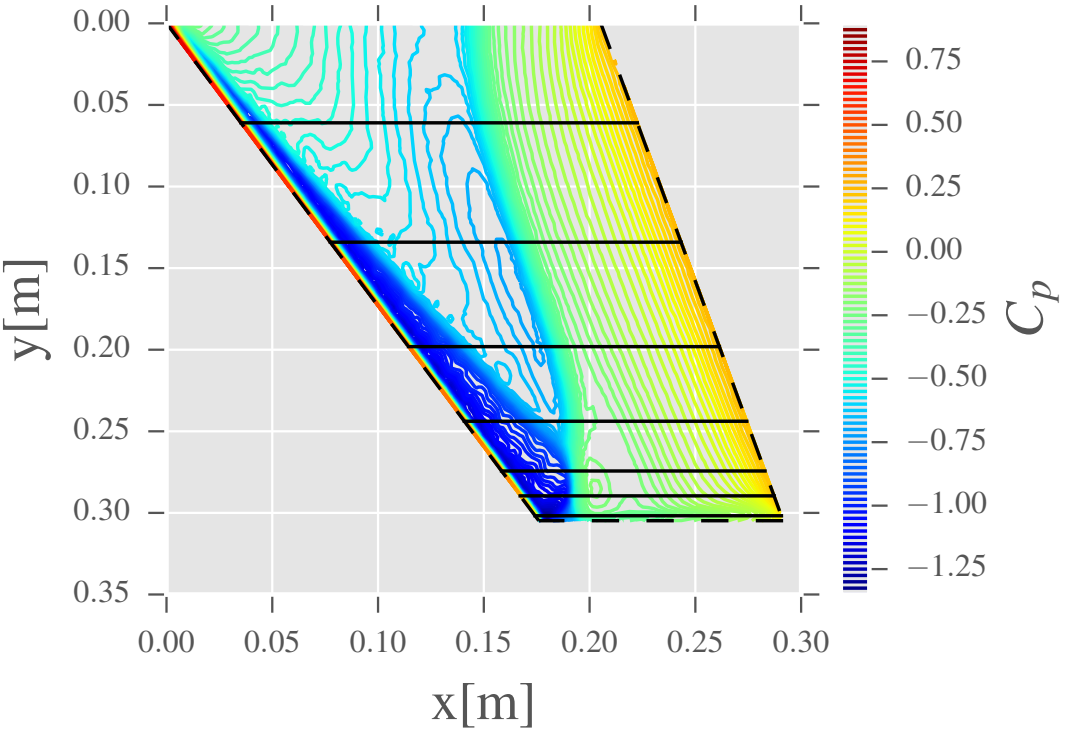
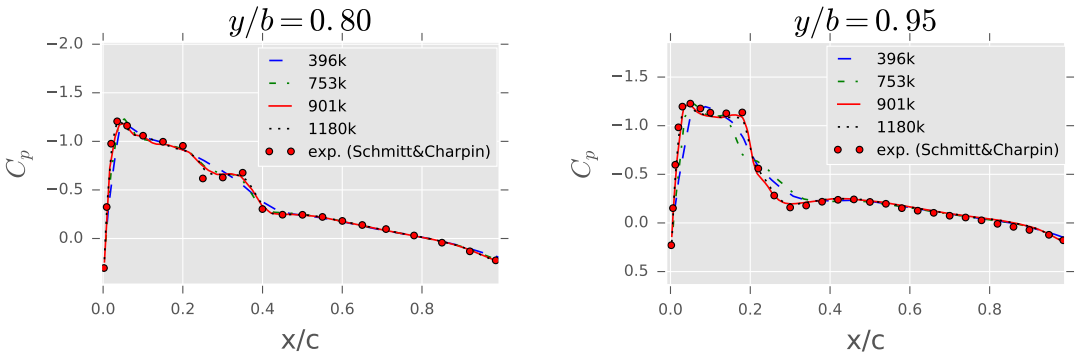


Fig. 6 A typical C_p distribution on the suction side of the ONERA M6 wing. The black lines correspond to the $y/b = [0.2, 0.44, 0.65, 0.80, 0.90, 0.95, 0.99]$ stations where C_p data is available from [39].

the numerical setup can be found in [42] and [43]. For these conditions experimental pressure data is available from Schmitt and Charpin [39] at several y/b stations. The error in the reported pressure data is $\Delta C_p = \pm 0.02$.

Before beginning the BMSA procedure, we performed a qualitative comparison of the C_p distributions using the MAP estimates of the 1400 case on 4 unstructured meshes of increasing size, namely $\{396, 753, 901, 1180\} \times 10^3$ elements. The results at $z/b = 0.80$ and $z/b = 0.95$ are shown in Figure 7. Note that the coarsest mesh smears the double shock at $y/b = 0.80$, and both mesh 1 and 2 smear the shock at $y/b = 0.95$. Between the two finest meshes we observed little variation, and selected therefore mesh 3 to perform the sampling with MAP estimates.

Our model set for this case consists of the $k - \varepsilon$ and SA model. We computed the mean C_p prediction (15) and the standard deviation based on (16) for the entire suction side of the M6



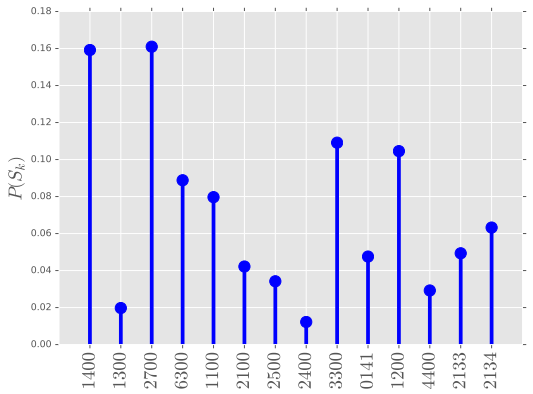
(a) The C_p distributions at $z/b = 0.8$.

(b) The C_p distributions at $z/b = 0.95$.

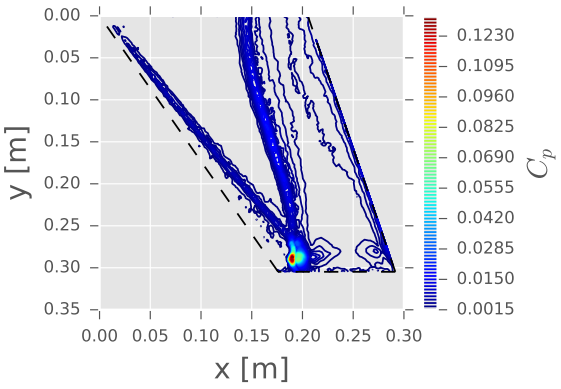
Fig. 7 The C_p distributions of the 4 meshes at two spanwise locations. All computations were done using the MAP estimates of calibration case 1400.

wing. The mean is is very similar to the results of Figure 6. More interesting is the uncertainty in the prediction, depicted in Figure 8(b). From this standard deviation is becomes clear that the uncertainty of the RANS prediction is localized in the region of the λ-shock, especially where the two shocks merge.

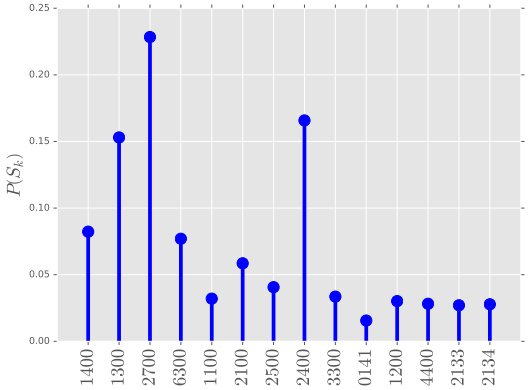
The corresponding $p(S_k)$ distribution is shown in Figure 8(a). Note that the distribution favours the results of MAP estimates from a zero (1400) and a favorable (2700) calibration scenario, as well as two adverse cases (3300 and 1200). This might be explained by the fact that the transonic flow experiences a wide range of pressure gradients, at different locations of the wing. Perhaps the weights $p(S_k)$ should ideally reflect the inhomogeneous nature of the uncertainty, rather than computing a scalar $p(S_k)$ for each k, and assuming it is applicable throughout the entire flow field. To investigate, we compute the BMSA prediction along different y/b stations of the wing. For each station, taking the local C_p distribution as QoI, we recompute $p(S_k)$. This allows us to build up a picture of how $p(S_k)$ varies from the wing root to the tip. A selection of results, at y/b stations where we have experimental validation data, are displayed in Figures 8(c)-8(f). First notice that the amount of uncertainty at $y/b = 0.44$ is very low, a result that could already be inferred from Figure 8(b). The $p(S_k)$ distribution at this station displays a preference for mostly those results computed using zero or favourable pressure-gradient MAP estimates. Note however, that the final C_p distribution is insensitive to $\mathbb{P}(S_k)$ in cases of very low uncertainty, seeing that all S_k



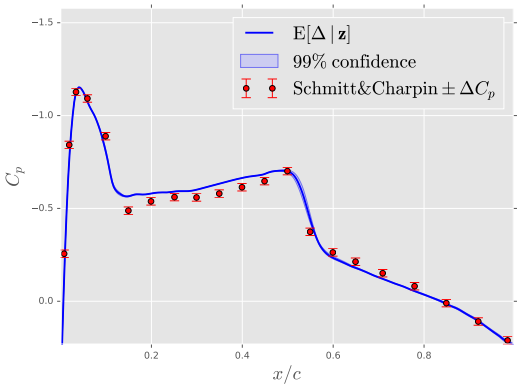
(a) The $p(S_k)$ distribution for the entire wing.



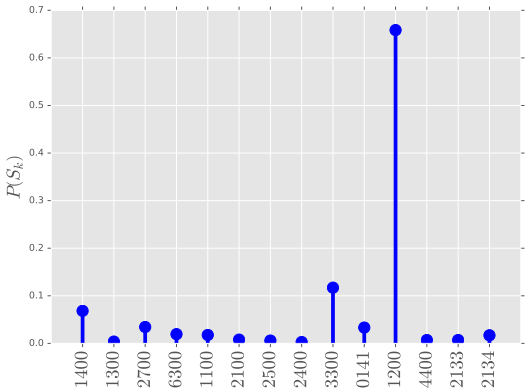
(b) The standard deviation of (16).



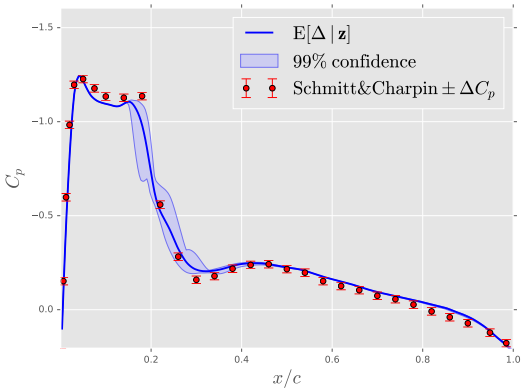
(c) The $\mathbb{P}(S_k)$ distribution at $y/b = 0.44$.



(d) The BMSA prediction at $y/b = 0.44$ plus 99% confidence region.



(e) The $\mathbb{P}(S_k)$ distribution at $y/b = 0.95$.



(f) The BMSA prediction at $y/b = 0.95$ plus 99% confidence region.

Fig. 8 The results for the ONERA M6 wing.

yield very similar predictions. Other computed $p(S_k)$ distributions are comparable, up until the neighborhood of $y/b = 0.95$, where the results are quite different. Here, a relatively large amount of uncertainty is present, accompanied by a different $p(S_k)$ distribution. There is a preference for the pressure computed with MAP estimates coming from a strongly-adverse calibration scenario (namely scenario 1200). These results suggest that incorporating a spatial dependence in $\mathbb{P}(S_k)$ does indeed reflect the local flow physics better. Specifying $p(S_k)$ locally through (11) assumes that there is no spatial correlation.

VI. Conclusion

A Bayesian Model-Scenario Averaging methodology was applied to estimate turbulence modelling uncertainties in RANS computations. Specifically, the goal of this paper was to assess the predictive capability of a BMSA model, trained on experimental velocity profiles for a set of flat plate boundary layers, when predicting other, more computationally expensive flow problems. We considered two severe validation cases, namely the separated flow over periodic hills, and a shocked transonic flow over a 3D wing. These contain physics that were not present during the training phase, such as compressibility effects and separation. The full BMSA approach requires the propagation of a set of posterior distributions, which proves to be computationally intractable for industrial-relevant test cases. We therefore proposed to approximate the posterior distributions with Dirac δ functions centered at the *Maximum A Posteriori* (MAP) estimates of the posterior distributions. This is the most drastic cost reduction possible, with just one code evaluation per closure model - calibration scenario pair.

The main advantage of an approach such as BMSA, is that one can make predictions with quantified uncertainty. Such information is not available with the baseline models, and provides insight into where, and to what extend, the closure models fail to be trustworthy. The confidence region generated by BMSA thus provides meaningful bounds on the QoI. In our selected cases we observed an overlap between the confidence intervals and the reference data, especially when the local flow physics resembled the attached boundary-layers of the calibration phase. Where there was no overlap the discrepancy between the propagated model-scenario pairs was small, which might be

remedied by including additional closure models in the set, e.g. nonlinear eddy viscosity models.

Acknowledgment

This research was partially supported by the European Union's Seventh Framework Programme (2007-2013) under grant agreement 605036, UMRIDA project and by French ANR under grant agreement 11-MONU-0008.

Bibliography

[1] Wilcox, D. C., Turbulence Modeling for CFD, DCW Industries, 2006.

[2] Cheung, S. H., Oliver, T. A., Prudencio, E. E., Prudhomme, S., and Moser, R. D., "Bayesian uncertainty analysis with applications to turbulence modeling," Reliability Engineering & System Safety, Vol. 96, No. 9, 2011, pp. 1137–1149.

[3] Guillas, S., Glover, N., and Malki-Epshtein, L., "Bayesian calibration of the constants of the k–ε turbulence model for a CFD model of street canyon flow," Computer Methods in Applied Mechanics and Engineering, Vol. 279, 2014, pp. 536–553.

[4] Kato, H. and Obayashi, S., "Approach for uncertainty of turbulence modeling based on data assimilation technique," Computers & Fluids, Vol. 85, 2013, pp. 2–7.

[5] Edeling, W., Cinnella, P., Dwight, R., and Bijl, H., "Bayesian estimates of parameter variability in the k-ε turbulence model," Journal of Computational Physics, Vol. 258, 2014, pp. 73–94.

[6] Launder, B. and Sharma, B., "Application of the energy-dissipation model of turbulence to the calculation of flow near a spinning disc," Letters in heat and mass transfer, Vol. 1, No. 2, 1974, pp. 131–137.

[7] Zhang, Z., Zhang, W., Zhai, Z., and Chen, Q., "Evaluation of various turbulence models in predicting airflow and turbulence in enclosed environments by CFD: Part 2 Comparison with experimental data from literature," Hvac&R Research, Vol. 13, No. 6, 2007, pp. 871–886.

[8] Harrison, K. and Bogard, D., "Comparison of RANS turbulence models for prediction of film cooling performance," ASME Turbo Expo 2008: Power for Land, Sea, and Air, American Society of Mechanical Engineers, 2008, pp. 1187–1196.

[9] Dow, E. and Wang, Q., "Quantification of structural uncertainties in the k–ω turbulence model," AIAA Paper, Vol. 1762, 2011, pp. 2011.

[10] Kennedy, M. and O'Hagan, A., "Bayesian calibration of computer models," Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 63, No. 3, 2001, pp. 425–464.

[11] Singh, A. and Duraisamy, K., "Using field inversion to quantify functional errors in turbulence closures," Physics of Fluids, Vol. 28, No. 4, 2016, pp. 045110.

[12] Xiao, H., Wu, J., Wang, J., Sun, R., and Roy, C., "Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach," Journal of Computational Physics, Vol. 324, 2016, pp. 115–136.

[13] Poroseva, S., Colmenares, J., and Murman, S., "RANS simulations of a channel flow with a new velocity/pressure-gradient model," AIAA Paper, Vol. 3067, 2015, pp. 2015.

[14] Poroseva, S., Colmenares F., J. D., and Murman, S., "On the accuracy of RANS simulations with DNS data," Physics of Fluids, Vol. 28, No. 11, 2016, pp. 115102.

[15] Poroseva, S. V., Hussaini, M. Y., and Woodruff, S. L., "Improving the predictive capability of turbulence models using evidence theory," AIAA journal, Vol. 44, No. 6, 2006, pp. 1220.

[16] Poroseva, S., Lay, N., and Hussaini, M., "Multimodel approach based on evidence theory for forecasting hurricane/typhoon tracks: Further improvements," Monthly Weather Review, Vol. 138, No. 2, 2009, pp. 405–420.

[17] Diomede, T., Davolio, S., Marsigli, C., Miglietta, M., Moscatello, A., Papetti, P., and Paccagnella, T. e. a., "Discharge prediction based on multi-model precipitation forecasts," Meteorology and atmospheric physics, Vol. 101, No. 3, 2008, pp. 245–265.

[18] Duan, Q., Ajami, N., Gao, X., and Sorooshian, S., "Multi-model ensemble hydrologic prediction using Bayesian model averaging," Advances in Water Resources, Vol. 30, No. 5, 2007, pp. 1371–1386.

[19] Tebaldi, C. and Knutti, R., "The use of the multi-model ensemble in probabilistic climate projections," Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, Vol. 365, No. 1857, 2007, pp. 2053–2075.

[20] Draper, D., "Assessment and propagation of model uncertainty," Journal of the Royal Statistical Society. Series B (Methodological), 1995, pp. 45–97.

[21] Rojas, R., Kahunde, S., Peeters, L., Batelaan, O., Feyen, L., and Dassargues, A., "Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling," Journal of Hydrology, Vol. 394, No. 3, 2010, pp. 416–435.

[22] Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C., "Bayesian model averaging: a tutorial," Statistical science, 1999, pp. 382–401.

[23] Edeling, W., Cinnella, P., and Dwight, R., "Predictive RANS simulations via bayesian model-scenario averaging," Journal of Computational Physics, Vol. 275, 2014, pp. 65–91.

[24] Oliver, T. and Moser, R., "Bayesian uncertainty quantification applied to RANS turbulence models,"

Journal of Physics: Conference Series, Vol. 318, IOP Publishing, 2011, p. 042032.

[25] Margheri, L., Meldi, M., Salvetti, M., and Sagaut, P., "Epistemic uncertainties in RANS model free coefficients," Computers & Fluids, Vol. 102, 2014, pp. 315–335.

[26] Papadimitriou, D. and Papadimitriou, C., "Bayesian uncertainty quantification of turbulence models based on high-order adjoint," Computers & Fluids, Vol. 120, 2015, pp. 82–97.

[27] Hastings, W. K., "Monte Carlo sampling methods using Markov chains and their applications," Biometrika, Vol. 57, No. 1, 1970, pp. 97–109.

[28] Eldred, M. and Burkardt, J., "Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification," AIAA paper, Vol. 976, No. 2009, 2009, pp. 1–20.

[29] Cheung, S. and Beck, J., "Calculation of posterior probabilities for Bayesian model class assessment and averaging from posterior samples based on dynamic system data," Computer-Aided Civil and Infrastructure Engineering, Vol. 25, No. 5, 2010, pp. 304–321.

[30] Edeling, W. N., Cinnella, P., and Dwight, R., "MCMC traces of posterior distributions of closure coefficients for 6 turbulence models [online database]," `www.researchgate.net/publication/308609583_MCMC_traces_of_posterior_distributions_of_closure_coefficients_for_6_turbulence_models_ASCII_format`, 2016, [Retrieved 25-September-2016].

[31] Beck, J., "Bayesian system identification based on probability logic," Structural Control and Health Monitoring, Vol. 17, No. 7, 2010, pp. 825–847.

[32] Chib, S. and Jeliazkov, I., "Marginal likelihood from the Metropolis–Hastings output," Journal of the American Statistical Association, Vol. 96, No. 453, 2001, pp. 270–281.

[33] Edeling, W., Dwight, R., and Cinnella, P., "Simplex-stochastic collocation method with improved scalability," Journal of Computational Physics, Vol. 310, 2016, pp. 301–328.

[34] Coles, D. and Hirst, E., "Computation of turbulent boundary layers," Proceedings of AFOSR-IFP Stanford Conference, Vol. 2, 1968.

[35] Jakirlic, S., "Extended excerpt related to the test case: Flow over a periodical arrangement of 2D hills," Tech. Rep. June, 2012.

[36] OpenFOAM, OpenFOAM User Guide, OpenFOAM Foundation, February 2014.

[37] Breuer, M., Peller, N., Rapp, C., and Manhart, M., "Flow over periodic hills - Numerical and experimental study in a wide range of Reynolds numbers," Computers and Fluids, Vol. 38, No. 2, 2009, pp. 433–457.

[38] Kuzmin, A., "On the lambda-shock formation on ONERA M6 wing," International Journal of Applied Engineering Research, Vol. 9, No. 20, 2014, pp. 7029–7038.

[39] Schmitt, V. and Charpin, F., "Pressure distributions on the ONERA-M6-Wing at transonic Mach numbers," Experimental data base for computer program assessment, Vol. 4, 1979.

[40] Lerat, A., Cinnella, P., Michel, B., and Falissard, F., "High-order residual-based compact schemes for aerodynamics and aeroacoustics," Computers & Fluids, Vol. 61, 2012, pp. 31–38.

[41] Huang, Y., Cinnella, P., and Lerat, A., "A third-order accurate centered scheme for turbulent compressible flow calculations in aerodynamics," Numerical Methods for Fluid Dynamics, Vol. 6, 1998, pp. 355–361.

[42] NASA, "NPARC Alliance Validation Archive, ONERA M6 wing," `https://www.grc.nasa.gov/www/wind/valid/m6wing/m6wing.html`, 2017, [Retrieved 2-October-2017].

[43] Cornell-University, "FLUENT - 3D Transonic Flow Over a Wing," `https://confluence.cornell.edu/display/SIMULATION/FLUENT+-+3D+Transonic+Flow+Over+a+Wing`, 2017, [Retrieved 9-October-2017].

[44] Kline, S., "The 1980-81 AFOSR-HTTM-Stanford conference on complex turbulent flows: Comparison of computation and experiment," Objectives, evaluation of data, specifications of test cases, discussion and position papers, Vol. 1, 1981.

[45] Underlying flow regime (UFR) 3-30, 2D Periodic Hill Flow: `http://qnet-ercoftac.cfms.org.uk`

APPENDIX A: TABLES OF MAP ESTIMATES, POSTERIOR MODEL PROBABILITIES, AND CALIBRATION CASES

Here we report the MAP estimates of the closure coefficients for the three turbulence models of Section 2 (tables 1-3, as well as the corresponding posterior model probabilities (Tables 4,5). We also give a description of the calibration cases.

In the case of the $k-\varepsilon$ mode we use (A1)-(A2) to fix the value of σ_ε and $C_{\varepsilon 1}$, where we set $\mathcal{P}/\varepsilon = 2.09$ for the latter, see [5] for a discussion. We use other well-known algebraic relations for the $k-\omega$ and SA models in (A3) and (A4) respectively [1].

$$\sigma_\varepsilon = \frac{\kappa^2}{C_\mu^{1/2}\left(C_{\varepsilon 2} - C_{\varepsilon 1}\right)}. \tag{A1}$$

$$\left(\frac{\mathcal{P}}{\varepsilon}\right) = \frac{C_{\varepsilon 2} - 1}{C_{\varepsilon 1} - 1}. \tag{A2}$$

$$\alpha = \frac{\beta}{\beta^*} - \frac{\kappa^2}{2\sqrt{\beta^*}}. \tag{A3}$$

$$C_{w1} = \frac{C_{b1}}{\kappa^2} + \frac{1 + C_{b2}}{\sigma}. \tag{A4}$$

Table 1 MAP estimates of the $k-\varepsilon$ model for the different calibration scenarios. The nominal values are depicted on the first row.

Scenario	$C_{\varepsilon 2}$	C_μ	σ_k	κ
nominal	1.92	0.09	1.0	0.41
S_1	2.0429	0.0646	1.4032	0.428
S_2	2.0556	0.0726	1.1737	0.4551
S_3	1.911	0.0655	0.7082	0.451
S_4	2.0634	0.0715	1.4205	0.4621
S_5	1.9334	0.0618	1.0796	0.3604
S_6	2.0456	0.0654	1.4156	0.3865
S_7	1.9454	0.0604	1.0314	0.4365
S_8	1.9919	0.0841	1.2869	0.4596
S_9	1.8205	0.0699	0.9999	0.36
S_{10}	1.8407	0.06	1.4067	0.3559
S_{11}	1.9373	0.0599	1.1435	0.3282
S_{12}	1.8258	0.0625	1.2562	0.3259
S_{13}	1.6687	0.0642	0.8093	0.2964
S_{14}	1.604	0.0578	0.7327	0.2902

Table 2 MAP estimates of the Wilcox 2006 $k-\omega$ model for the different calibration scenarios. The nominal values are depicted on the first row.

Scenario	α	β_0	β^*	σ	σ^*
nominal	0.52	0.0708	0.09	0.5	0.6
S_1	0.4557	0.0668	0.0768	0.4953	0.4107
S_2	0.4369	0.0728	0.0848	0.5186	0.5177
S_3	0.4383	0.0687	0.0828	0.5017	0.4107
S_4	0.4703	0.0737	0.0909	0.4945	0.5231
S_5	0.5906	0.0665	0.0858	0.5089	0.4094
S_6	0.6045	0.0743	0.0895	0.5057	0.5825
S_7	0.4786	0.0644	0.0784	0.4985	0.4067
S_8	0.4306	0.0738	0.0744	0.6887	0.4074
S_9	0.4946	0.0629	0.0744	0.6927	0.406
S_{10}	0.4682	0.0647	0.0844	0.5234	0.4087
S_{11}	0.592	0.069	0.0979	0.4961	0.4167
S_{12}	0.6129	0.0696	0.0904	0.5073	0.4268
S_{13}	0.592	0.0649	0.0887	0.5106	0.4147
S_{14}	0.6087	0.0644	0.1043	0.4985	0.5853

Table 3 MAP estimates of the SA model for the different calibration scenarios. The nominal values are depicted on the first row.

Scenario	C_{b1}	C_{b2}	C_{v1}	σ	C_{w2}	C_{w3}
nominal	0.1355	0.622	7.1	0.67	0.3	2.0
S_1	0.1354	0.5377	7.1833	0.8009	0.3664	2.0619
S_2	0.1185	0.6152	6.881	0.602	0.349	1.8144
S_3	0.1264	0.6302	7.1017	0.5711	0.2702	2.4675
S_4	0.144	0.5629	7.1384	0.7964	0.3731	2.4636
S_5	0.1299	0.7561	7.6842	0.746	0.3682	1.7774
S_6	0.1305	0.5678	6.6114	0.658	0.2931	1.8615
S_7	0.1606	0.6559	9.0895	0.8144	0.3153	2.3111
S_8	0.1048	0.519	7.7523	0.6889	0.2468	2.047
S_9	0.1353	0.5042	8.5117	0.5256	0.3033	2.054
S_{10}	0.1693	0.6837	9.0427	0.7917	0.2384	2.0496
S_{11}	0.1286	0.5441	6.8871	0.7676	0.3671	1.5532
S_{12}	0.1214	0.7211	7.022	0.5924	0.257	1.641
S_{13}	0.1062	0.6327	7.9545	0.6648	0.2601	1.5889
S_{14}	0.0983	0.5603	6.5696	0.5643	0.2938	2.1707

Table 4 The posterior model probabilities $p(M_i|\boldsymbol{z}_k)$ when $\mathcal{M}$ consists of the $k-\varepsilon$ and SA model. We utilized a uniform prior $p(M_i)$.

Scenario	$k-\varepsilon$	SA
S_1	6.54763e-02	9.34524e-01
S_2	1.31256e-01	8.68744e-01
S_3	9.98800e-01	1.19985e-03
S_4	2.09486e-01	7.90514e-01
S_5	9.95836e-01	4.16372e-03
S_6	2.36294e-01	7.63706e-01
S_7	1.00000e+00	4.51629e-07
S_8	1.00000e+00	4.34707e-11
S_9	1.00000e+00	0.00000e+00
S_{10}	1.00000e+00	0.00000e+00
S_{11}	9.99867e-01	1.32751e-04
S_{12}	9.98324e-01	1.67614e-03
S_{13}	0.00000e+00	1.00000e+00
S_{14}	0.00000e+00	1.00000e+00

Table 5 The posterior model probabilities $p(M_i|\boldsymbol{z}_k)$ when $\mathcal{M}$ consists of the $k-\varepsilon$, $k-\omega$ and SA models. We utilized a uniform prior $p(M_i)$

Scenario	$k-\varepsilon$	$k-\omega$	SA
S_1	1.04824e-02	6.76163e-03	9.82756e-01
S_2	5.15244e-02	8.73518e-02	8.61124e-01
S_3	4.11104e-01	5.88478e-01	4.18024e-04
S_4	7.15142e-02	4.21388e-02	8.86347e-01
S_5	9.73137e-01	2.57038e-02	1.15898e-03
S_6	1.04312e-01	1.26575e-03	8.94423e-01
S_7	1.00000e+00	1.94074e-10	2.35604e-08
S_8	9.99999e-01	9.14473e-07	1.98968e-11
S_9	9.98710e-01	1.29035e-03	2.34332e-21
S_{10}	1.00000e+00	0.00000e+00	4.95032e-13
S_{11}	3.78159e-01	6.21680e-01	1.61533e-04
S_{12}	9.42438e-01	5.36331e-02	3.92870e-03
S_{13}	0.00000e+00	3.69311e-01	6.30689e-01
S_{14}	0.00000e+00	9.74174e-01	2.58255e-02

Table 6 Flow descriptions of training database. Identification numbers are taken from the source [34]. The 'Type' column describes the pressure gradient, which ranges from favourable to strongly adverse.

Scenario	Identification	Type	Description
S_1	1400	Zero	Equilibrium boundary layer at constant pressure
S_2	1300	Fav	Near-equilibrium boundary layer in moderate negative pressure gradient
S_3	2700	Fav	Equilibrium boundary layer in mild negative pressure gradient
S_4	6300	Fav	Near-equilibrium boundary layer growing beneath potential flow on model spillway
S_5	1100	Mild adv	Boundary layer in diverging channel
S_6	2100	Div	Boundary layer on large airfoil-like body; pressure gradient mildly negative
S_7	2500	Mild adv	Equilibrium boundary layer in mild positive pressure gradient
S_8	2400	Div	Initial equilibrium boundary layer in moderate positive pressure gradient; pressure gradient abruptly decreases to zero, and flow relaxes to new equilibrium
S_9	3300	Mod adv	Boundary layer, initially at constant pressure, developing into equilibrium flow in moderate positive pressure gradient
S_{10}	0141	Str adv	Boundary-layer with strong adverse pressure gradient, source [44]
S_{11}	1200	Str adv	Boundary layer in diverging channel with eventual separation
S_{12}	4400	Str adv	Boundary layer in strong positive pressure gradient
S_{13}	2133	Str adv	Boundary layer on large airfoil-like body; pressure gradient strongly positive
S_{14}	2134	Str adv	Boundary layer on large airfoil-like body; pressure gradient strongly positive, close to separation