



### **Science Arts & Métiers (SAM)**

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>  
Handle ID: <http://hdl.handle.net/10985/10035>

#### **To cite this version :**

Paola CINNELLA - Predictive RANS simulations via Bayesian Model-Scenario Averaging -  
Journal of Computational Physics - Vol. 275, p.65-91 - 2014

Any correspondence concerning this service should be sent to the repository

Administrator : [scienceouverte@ensam.eu](mailto:scienceouverte@ensam.eu)



# Predictive RANS simulations via Bayesian Model-Scenario Averaging

W.N. Edeling<sup>a,b</sup>, P. Cinnella<sup>a</sup>, R.P. Dwight<sup>b,\*</sup>

<sup>a</sup>*Arts et Métiers ParisTech, DynFluid laboratory, 151 Boulevard de l'Hospital, 75013 Paris, France*

<sup>b</sup>*Delft University of Technology, Faculty of Aerospace Engineering, Kluyverweg 2, Delft, The Netherlands*

---

## Abstract

The turbulence closure model is the dominant source of error in most Reynolds Averaged Navier-Stokes simulations, yet no reliable estimators for this error component currently exist. Here we develop a stochastic, *a posteriori* error estimate, calibrated to specific classes of flow. It is based on variability in model closure coefficients across multiple flow scenarios, for multiple closure models. The variability is estimated using Bayesian calibration against experimental data for each scenario, and Bayesian Model-Scenario Averaging (BMSA) is used to collate the resulting posteriors, to obtain a stochastic estimate of a Quantity of Interest (QoI) in an unmeasured (prediction) scenario. The scenario probabilities in BMSA are chosen using a sensor which automatically weights those scenarios in the calibration set which are similar to the prediction scenario. The methodology is applied to the class of turbulent boundary-layers subject to various pressure gradients. For all considered prediction scenarios the standard-deviation of the stochastic estimate is consistent with the measurement ground truth. Furthermore, the mean of the estimate is more consistently accurate than the individual model predictions.

**Keywords:** Bayesian Model Averaging; Bayesian Model-Scenario Averaging; RANS models, model inadequacy; uncertainty quantification; calibration; boundary-layers; error estimation

---

## 1. Introduction

Numerical predictions of continuum mechanics are affected by several sources of error: discretization error, parametric uncertainty, and physical modelling error, are the three most significant. Yet only the latter can not be estimated and controlled with standard techniques. As such it represents the bottleneck to a robust predictive capability in many fields.

In industrial Computational Fluid Dynamics (CFD) simulations, the Reynolds-Averaged Navier-Stokes (RANS) equations are the workhorse tool for modelling turbulent flow

---

\*Corresponding author

*Email addresses:* W.N.Edeling@tudelft.nl (W.N. Edeling), P.Cinnella@ensam.eu (P. Cinnella), R.P.Dwight@tudelft.nl (R.P. Dwight)

*Preprint submitted to Elsevier*

*July 3, 2014*

fields. The equations require a turbulence closure model, of which there are many [36]; each of which requires empirical closure coefficients parameterising modelling assumptions. Standard practice is to calibrate these coefficients against simple flow configurations (generally, incompressible homogeneous isotropic turbulence, free shear layers, and zero pressure gradient flat plate boundary layers [25]) using available experimental data. Even so there is no consensus on the best coefficient values for most models, as evidenced by the wide range of coefficients seen in the literature. As a consequence model predictions can differ significantly [34, 38].

In the Bayesian framework, model error is typically accounted for with a stochastic term added to (or multiplying) the output of the simulation. For instance Kennedy and O’Hagan [13] define “model inadequacy” as the discrepancy that remains between the true process and the prediction of the simulation code evaluated using the best-fit parameter values. We believe however that modelling this discrepancy directly is inappropriate when the simulation is an approximation of continuum mechanics, which is characterized by perfectly known conservation laws (e.g. conservation of momentum), containing imperfect, subordinate, empirical models, such as turbulence closure models in RANS. In this context, it makes much more sense to include stochastic modelling of inadequacy at the level of the empirical model, not at the level of the simulation code output. This choice has the ancillary benefit that predictions of quantities of other type than the calibration data can be made – and in other geometries – while still incorporating estimates of model inadequacy. This is not possible in the original Kennedy and O’Hagan framework.

There are two natural ways to incorporate stochastic model inadequacy into turbulence closure models: via the Reynolds-stress tensor [9], and via the closure coefficients [8]. It is the latter approach that we will use; and our goal is to quantify model error in the predictions of RANS codes.

Our novel approach is based on Bayesian statistics [10], and can be summarized as follows: we first choose a *class* of flows for which we wish to make reliable predictions with quantified model error, in this work the class is flat-plate boundary-layers at a variety of pressure gradients. We select a number of examples of this class for which we have reliable experimental data; these flows form our *calibration scenarios*. Finally we select a set of turbulence closure models, that are both expected to perform acceptably on the flow-class, but which are also heterogeneous in their modelling approaches, in this work we use  $k - \varepsilon$ ,  $k - \omega$ , Spalart-Allmaras, Baldwin-Lomax and stress- $\omega$  models [36]. For each model and each scenario we perform Bayesian calibration to obtain posteriors on closure coefficients [3] as well as on model probabilities, i.e. probabilities of recovering the data by using a given model with the calibrated coefficients. For all models a strong dependence of coefficients on scenario is observed. This variation is interpreted as the extent to which the coefficients must be adjusted to match scenarios in the calibration set; informally a kind of coefficient uncertainty. Precisely, to make predictions for some Quantity of Interest (QoI) in some unmeasured scenario, we use Bayesian Model Averaging (BMA) [12] to collate the individual coefficient posteriors. A key step is propagating these posteriors through the simulation code for the prediction scenario. By using the freedom inherent in BMA to define prior scenario probabilities, we automatically assign a higher weight to those calibration scenarios which allow recovering consistent predictions for a new scenario with all alternative models considered; these are often found to be similar to the prediction scenario. The prediction is summarized by the mean and

variance of the posterior predictive distribution [6, 27]. Because in our framework the variation between scenarios is at least as important as the variation between models, we speak of *Bayesian Model-Scenario Averaging* (BMSA), rather than BMA.

There is very little work in the literature on numerical methods for estimating RANS closure model error. The authors of [22] perform calibration with multiple turbulence models and model-inadequacy terms on the simulation output, but they do not combine them to make reliable predictions. Other authors have applied Bayesian approaches to problems in aeroelasticity, but rely on expert judgement [26]. Recently Emory et al. [9] proposed a physically-motivated method for perturbing the Reynolds stress tensor, but again the uncertainty is determined by physically possible states, and not informed by experimental data. Finally in [8] the present authors used Bayesian statistics to estimate the model error of the  $k - \varepsilon$  turbulence model [15], via multiple posterior distributions of the closure coefficients. However our use of p-boxes limited the flexibility of the predictions, leading to model error estimates that contained the true solution, but which were impractically large.

This work proposes one of the first error estimates for turbulence closure models. However the computational cost of the method investigated here is extremely high, if it would be applied to complex flow topologies. Our primary goal in this article is to investigate the potential of our framework for estimating model error, in the absence of simplifying approximations. Success at this stage motivates investigation of both numerical and statistical techniques for making the framework computationally feasible for expensive simulation codes. As a proof of concept, one such a technique is already proposed in the current article.

The paper is laid out as follows: the four turbulence models are described in Section 2, and the scenarios together with experimental data in Section 3. Section 4.1 lays out our calibration framework, Section 4.2 prediction with Bayesian Model-Scenario Averaging, and Section 4.3 automatic scenario weighting. Section 5 discusses results.

## 2. RANS Turbulence Modelling

### 2.1. Governing equations

Directly solving the Navier-Stokes equations for a turbulent flow at a high Reynolds number is computationally intractable, since the range of scales present increases rapidly with the Reynolds number [25]. Capturing all these scales of motion in a Direct Numerical Simulation (DNS) would require a very fine grid and a very small time step such that the computational resources required are beyond the capability of currently available computing clusters. For this reason, DNS is presently a research tool limited to low-Reynolds number flows, not a brute-force option for engineering problems. For the latter problems often knowledge of the mean-flow only is sufficient, hence a common option is to employ the Reynolds-Averaged Navier-Stokes (RANS) equations coupled with an eddy-viscosity closure model. The governing RANS equations, once a particular eddy-viscosity closure model is adopted, for an incompressible boundary-layer flow are given

by

$$\frac{\partial \bar{u}_1}{\partial x_1} + \frac{\partial \bar{u}_2}{\partial x_2} = 0, \quad (1)$$

$$\frac{\partial \bar{u}}{\partial t} + \bar{u}_1 \frac{\partial \bar{u}_1}{\partial x_1} + \bar{u}_2 \frac{\partial \bar{u}_1}{\partial x_2} = -\frac{1}{\rho} \frac{\partial \bar{p}}{\partial x_2} + \frac{\partial}{\partial x_2} \left[ (\nu + \nu_T) \frac{\partial \bar{u}_1}{\partial x_2} \right], \quad (2)$$

where  $\rho$  is the constant density,  $\bar{u}_i$  is the mean velocity in  $x_i$  direction and  $\nu$  is the kinematic viscosity. The eddy viscosity  $\nu_T$  is meant to represent the effect of turbulent fluctuations on the mean flow, and is calculated by means of a turbulence model.

## 2.2. Turbulence models

The models we use in our analysis are given below. For conciseness, we do not give full details on these models. Rather, we refer to [36] for a more complete description of their mathematical structure.

### 2.2.1. The $k - \varepsilon$ model with Launder-Sharma damping functions

The  $k - \varepsilon$  model is a so-called two-equation model in which the eddy viscosity is calculated as  $\nu_T = C_\mu k / \varepsilon$ . Here,  $k$  is the turbulent kinetic energy and  $\varepsilon$  the turbulent dissipation, each of which has its own transport equation that has to be solved in order to compute  $\nu_T$ . These equations, and the expression for  $\nu_T$  itself, contain the following empirically determined closure coefficients:  $C_\mu$ ,  $C_{\varepsilon 1}$ ,  $C_{\varepsilon 2}$ ,  $\sigma_k$  and  $\sigma_\varepsilon$ . The values of these coefficients are traditionally determined by calibration on simple flow problems, and different versions of the  $k - \varepsilon$  model assume different values [36]. By introducing empirical damping functions which improve the predictive capability close to a wall, the Launder-Sharma  $k - \varepsilon$  model [15] is arrived at.

A common fundamental flow problem is fully developed channel flow. The resulting simplified governing equations lead to the following constraint amongst several parameters [25]

$$\kappa^2 = \sigma_\varepsilon C_\mu^{1/2} (C_{\varepsilon 2} - C_{\varepsilon 1}), \quad (3)$$

where  $\kappa$  is the von Karman constant. Likewise, the balance between the production  $\mathcal{P}$  and dissipation  $\varepsilon$  of turbulent kinetic energy in free shear layers gives the constraint [25]

$$\left( \frac{\mathcal{P}}{\varepsilon} \right) = \frac{C_{\varepsilon 2} - 1}{C_{\varepsilon 1} - 1}. \quad (4)$$

We use (3)-(4) to fix the value of  $\sigma_\varepsilon$  and  $C_{\varepsilon 1}$ , where we follow the advice of the authors of [24] and set  $\mathcal{P}/\varepsilon = 2.09$  for the latter. This is the value obtained when inserting the nominal values of  $C_{\varepsilon 1}$  and  $C_{\varepsilon 2}$  in (4). Experimental values of  $\mathcal{P}/\varepsilon$  vary, but they are in general closer to 1.7 than 2.09 [33]. But, as we shall see in Section 5.1, using  $\mathcal{P}/\varepsilon = 2.09$  will center the posterior  $C_{\varepsilon 2}$  distributions close to its standard value of 1.92. In this case we expect greater stability of the  $k - \varepsilon$  model, especially in the prediction phase (see Section 4.2) where we will propagate multiple posterior parameter distributions through the RANS code. Alternatively, we could calibrate for  $\mathcal{P}/\varepsilon$  instead. In this case we can fix either  $C_{\varepsilon 1}$  or  $C_{\varepsilon 2}$  to infer the value of the other via relation (4), leading to a similar situation as the present setup. Although not investigated in the current article, a final

option would be to make both  $C_{\varepsilon 2}$  and  $\mathcal{P}/\varepsilon$  random, in which case  $C_{\varepsilon 1}$  becomes a function of 2 random variables.

We are now left with 4 calibration parameters, namely  $C_{\varepsilon 2}$ ,  $C_\mu$ ,  $\sigma_k$  and  $\kappa$ . The nominal values of these coefficients as traditionally determined in the literature are given in Table 1.

### 2.2.2. The Wilcox (2006) $k - \omega$ model

This model is also a two-equation model, which computes the eddy viscosity as  $\nu_T = k/\tilde{\omega}$ , where  $\tilde{\omega}$  is a quantity related to the specific dissipation  $\omega = \varepsilon/k$  [36]. The following closure coefficients are present:  $\alpha$ ,  $\beta_o$ ,  $\beta^*$ ,  $\sigma$ ,  $\sigma^*$  and  $\sigma_{do}$ . Again, a constraint between parameters, equivalent to (3), can be found [36]

$$\alpha = \frac{\beta}{\beta^*} - \frac{\kappa^2}{2\sqrt{\beta^*}}, \quad (5)$$

which we use to fix  $\alpha$ . The nominal values of the remaining coefficients can be found in Table 1.

### 2.2.3. The Spalart-Allmaras model

This is a one-equation model where a transport equation is solved for a viscosity-like variable  $\tilde{\nu}$  [30]. It contains the following seven closure coefficients:  $C_{b1}$ ,  $C_{b2}$ ,  $\sigma$ ,  $C_{w2}$ ,  $C_{w3}$ ,  $C_{v1}$  and  $\kappa$ . The coefficient  $C_{w1}$  is constraint by the values of the other coefficients as

$$C_{w1} = \frac{C_{b1}}{\kappa^2} + \frac{1 + C_{b2}}{\sigma}, \quad (6)$$

the other 7 parameters are considered for calibration.

### 2.2.4. The Baldwin-Lomax model

Unlike the preceding models, which require the solution of one or two additional transport equations, this model assumes an algebraic expression for  $\nu_T$  [1]. It is therefore the simplest model in our set. Like the other models it also contains closure coefficients, namely  $A_0^+$ ,  $C_{cp}$ ,  $C_{kleb}$ ,  $C_{wk}$ ,  $\alpha$  and  $\kappa$ , which are all considered for calibration. Again, their nominal values can be found in Table 1.

### 2.2.5. The Stress- $\omega$ model

All preceding models utilize the Boussinesq hypothesis to close the RANS equations, which assumes that the principal axes of the Reynolds-stress tensor  $\tau_{ij}$  are coincident with those of the mean strain-rate tensor  $S_{ij}$  [36]. The constant of proportionality between  $\tau_{ij}$  and  $S_{ij}$  is the eddy viscosity  $\nu_T$ , which is calculated by means of a turbulence model, and appears in (1). However, the validity of the Boussinesq hypothesis is questionable, see e.g. [28] for a discussion. A class of (more complex) turbulence models attempting to remedy these shortcomings are stress-transport models. These directly write a model transport equation for the Reynolds-stress tensor  $\tau_{ij}$ , rather than relying on the linear relationship between  $\tau_{ij}$  and  $S_{ij}$  suggested by the Boussinesq hypothesis. An ancillary transport equation for a turbulent length scale is also required, (see [36] for further details). One such a model is the stress- $\omega$  model of Wilcox [36]. As it uses  $\omega$  from the

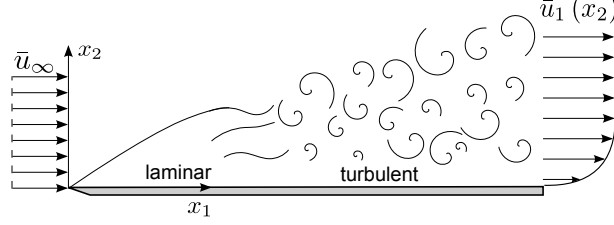


Figure 1: A schematic overview of a flat-plate turbulent boundary layer. Shown are the uniform inflow velocity  $\bar{u}_\infty$ , a schematic visualisation of the instantaneous flow field, and the averaged velocity  $\bar{u}_1(x_2)$

$k - \omega$  model for certain terms in the  $\tau_{ij}$  transport equation, it contains all the same closure coefficients as the Wilcox (2006)  $k - \omega$  model, plus two additional coefficients, denoted as  $C_1$  and  $C_2$ .

Table 1: Traditional closure coefficient values [36].

$k - \varepsilon$		$k - \omega$		SA		BL		stress- $\omega$	
$\kappa$	0.41	$\kappa$	0.41	$\kappa$	0.41	$\kappa$	0.4	$\kappa$	0.41
$C_\mu$	0.09	$\beta^*$	0.09	$c_{v1}$	7.1	$A_0^+$	26.0	$\beta^*$	0.09
$C_{\varepsilon 2}$	1.92	$\beta_o$	0.0708	$c_{b1}$	0.1355	$\alpha$	0.0168	$\beta_o$	0.0708
$\sigma_k$	1.0	$\sigma^*$	0.6	$c_{b2}$	0.622	$C_{cp}$	1.6	$\sigma^*$	0.6
		$\sigma$	0.5	$\sigma$	2/3	$C_{kleb}$	0.3	$\sigma$	0.5
		$\sigma_{do}$	0.125	$c_{w2}$	0.3	$C_{wk}$	1.0	$\sigma_{do}$	0.125
				$c_{w3}$	2.0			$C_1$	9/5
								$C_2$	10/19

### 3. Turbulent boundary-layer configuration

In the following, we investigate the predictive capabilities of our Bayesian methodology for a specific class of flows, namely incompressible turbulent boundary layers subject to different pressure gradients. Figure 1 depicts a sketch of a turbulent boundary layer.

#### 3.1. Numerical solutions

As in [8], we use the fast boundary-layer code EDDYBL[35] to obtain efficient numerical solutions to the boundary-layer problem (1), supplemented with one of the described turbulence models. To make sure that the discretization error does not dominate over the uncertainties that we wish to quantify we use a computational grid in which the first grid point satisfies  $\Delta y^+ < 1$ . For more information on the finite-difference scheme and discretization error of EDDYBL, we refer to [8, 35]. Here,  $y^+ := x_2/\delta_\nu$  represents the distance to the wall  $x_2$ , normalized by the length-scale of the viscous layer  $\delta_\nu := \nu/\sqrt{\tau_w/\rho}$ ,  $\tau_w$  being the wall shear stress. The velocity scale of the viscous layer, i.e. the so-called friction velocity, is denoted as  $u_\tau := \sqrt{\tau_w/\rho}$ .

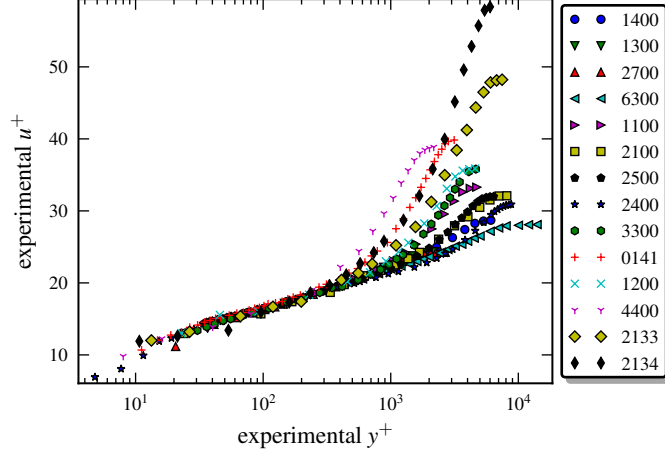


Figure 2: Our experimental data set.

### 3.2. Experimental boundary-layer data

EDDYBL comes with configuration files which mimic the experiments described in the 1968 AFOSR-IFP-Stanford conference proceedings [5]. From this data source, we selected one zero pressure-gradient flow, and 13 flows from other types of available pressure gradients, which range from favorable ( $d\bar{p}/dx_1 < 0$ ) to strongly adverse ( $d\bar{p}/dx_1 > 0$ ) gradients. These 14 flows are described in table 2. The identification number of each flow is copied from [5]. We plotted the experimentally determined, non-dimensional, streamwise velocity profiles in Figure 2. As usual, the normalized streamwise velocity is defined as  $u^+ \equiv \bar{u}_1/u_\tau$ . Too much weight should not be given to the classifications of the severity of the adverse gradients in Table 2, since some flows (such as 2100 and 2400) experience multiple gradient types along the spanwise direction.

Therefore, it can be more informative to also inspect which flows are in equilibrium. There are two main forces acting on a boundary-layer. First, the pressure gradient acts with a force per unit length and width as  $\delta^* d\bar{p}/dx_1$ , where  $\delta^*$  is the boundary-layer displacement thickness. The second force on the boundary layer comes from the wall-shear stress  $\tau_w$ . A boundary-layer is said to be in equilibrium when the ratio of these two forces is constant along the length of the boundary layer, i.e. when the so-called equilibrium parameter  $\beta_T := (\delta^*/\tau_w)(d\bar{p}/dx_1)$  is constant [4]. We plot the absolute value of the experimentally determined  $\beta_T$  values in figure 3. Flows such as 6300 with constant  $\beta_T$  are simple flows for which the turbulence models should be able to make accurate predictions. Other flows are more complex. We especially expect that models will encounter difficulties in the last part of flow 2100, i.e. flows 2133 and 2134.

We calibrate each model of section 2 for each flow of table 2 separately, using one velocity profile as experimental data. We omit any experimental data in the viscous wall region. Since some models have damping functions which dominate over the effect of the closure coefficients in this region, little information is obtained from the measurements here. We investigated the effect of not including near-wall data on the posterior distributions in [8]. We found that the posteriors were not significantly affected by this



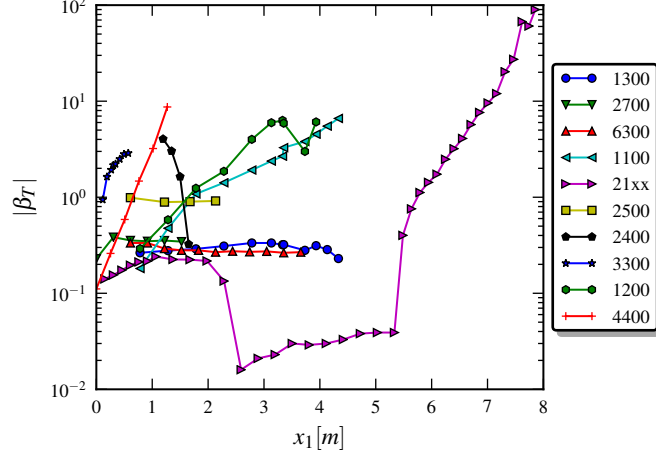


Figure 3: Experimental  $|\beta_T|$  values for all flows but 1400 and 0141. Source [5].

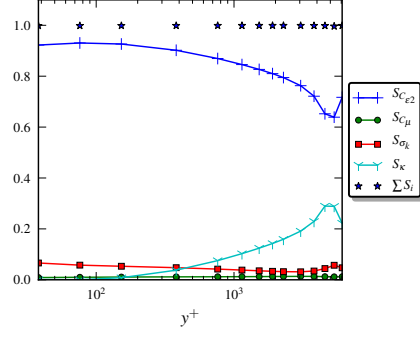
exclusion.

### 3.3. Sensitivity analysis of boundary-layer problem

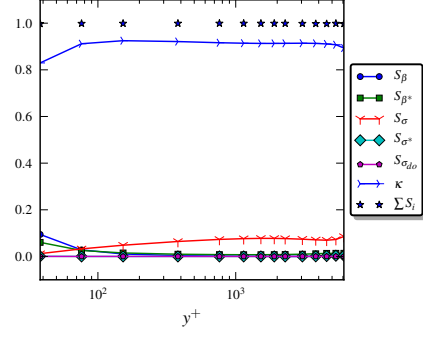
To get an idea of which coefficients can be informed by our data, we perform a variance-based, global-sensitivity analysis before we start the calibrations. We expect that we can obtain informative posterior distributions for the parameters with a high sensitivity measure, whereas the ones with low sensitivities are likely to yield posterior distributions that do not differ much from the uniform prior distributions. To this end we call  $\boldsymbol{\theta} \in \mathbb{R}^Q$  a random vector, the components of which are  $Q$  uncertain closure coefficients. We carry out a sensitivity analysis of the output quantity  $u^+ = u^+(\mathbf{y}^+; \boldsymbol{\theta})$  by computing the Sobol indices  $S_w$ , defined as [29]:

$$\begin{aligned} D &= \text{Var}_{\mathcal{W}} \{u^+\}, \\ D_w &= \text{Var}_w \{\mathbb{E}_{w'}(u^+ | \boldsymbol{\theta}_w)\}, \\ S_w &= D_w / D \end{aligned} \tag{7}$$

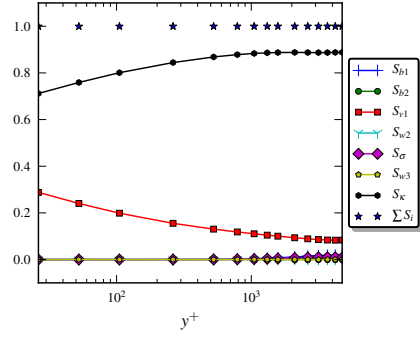
where  $w \subset \{1, \dots, Q\} = \mathcal{W}$  indexes the components of  $\boldsymbol{\theta}$ ,  $w' = \mathcal{W} \setminus w$ , and  $\text{Var}_w\{\cdot\}$  indicates variance taken over priors of  $\boldsymbol{\theta}_w$ , etc. The indices  $S_w$  satisfy  $\sum_{\mathcal{P}(\mathcal{W})} S_w = 1$ , where  $\mathcal{P}(\mathcal{W})$  is the power set of  $\mathcal{W}$ . A value close to unity for  $S_i$  can be interpreted as the coefficient corresponding to  $i \in \mathcal{W}$  being responsible for most of the total variance in  $u^+$  on its own (also without interaction effects with other parameters). A value close to zero indicates an uninfluential parameter. The  $D_w$  are computed using a 3rd-order polynomial approximation to the response in the stochastic space [16], given which the expectation and variance can be evaluated analytically [31, 32]. The method of [16] requires running  $(p+1)^d$  deterministic simulations, where  $p$  is the polynomial order of the approximation and  $d$  is the number of uncertain parameters. We contrasted the results with those obtained from a 2nd-order polynomial approximation, and found no



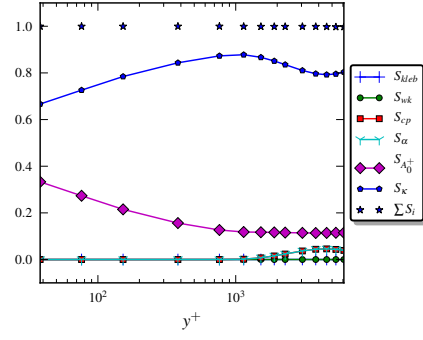
(a) The  $k - \varepsilon$  model.



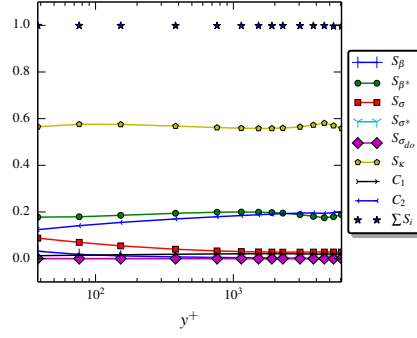
(b) The  $k - \omega$  model.



(c) The Spalart-Allmaras model.



(d) The Baldwin-Lomax model.



(e) The stress- $\omega$  model.

Figure 4: Sobol indices of the considered turbulence models for flow case 1400.

Table 2: Flow descriptions, source [5].

Identification	Type	Description
1400	Zero	Equilibrium boundary layer at constant pressure
1300	Fav	Near-equilibrium boundary layer in moderate negative pressure gradient
2700	Fav	Equilibrium boundary layer in mild negative pressure gradient
6300	Fav	Near-equilibrium boundary layer growing beneath potential flow on model spillway
1100	Mild adv	Boundary layer in diverging channel
2100, 2133, 2134	Div	Boundary layer on large airfoil-like body; pressure gradient first mildly negative, then strongly positive, with eventual separation
2500	Mild adv	Equilibrium boundary layer in mild positive pressure gradient
2400	Div	Initial equilibrium boundary layer in moderate positive pressure gradient; pressure gradient abruptly decreases to zero, and flow relaxes to new equilibrium
3300	Mod adv	Boundary layer, initially at constant pressure, developing into equilibrium flow in moderate positive pressure gradient
0141	Str adv	Boundary-layer with strong adverse pressure gradient, source [14]
1200	Str adv	Boundary layer in diverging channel with eventual separation
4400	Str adv	Boundary layer in strong positive pressure gradient

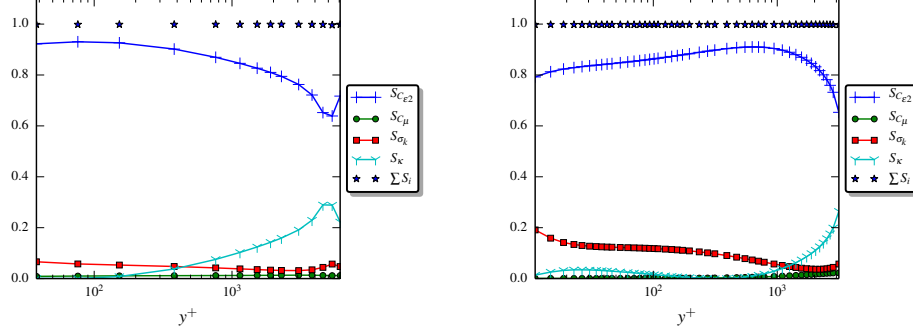
significant deviations. This makes us confident about the convergence of the stochastic method with respect to the polynomial order.

We considered uniform input pdfs for all coefficients, with the boundaries located at  $\pm 10\%$  of the nominal values displayed in Table 1. The results can be found in Figure 4. In case of the  $k - \varepsilon$  model with the constraints (3)-(4) applied, the most influential parameter is  $C_{\varepsilon 2}$ , followed by  $\kappa$ . The parameter  $C_{\varepsilon 2}$  is the proportionality constant of the dissipation term in the transport equation of  $\varepsilon$ . The von Karman constant  $\kappa$  enters the model through the constraint on  $\sigma_\varepsilon$  (3), which is the proportionality constant of the diffusion term in the  $\varepsilon$  equation.

The Sobol indices of the  $k - \omega$  model were calculated with the constraint (5) applied. We can see that  $\kappa$  is by far the most influential parameter. Again,  $\kappa$  enters the model through a constraint. This time the constraint is on  $\alpha$ , which regulates the production of  $\omega$ . The second parameter that has an impact on the computed  $u^+$  is  $\sigma$ , which is the proportionality constant of the diffusion term of the  $\omega$  transport equation.

For the SA model we see that again  $\kappa$  is the most influential parameter when it comes to  $u^+$  profiles. And, like the two preceding models, it also enters the model through a constraint. Here, the constraint (6) is on  $C_{w1}$ , which regulates the dissipation of  $\tilde{\nu}$ . A second influential parameter is  $C_{v1}$ , which appears in the function  $f_{v1}$ . In turn, this function relates  $\tilde{\nu}$  to the eddy viscosity as  $\nu_T = f_{v1}\tilde{\nu}$ .

For the Baldwin-Lomax model the most influential parameter is  $\kappa$  as well, although it enters the model directly through algebraic expressions of the eddy viscosity  $\nu_T$ . The  $u^+$  profiles are sensitive to  $A_0^+$  as well, which is a parameter in the van Driest damping function [36].



(a) Sobol indices for  $k - \epsilon$  model for case 1400. (b) Sobol indices for  $k - \epsilon$  model for case 0141.

Figure 5: Variation of Sobol indices over flow cases for the  $k - \epsilon$  model.

Finally, for the stress- $\omega$  model we find that again  $\kappa$ , which performs exactly the same function as for the  $k - \omega$  model, is the most influential, although to a lesser extend compared to the other 4 models. Unlike the  $k - \omega$  model however, the parameters  $\beta$  and  $\beta^*$  also carry some weight.

So far we have examined Sobol indices for one particular flow case, i.e. 1400. The variation over different flow cases is not very large, although small changes can occur, see Figure 5 for an example, which compares Sobol indices associated to the closure coefficients of  $k - \epsilon$  computed for flow case 1400, and for the strong adverse pressure gradient case 0141. In both cases  $C_{\epsilon 2}$  is by far the most influential. And in both cases  $\kappa$  becomes more influential than  $\sigma_k$  when we move away from the wall. However, in the strong adverse case the intersection point is at a higher  $y^+$  value, and before that  $\sigma_k$  has a higher Sobol index than it had in case 1400.

#### 4. Statistical Methodology

The first step in our method consists of calibrating the closure coefficients of each turbulence model of Section 2.2 against different sets of experimental data. The calibration framework is described in Section 4.1. We obtain 14 sets of posterior parameter distributions for each model by calibrating our computer code to each flow of Section 3.2. We summarize this large amount of information by calculating Highest Posterior Density intervals, see Section 4.5. The final phase consists of using the posterior parameter distributions to make predictions for a new flow case. To achieve this we use the BMA framework described in Section 4.2.

##### 4.1. Calibration framework

The essential step in Bayesian inference is the application of Bayes' rule, given by

$$p(\boldsymbol{\theta} | \mathbf{z}) = \frac{p(\mathbf{z} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{z})}, \quad (8)$$

where  $\mathbf{z}$  is the vector of experimental data and  $\boldsymbol{\theta}$  is the vector containing the random input parameters, i.e. the closure coefficients in our case. The law of total probability states that  $p(\mathbf{z}) = \int p(\mathbf{z} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$ . Furthermore,  $p(\boldsymbol{\theta} | \mathbf{z})$  is the joint posterior pdf of the coefficients given the observed data,  $p(\boldsymbol{\theta})$  represents the prior belief on  $\boldsymbol{\theta}$  and  $p(\mathbf{z} | \boldsymbol{\theta})$  is the likelihood function which represents our statistical model. Details of the calibration framework can be found in [8], so for brevity we repeat only the key elements here.

Our statistical model is based on the framework of Cheung et al. [3]:

$$\begin{aligned}\mathbf{z} &= \zeta(\mathbf{y}^+) + \mathbf{e} \\ \zeta(\mathbf{y}^+) &:= \eta(\mathbf{y}^+) \cdot u^+(\mathbf{y}^+, \mathbf{S}; \boldsymbol{\theta}), \\ \mathbf{e} &\sim \mathcal{N}(\mathbf{0}, \Lambda), \quad \Lambda := \text{diag}(\boldsymbol{\lambda}) \\ \eta &\sim \text{GP}(1, c_\eta).\end{aligned}\tag{9}$$

Here  $\mathbf{z} \in \mathbb{R}^N$  is a vector of measurements of the flow speed at locations distances  $\mathbf{y}^+ \in \mathbb{R}^N$  from the wall. The specific flow-case, or *scenario*, is completely described by the known vector of deterministic parameters  $\mathbf{S}$ . This includes boundary conditions, material parameters, etc. The vector  $\boldsymbol{\theta}$  contains the turbulence model closure coefficients. The computer code output  $u^+(\mathbf{y}^+, \mathbf{S}; \boldsymbol{\theta})$ , represents the prediction of the flow speed at  $\mathbf{y}^+$ , i.e. our simulation's approximation to  $\mathbf{z}$ . The measurement error  $\mathbf{e}$  is modelled as uncorrelated, unbiased, with known variance  $\boldsymbol{\lambda}$ . Remaining model error is modelled with the Gaussian-process  $\eta$ .

This statistical model leads directly to the likelihood:

$$\begin{aligned}p(\mathbf{z} | \boldsymbol{\theta}, \gamma) &= \frac{1}{\sqrt{(2\pi)^N |K|}} \exp \left[ -\frac{1}{2} \mathbf{d}^T K^{-1} \mathbf{d} \right], \\ \mathbf{d} &:= \mathbf{z} - u^+(\mathbf{y}^+, \mathbf{S}; \boldsymbol{\theta}), \\ K &:= \Lambda + K_\zeta.\end{aligned}\tag{10}$$

The covariance matrix  $K$  is specified via

$$[K_\zeta]_{ij} := c_\zeta(y_i^+, y_j^+ | \boldsymbol{\theta}, \gamma) \tag{11}$$

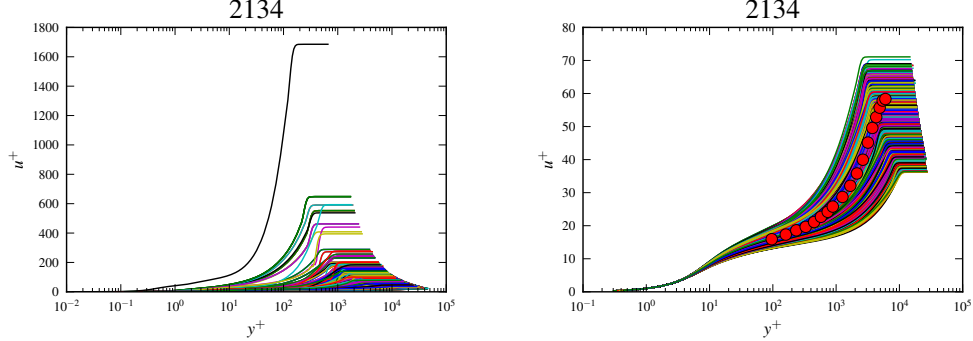
$$= u^+(y_i^+, \mathbf{S}; \boldsymbol{\theta}) c_\eta(y_i^+, y_j^+ | \gamma) u^+(y_j^+, \mathbf{S}; \boldsymbol{\theta}), \tag{12}$$

where  $\gamma := (\sigma_{hp}, \alpha_{hp})$  is a vector of unknown hyper-parameters, which will be calibrated alongside  $\boldsymbol{\theta}$ . Finally, the covariance function of  $\eta$  is taken as

$$c_\eta(y_i^+, y_j^+ | \gamma) := \sigma_{hp}^2 \exp \left[ -\left( \frac{y_i^+ - y_j^+}{10^{\alpha_{hp}} l} \right)^2 \right]. \tag{13}$$

The magnitude of the correlation length  $10^{\alpha_{hp}} l$  is determined by the hyper parameter  $\alpha_{hp}$ , where  $l$  is a user-specified length scale. We set  $l$  to 5  $y^+$  units, which signifies the end of the viscous-wall region. The magnitude of  $c_\eta$  itself is determined by  $\sigma_{hp}$ . We prescribe uniform priors for both  $\sigma_h$  and  $\alpha_h$ , with ranges  $[0.0, 1.0]$  and  $[0.0, 4.0]$  respectively.

To obtain samples from the posterior parameter distributions, we employ the Markov-chain Monte-Carlo (MCMC) method [11]. We subsequently approximate the marginal pdf of each closure coefficient using kernel-density estimation, using the last 5,000 (out



(a) 5000 samples obtained by propagating  $\pm 50\%$  priors through the stress- $\omega$  model configured for case 2134. Clearly some extremely non-physical samples are present considering that the maximum experimental  $u^+$  value is 58.3.

(b) 5000 samples obtained by propagating  $\pm 20\%$  priors through the stress- $\omega$  model configured for case 2134. The red dots are the experimental data.

Figure 6: Outlier analysis as a function of the prior domain.

of a total of 40,000) samples from the Markov chain. Visual inspection of the chain is used to judge convergence. Those cases not achieving convergence at 35,000 samples, are re-sampled using a larger number of total samples.

#### 4.1.1. Priors for $\theta$ and $\gamma$

All priors, for both the closure coefficients  $\theta$  and hyper-parameters  $\gamma$ , are independent uniform distributions. The choice of interval end-points was mainly on the range of coefficients for which the solver was stable and provides physical solutions. By physical solutions we mean  $u^+$  profiles that are not excessively large. Take for instance the stress- $\omega$  model. If we assume a uniform prior where we perturb the coefficients by  $\pm 50\%$  from their nominal values, we found  $u^+$  values that were extremely large. The worst case we encountered is depicted in Figure 6(a). In the calibration phase this does not pose a problem. The MCMC chain will automatically reject the bad samples and move to a part of the stochastic domain that does generate samples close to the data. However, our predictive methodology (Section 4.2), involves the propagation of multiple posterior distributions through the code at validation settings. In this case the extremely unphysical samples can re-occur, although with low probability. This was observed especially in the case of the stress- $\omega$  model. Therefore we chose the prior domain such that these extreme outliers are not possible. At the same time we checked that the solution variation, obtained by prior propagation, was still large enough to easily encompass the data, see e.g. Figure 6(b).

For the  $k - \varepsilon$  model, which does not suffer from extreme outliers, we re-used our posterior distributions from [8]. For the stress- $\omega$  and  $k - \omega$  models the prior domain boundaries of each coefficient were set at  $\pm 20\%$  of their respective nominal values. For the Spalart-Allmaras and Baldwin-Lomax models we used  $\pm 30\%$ .

#### 4.2. Bayesian scenario averaging: Prediction

Calibration according to Section 4.1 gives values of  $\theta$  that reproduce the single scenario  $\mathbf{S}$  accurately. The variance of  $p(\theta|z)$  only contains information about how well  $\theta$  has been identified by  $\mathbf{z}$ , but nothing about the variability needed in  $\theta$  when varying the scenario  $\mathbf{S}$ . It is this latter information that is relevant for predicting model inadequacy in some unmeasured scenario  $\mathbf{S}'$ . In the remaining derivation the effect of  $\gamma$  is integrated out, that is we work with marginal posteriors of  $\theta$  only.

To capture this we consider a set of  $K$  measured scenarios  $\mathcal{S} = \{\mathbf{S}_1, \dots, \mathbf{S}_K\}$  chosen to encompass the class of flows we wish to predict. In this work they are a subset of the 14 cases described in Section 3.2. The class of flows we may expect to reliably predict is therefore the class of attached flat-plate boundary-layers at any pressure-gradient within the range of cases in  $\mathcal{S}$ . However, the predictive capability of our  $\theta$  for a QoI outside this class is investigated in Section 5.4. The measurements associated with these scenarios are denoted  $\mathbf{z} = \{\mathbf{z}_1, \dots, \mathbf{z}_K\}$ . Furthermore, to capture some model error due to the form of the turbulence model, we consider the set of  $I$  models  $\mathcal{M} = \{M_1, \dots, M_I\}$ . Here we use the 5 models described in Section 2.2.

For each turbulence model and each scenario the calibration procedure results in a posterior on  $\theta$ :

$$\hat{\theta}_{i,k} \sim \theta | M_i, \mathbf{S}_k, \mathbf{z}_k, \quad i = 1, \dots, I, \quad k = 1, \dots, K,$$

whereby the nature of  $\theta$  depends on the model. Formally we write  $\theta \in \Theta_i$  for  $i = 1, \dots, I$  depending on the context.

Now let  $\Delta$  be the Quantity of Interest in an unmeasured scenario  $\mathbf{S}' \notin \mathcal{S}$ . This is a velocity at  $y^{+'}$ , modelled as

$$\Delta = u_i^+(y^{+'}, \mathbf{S}'; \theta), \quad (14)$$

but it could equally be a quantity of entirely different type, such as turbulent kinetic energy. In (14),  $u_i^+(\cdot)$  indicates the simulation code under turbulence model  $M_i$ . Evaluating  $p(\Delta|\mathbf{z})$  conditional on  $\mathcal{M}$  and  $\mathcal{S}$  is commonly called Bayesian Model Averaging (BMA) – but in our case could equally be called Bayesian Model-Scenario Averaging (BMSA), since the choice of  $\mathcal{S}$  is critical to the validity of the resulting estimate. In particular

$$\begin{aligned} p(\Delta|\mathbf{z}) &\stackrel{(a)}{=} \sum_{i=1}^I \sum_{k=1}^K \int_{\Theta_i} p(\Delta, M_i, \mathbf{S}_k, \theta | \mathbf{z}_k) d\theta \\ &\stackrel{(b)}{=} \sum_{i=1}^I \sum_{k=1}^K \int_{\Theta_i} p(\Delta | M_i, \theta) p(\theta | M_i, \mathbf{S}_k, \mathbf{z}_k) \mathbb{P}(M_i, \mathbf{S}_k | \mathbf{z}_k) d\theta \\ &\stackrel{(c)}{=} \sum_{i=1}^I \sum_{k=1}^K \int_{\Theta_i} \underbrace{p(\Delta | M_i, \theta) p(\hat{\theta}_{i,k})}_{p(\hat{\Delta}_{i,k})} \mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k) \mathbb{P}(\mathbf{S}_k) d\theta. \end{aligned} \quad (15)$$

Here we use  $\mathbb{P}(\cdot)$  to denote a probability mass function (pmf) rather than a pdf. The equalities in the above are established using: (a) the law of total probability, and the fact that scenario  $k$  depends only on  $\mathbf{z}_k$ ; (b) the definition of conditional probability (twice), and the conditional independence of  $\Delta$  from  $\mathbf{z}$  and  $\mathbf{S}_k$  given  $\theta$ ; (c) the definition of conditional probability, and the conditional independence of  $\mathbf{S}_k$  and  $\mathbf{z}_k$ . The braced

term  $\hat{\Delta}_{i,k}$  is equivalent to the posterior  $\hat{\theta}_{i,k}$  propagated through the simulation code for  $\Delta$  (14). The weights  $\mathbb{P}(M_i|\mathbf{S}_k, \mathbf{z}_k)$  are commonly interpreted as the level of evidence for a given model [12], in this case specific to a given scenario. Finally we are free to choose  $\mathbb{P}(\mathbf{S}_k)$ , and it is reasonable to preferentially weight those scenarios that are (in some sense) similar to the prediction scenario  $\mathbf{S}'$ .

The posterior probability of model  $M_i$  can be calculated through another application of Bayes' rule

$$\mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k) = \frac{p(\mathbf{z}_k | M_i, \mathbf{S}_k) \mathbb{P}(M_i | \mathbf{S}_k)}{\sum_{j=1}^J p(\mathbf{z}_k | M_j, \mathbf{S}_k) \mathbb{P}(M_j | \mathbf{S}_k)}, \quad (16)$$

$\forall k \in \{1, 2, \dots, K\}$ , where additionally

$$p(\mathbf{z}_k | M_i, \mathbf{S}_k) = \int_{\Theta_i} p(\mathbf{z}_k | \theta, M_i, \mathbf{S}_k) p(\theta | M_i, \mathbf{S}_k) d\theta. \quad (17)$$

Furthermore, the axioms of probability require that

$$\sum_{i=1}^I \mathbb{P}(M_i | \mathbf{S}_k) = 1, \quad \text{and} \quad \sum_{k=1}^K \mathbb{P}(\mathbf{S}_k) = 1. \quad (18)$$

Following Draper [6], and the derivation in [19], the leading moments of  $p(\Delta|\mathbf{z})$  can be written as

$$\mathbb{E}[\Delta | \mathbf{z}] = \sum_{i=1}^I \sum_{k=1}^K \mathbb{E}[\hat{\Delta}_{i,k}] \mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k) \mathbb{P}(\mathbf{S}_k), \quad (19)$$

$$\begin{aligned} \text{Var}[\Delta | \mathbf{z}] &= \sum_{i=1}^I \sum_{k=1}^K \text{Var}[\hat{\Delta}_{i,k}] \mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k) \mathbb{P}(\mathbf{S}_k) \\ &+ \sum_{i=1}^I \sum_{k=1}^K \left( \mathbb{E}[\hat{\Delta}_{i,k}] - \mathbb{E}[\Delta | \mathbf{S}_k, \mathbf{z}_k] \right)^2 \mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k) \mathbb{P}(\mathbf{S}_k) \\ &+ \sum_{k=1}^K \left( \mathbb{E}[\Delta | \mathbf{S}_k, \mathbf{z}_k] - \mathbb{E}[\Delta | \mathbf{z}] \right)^2 \mathbb{P}(\mathbf{S}_k), \end{aligned} \quad (20)$$

Here,  $\mathbb{E}[\hat{\Delta}_{i,k}]$  and  $\text{Var}[\hat{\Delta}_{i,k}]$  are the mean and variance obtained by propagating  $\hat{\theta}_{i,k}$  through the code of  $\Delta$ . Furthermore,

$$\mathbb{E}[\Delta | \mathbf{S}_k, \mathbf{z}_k] = \sum_{i=1}^I \mathbb{E}[\hat{\Delta}_{i,k}] \mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k).$$

This decomposition of the variance into 3 (positive) terms, allows for additional interpretation. The first term on the right-hand side is called the *within-model, within-scenario* variance. This can be considered a measure of the effect of the variance in all the  $\hat{\theta}_{i,k}$  on the variance in  $\Delta|\mathbf{z}$  – specifically an average weighted according to the importance of



each model and scenario. The second term is the *between model, within scenario* variance. It is large when when different models applied to the same scenario give different predictions. The last term is the *between scenario* variance. It accounts for the fact that the same model  $M_i$ , calibrated under different scenarios, results in different posterior  $\theta$  and therefore different predictions of  $\Delta$  [8].

#### 4.3. Smart scenario weighting

In order to close the above system we must specify  $\mathbb{P}(M_i|\mathbf{S}_k)$  and  $\mathbb{P}(\mathbf{S}_k)$ . As we will show in Section 5.3.1, simply specifying a uniformly distributed  $\mathbb{P}(\mathbf{S}_k)$  yields an unacceptably large variance. This is due to the large spread of  $\hat{\theta}_{i,k}$  between scenarios, see Section 5.2, resulting in a large spread of  $\Delta$ . We would like to preferentially weight those scenarios in  $\mathcal{S}$  which are – in some relevant sense – similar to  $\mathbf{S}'$ . Furthermore we would like to do this in an automatic, geometry- and flow-independent way.

We propose choosing  $\mathbb{P}(\mathbf{S}_k)$  to preferentially weight those scenarios for which all models in  $\mathcal{M}$  give similar mean predictions of  $\Delta$ , under the coefficients  $\hat{\theta}_{i,k}$ . Equally, we assign low probabilities to scenarios with large scatter in the model predictions of  $\Delta$ . The rationale is that, if  $\mathbf{S}'$  is very similar to  $\mathbf{S}_k$ , the 4 models are expected to give similar predictions under  $\hat{\theta}_{i,k}$ , as they have each been calibrated to the same data  $\mathbf{z}_k$ , containing physical processes similar to  $\mathbf{S}'$ . If  $\mathbf{S}'$  is completely different to  $\mathbf{S}_k$ , the choice of  $\hat{\theta}_{i,k}$  is likely to be inappropriate, and the models – provided they are sufficiently heterogeneous – will deliver different predictions.

In particular we define the scenario probabilities as

$$\begin{aligned}\mathcal{E}_k &= \sum_{i=1}^I \|\mathbb{E}[\hat{\Delta}_{i,k}] - \mathbb{E}[\Delta | \mathbf{S}_k, \mathbf{z}_k]\|_2 \\ \mathbb{P}(\mathbf{S}_k) &= \frac{\mathcal{E}_k^{-p}}{\sum_{k=1}^K \mathcal{E}_k^{-p}}, \quad \forall \mathbf{S}_k \in \mathcal{S}.\end{aligned}\tag{21}$$

Here  $\mathcal{E}_k$  represents the measure for the prediction similarity, and  $p$  is a tuning parameter, controlling the degree to which the preferred scenarios are weighted. Setting  $p = 0$  yields uniform probabilities, and as  $p \rightarrow \infty$  a single scenario is selected (provided that the  $\mathcal{E}_k$  have distinct values).

#### 4.4. Numerical evaluation

In CFD evaluating the likelihood involves running a CFD code. Here, the posteriors  $\hat{\theta}_{i,k}$  are approximated with Markov-chain Monte-Carlo (MCMC) methods [18, 11], which provide samples  $\hat{\theta}_{i,k}^n, n = 1, 2, \dots, N$  from  $\hat{\theta}_{i,k}$ . This chain approximation is thereafter used to approximate everything else by Monte-Carlo, e.g.:

$$\mathbb{E}(\hat{\Delta}_{i,k}) \approx \frac{1}{N} \sum_{n=1}^N u_i^+ \left( y^{+'}, \mathbf{S}'; \hat{\theta}_{i,k}^n \right).\tag{22}$$

Clearly the process outlined here is impractically costly for an expensive CFD simulation code applied to a complex geometry – here it is possible due to the use of a boundary-layer code. Our intention in this paper is to explore the statistical framework, and determine to what extent prediction of model inadequacy is possible, under the assumption that numerical costs can be handled. Success will motivate us to study numerical techniques for improving the computational efficiency of the procedure. There are many possible approaches, including: surrogate modelling [7], Occam’s window [12], zero-variance MCMC methods [20], etc. Also, a means of reducing the computational cost is described in Section 5.5. Furthermore, the calibration step is perhaps necessary for simple geometries only, giving a database of coefficient distributions  $\hat{\theta}_{i,k}$  which can then be applied to more complex geometries. A flowchart of our entire procedure, from calibration to BMA prediction, can be found in Appendix A.

#### 4.5. Summarizing posteriors: HPD intervals

To summarize the 60 posterior parameter distributions of  $\hat{\theta}_{i,k}$  for the purposes of visualization we construct Highest Posterior Density (HPD) intervals on the 1D marginals. An HPD interval is a Bayesian credible interval which satisfies two main properties, i.e.:

1. The probability density for every point inside the interval is greater than that for every point outside the interval.
2. For a given probability content  $1 - \beta_{hpd}$ ,  $\beta_{hpd} \in (0, 1)$ , the interval is of the shortest length.

We use the algorithm of Chen et al. [2] to approximate the HPD intervals using the MCMC samples, and we set  $\beta_{hpd} = 0.5$ . This summary is appropriate provided the posteriors are not strongly multi-modal, and there is no strong correlation structure between coefficients. Both these conditions are satisfied for all  $\hat{\theta}_{i,k}$ .

## 5. Results

### 5.1. HPD intervals of coefficient posteriors

To evaluate the success of the individual calibrations, we inspect the coefficient posterior distributions  $\hat{\theta}_{i,k}$ . Instead of plotting full multi-dimensional distributions, we plot HPD intervals of the two most sensitive parameters for each model, shown in Figures 7 and 8. The scenarios are presented roughly in order of increasing pressure-gradient, but with the zero-pressure-gradient case (1400) first. The coefficient  $\kappa$  was the most well-informed coefficient for all models except  $k - \varepsilon$ , for which it was the second most well-informed coefficient. As can be seen by the small width of the individual HPD intervals compared to the prior width, these coefficients have been informed by the data. This was to be expected considering the results of the sensitivity analysis of Section 3.3. The remaining closure coefficients were not informed well enough by the data to be interesting.

The four  $\kappa$ ’s presented in Figure 7 are not the same  $\kappa$ : the coefficient has a different meaning and influence in each model. Neither should they be confused with any model-independent, or “true” von Karman constant. Nonetheless, ideally  $\kappa$  would be independent of scenario for a given model. Figure 7 clearly shows statistically significant variation in calibrated coefficients across scenarios. This may be interpreted as

pressure-gradient dependence in the models, and we might conclude, for example, that Spalart-Allmaras can reproduce attached boundary-layer flows with a single set of coefficients better than Baldwin-Lomax can.

This observed coefficient variation also supports our choice of a statistical model with scenario-dependent coefficients. A standard calibration approach is to assume there exists a single best value of model coefficients  $\theta^*$ , and then use all available data to inform this value, with model inadequacy captured by an  $\eta$ -like term only [13]. These results question the validity of that assumption in the case of turbulence closure models.

Figure 8 shows the HPD intervals of the 2nd most sensitive coefficients (or the 1st most sensitive in the case of  $k - \varepsilon$ ). In general coefficients are less well informed and the scenario scatter is greater, supporting the view that optimal closure coefficients are strongly dependent upon the pressure gradient.

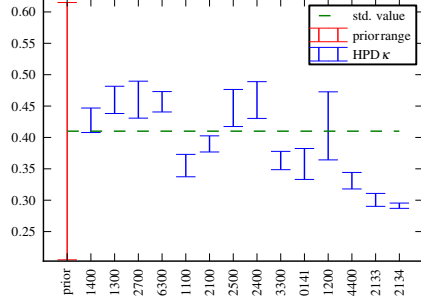
The HPD intervals of the hyper-parameter  $\sigma_{hp}$  give insight into the model error that remains after optimal closure coefficients have been found for each scenario, see Figure 9. HPD interval values close the the 0.1 boundary indicates that the turbulence model has to rely on the  $\eta$  term in order to capture the data better. All models encounter flow cases where this occurs, although some more than others. Especially the Spalart-Allmaras model shows high HPD intervals of  $\sigma_{hp}$  in most cases. This indicates that, of the models in our set  $\mathcal{M}$ , it is worst at matching the experimental velocity profiles exactly. As we shall see in the next section, BMA provides a coherent mechanism for penalizing models that do not perform well on the calibration dataset.

## 5.2. Posterior model probability

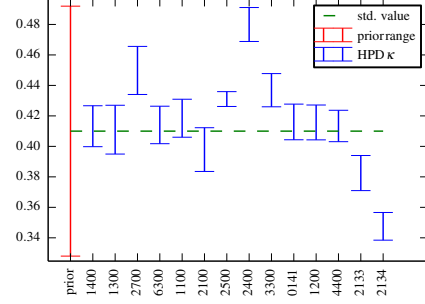
For a given flow case, once we have calibrated each model in the set  $\mathcal{M}$  on the data  $\mathbf{z}_k$  we can calculate the posterior model probability (16). This posterior model probability  $\mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k)$  can be interpreted as the evidence for model  $M_i$  given the available data, and the other models in  $\mathcal{M}$ . It should be emphasized that the posterior model probabilities are conditional on the choice of models in  $\mathcal{M}$ , i.e.  $\mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k, \mathcal{M})$ . Since in our analysis the set  $\mathcal{M}$  remains fixed we drop it from the notation for the sake of brevity.

We compute  $\mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k)$  for each flow case  $k$  in the set  $\mathcal{S}$ , assuming a uniform probability mass function for  $\mathbb{P}(M_i | \mathbf{S}_k)$ . The results can be found in Figure 10. From this figure it becomes clear that also the posterior model probability is dependent upon the applied flow case. Which model has the highest posterior probability given the data can change significantly from one flow case to another. Thus, both the posterior closure coefficients and the posterior model probabilities are functions of the pressure gradient. There is no clear “best”  $\theta$  or  $M_i$  that will outperform its competitors in  $\mathcal{M}$  for every scenario in  $\mathcal{S}$ . For a predictive case that is not in  $\mathcal{S}$  it would be hard to select the best model and coefficient set *a priori*, given that in prediction we do not have access to experimental data by definition.

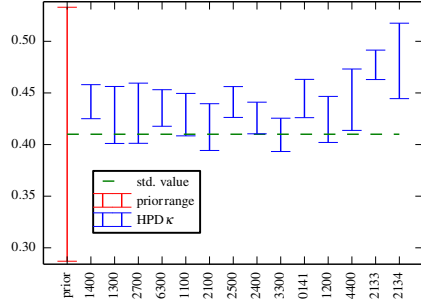
More speculatively, these results could be interpreted as a explanation for why it is that no clear “winner” amongst turbulence closure models has been found, see e.g. [38], and why there are a wide spread of closure coefficients recommended and used in literature and industry. Note in particular that no clear superiority of the more advanced Reynolds-Stress model over the other is noticed, even if it does exhibit somewhat higher probabilities for strong adverse pressure gradient cases. On the other hand, no clear



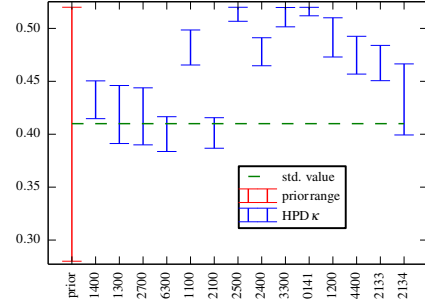
(a)  $\kappa$  of the  $k - \epsilon$  model.



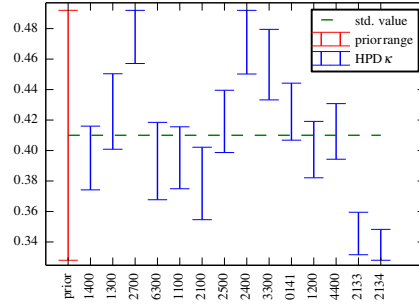
(b)  $\kappa$  of the  $k - \omega$  model.



(c)  $\kappa$  of the Spalart-Allmaras model.

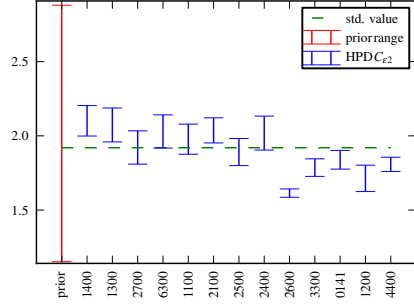


(d)  $\kappa$  of the Baldwin-Lomax model.

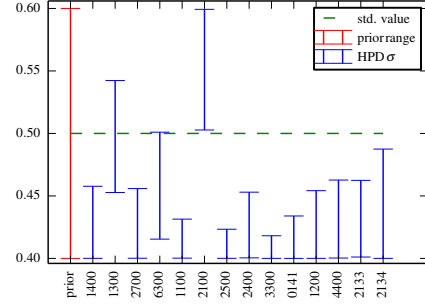


(e)  $\kappa$  of the stress- $\omega$  model.

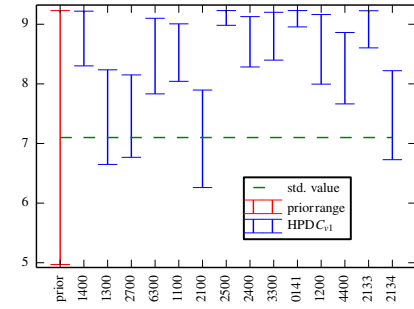
Figure 7: The HPD intervals of  $\kappa$  for all  $(M_i, \mathbf{S}_k)$  combinations.



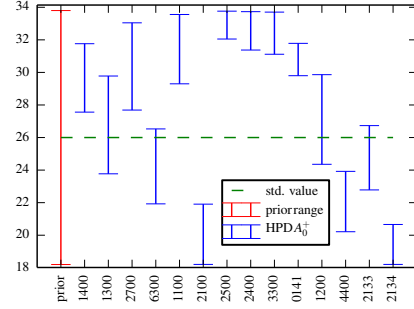
(a)  $C_{\epsilon 2}$  of the  $k - \epsilon$  model.



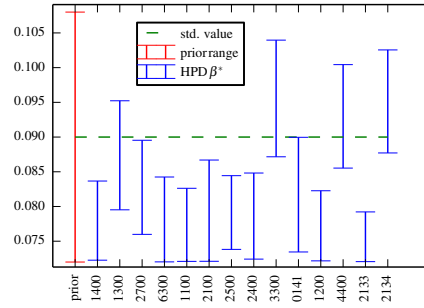
(b)  $\sigma$  of the  $k - \omega$  model.



(c)  $C_{v1}$  of the Spalart-Allmaras model.

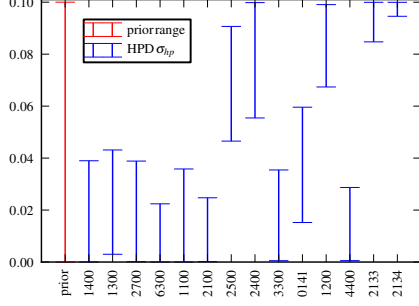


(d)  $A_0^+$  of the Baldwin-Lomax model.

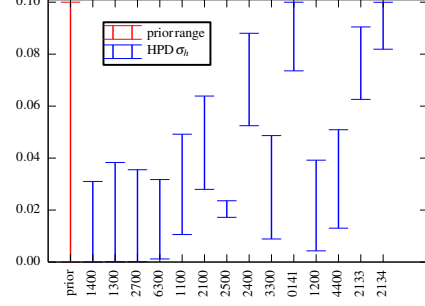


(e)  $\beta^*$  of the stress- $\omega$  model.

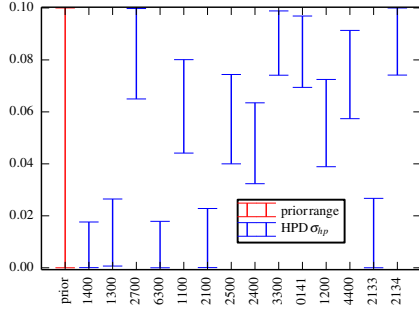
Figure 8: The HPD intervals of second-most well-informed closure coefficients for all  $(M_i, \mathbf{S}_k)$  combinations.



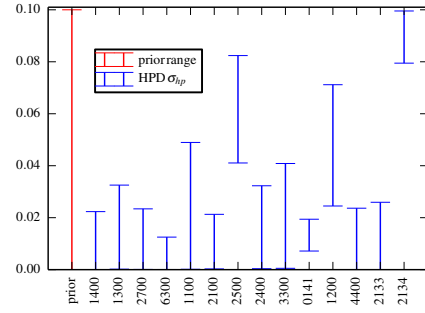
(a)  $\sigma_{hp}$  of the  $k - \epsilon$  model.



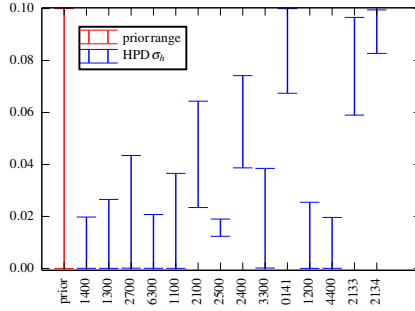
(b)  $\sigma_{hp}$  of the  $k - \omega$  model.



(c)  $\sigma_{hp}$  of the Spalart-Allmaras model.



(d)  $\sigma_{hp}$  of the Baldwin-Lomax model.



(e)  $\sigma_{hp}$  of the stress- $\omega$  model.

Figure 9: The HPD intervals hyper parameter  $\sigma_{hp}$  for all  $(M_i, \mathbf{S}_k)$  combinations.

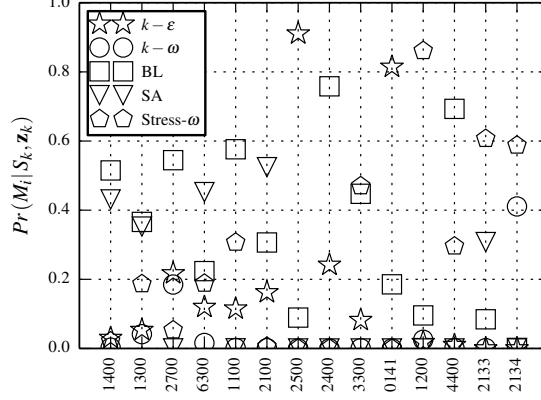


Figure 10: The distribution of  $\mathbb{P}(M_i | \mathbf{S}_k, \mathbf{z}_k)$  for all models and scenarios.

inferiority of the algebraic model is observed either, not even for adverse pressure gradient scenarios. Indeed – based on these results – the practice of tuning a set of deterministic closure coefficients for specific applications seems futile, even if that tuning is done in a rigorous way. Predictions made with existing models and deterministic coefficients seem to be subject to a degree of model error that is not substantially reducible by parameter tuning. For this reason we choose to make model- and scenario-averaged predictions, using our framework of Section 4.2.

### 5.3. Predictions with Bayesian Model-Scenario Averaging

In this section the methodology of Section 4.2 is applied to prediction of boundary-layer flows at pressure gradients for which (nominally) no data is available. Our goal is to assess the predictive capability of (19)-(20). Our predictions are in the form of mean and variance of velocity profiles for the unmeasured flows, that is  $\mathbb{E}(\Delta|\mathbf{z})$  and  $\text{Var}(\Delta|\mathbf{z})$  where  $\Delta = u^+(y^+)$ .

We proceed as follows: one case is selected from the 14 flows in Table 2 as validation flow,  $\mathbf{S}'$ . The experimental data for this case is completely excluded from the analysis, and only used for comparison purposes in the final plots. The remaining cases form the set of calibration scenarios  $\mathcal{S}$ , so that  $K = |\mathcal{S}| = 13$ . Each of these cases is calibrated independently, following Section 4.1, to give the coefficient posteriors per scenario and per model,  $\hat{\theta}_{i,k}$ . The moments  $\mathbb{E}[\hat{\Delta}_{i,k}]$  and  $\text{Var}[\hat{\Delta}_{i,k}]$  are obtained by propagating the  $K \times I$  MCMC chains from the calibration step through (14). The chains are our best available representation of the coefficient posteriors, and therefore no uncertainty information gets lost in this step. Finally (19)-(20) are evaluated.

To close (19) it remains only to specify the probability mass function (pmf) for  $\mathbb{P}(\mathbf{S}_k)$ . This could be based on expert opinion. For instance if one believes that the current prediction flow is subject the mildly-adverse pressure gradients, the posterior parameter distributions that were calibrated under similar circumstances could be favored through  $\mathbb{P}(\mathbf{S}_k)$ . Ideally an automatic procedure is preferred, or if no information is available one must admit ignorance and select a uniform pmf. In the following sections first predictions

based on this last choice are examined as a worst-case; then a smart automated selection of  $\mathbb{P}(\mathbf{S}_k)$  is developed

### 5.3.1. Results of uniform $\mathbb{P}(\mathbf{S}_k)$

In the worst-case we are completely ignorant of the relationship between  $\mathbf{S}'$  and the members of  $\mathcal{S}$ , and select a uniform pmf for  $\mathbb{P}(\mathbf{S}_k)$ , i.e. assuming that each pressure-gradient scenario is equally likely. We show the results for two validation flow cases in Figure 11, an “easy” case (flow 1300), and a “difficult” case subject to a strongly adverse pressure gradient (flow 4400). The mean prediction for case 1300 in Figure 11(a) is reasonable, it falls within the region of experimental uncertainty for all  $y^+$ . However the standard deviation of the prediction  $\sigma_\Delta := \sqrt{\text{Var}[\Delta | \mathbf{z}]}$  is almost  $2 u^+$  units. As can be seen from the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  quantiles, plotted in blue, this is very large, and a clear over-estimation of the true model error. Nonetheless the fact that the correct solution lies within the range of our prediction should be interpreted as a successful prediction – albeit perhaps not a very useful one.

Thanks to the variance decomposition in (20), the source of the large variance can be diagnosed, see Figure 11(b). It seems the contributions from all three sources are of the same order, but dominated somewhat by the *between scenario* variance. It is exactly this variance that stands to be reduced most by smart weighting of scenario probabilities.

The story for case 4400 is essentially the same, where again the variance is a substantial over-estimation.

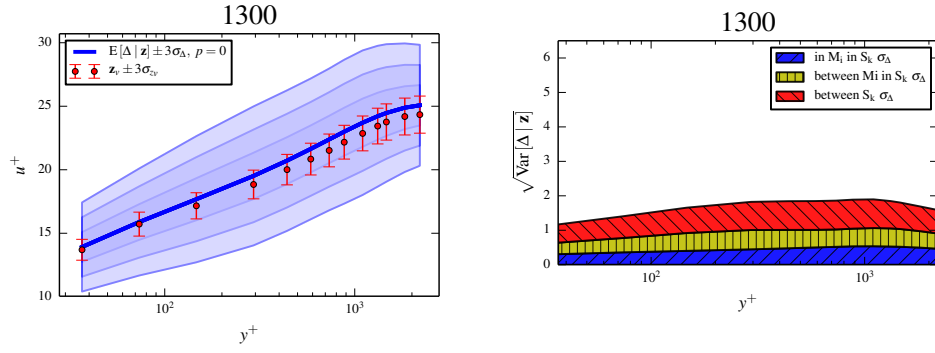
One could argue that it is better to be conservative in predictions of uncertainty in the QoI, however the amount of uncertainty should also not be so high as to make the prediction useless. We are of the opinion that the amount of uncertainty (in these relatively simple flows) crosses that line. For instance the results for flow case 4400 show an uncertain maximum  $3\sigma_\Delta$  range of  $\pm 15 u^+$  units in the defect layer, roughly 43% of the mean.

Ultimately the source of this large variance is the large spread of closure coefficient values across the calibration cases. As more extreme cases are added to  $\mathcal{S}$  – with the goal of extending the range of validity of the predictions – this variance will only increase. This result echoes our previous results using p-boxes in Edeling et al. [8] which also lead to very large error bars. Unlike p-boxes however, our Bayesian Scenario Averaging framework is not limited to a uniform scenario weighting.

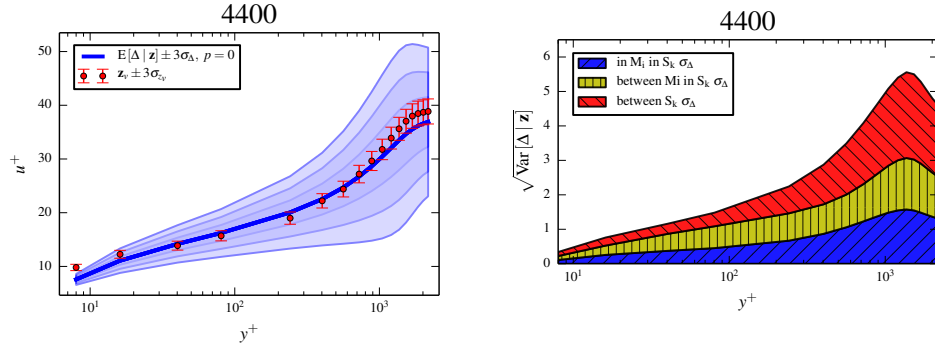
### 5.3.2. Results of smart $\mathbb{P}(\mathbf{S}_k)$ weighting

In order to bring  $\sigma_\Delta$  down to realistic levels we use the model prediction-spread based  $\mathbb{P}(\mathbf{S}_k)$  defined in (21). To clarify, we plot the model prediction spread for validation case 1300 in Figure 12. This corresponds to the envelope of predictions provided by the different models, using posterior coefficients associated to a particular calibration scenario. Figure 12 displays such a prediction spread for different calibrations scenarios. Note that the spread is quite small for predictions made with posteriors coming from cases similar to 1300, and that it steadily grows as we move to the predictions made with  $\hat{\theta}_{i,k}$  calibrated under strongly-adverse  $\mathbf{S}_k$ . It is this behavior that is used to inform  $\mathbb{P}(\mathbf{S}_k)$  via  $\mathcal{E}_k$ , and the effect is amplified by selecting a  $p \geq 1$ . The resulting scenario weights for  $p = 1, 4, 8$  are shown in 13. Notice that in both cases the sensor favours cases with pressure-gradients similar to the validation scenario. For case 1300 it selects favorable





(a) The BMSA prediction for validation case 1300. (b) The BMSA standard deviation for validation case 1300.



(c) The BMSA prediction for validation case 4400. (d) The BMSA standard deviation for validation case 4400.

Figure 11: Selected BMSA predictions with uniform  $\mathbb{P}(\mathbf{S}_k)$ .

pressure-gradients, and for case 4400 adverse gradients. In both cases with  $p = 8$  most scenarios have an almost-zero weight and can be neglected.

We now examine how this new  $\mathbb{P}(\mathbf{S}_k)$  affects the BMSA predictions, for different values of  $p$ . In Figure 14a we plot  $\mathbb{E}[\Delta|\mathbf{z}]$  for  $p = 0, 1, \dots, 8$  for case 1300. We see that for  $p = 0, 1$  the predictions are slightly off the experimental data. For all  $p > 1$  the predictions overlap and match the data quite well. However, in most cases we found that the value of  $p$  had no significant effect on the BMSA prediction  $\mathbb{E}[\Delta|\mathbf{z}]$ , see e.g. Figure 14b. The posterior variance (20) on the other hand is significantly affected by  $p$ . This can be inferred from Figure 15, where we show the same predictions with quantified uncertainty as in the previous section, but using the smart weighting (with  $p = 8$ ). As discussed, the BMSA prediction mean for 1300 is improved, and now lies right on top of the measurement points. In addition the maximum  $\sigma_\Delta$  value has more than halved from 2.0 to 0.8. It is now comparable to the level of measurement error indicating that our prediction is as accurate as our reference experimental data. For 4400 the prediction with  $p = 0$  was already reasonable, and remains unchanged for  $p > 0$ . Again, the maximum  $\sigma_\Delta$  has decreased, in this case from 5.5 to roughly  $3.7 u^+$ -units.

From Figures 15(b) and (d), we observe the variance reduction was most strong for *between scenario* variance as expected, but *between model, within scenario* variance also took a substantial hit. The remaining variance is dominated by the term representing the lack of identification of closure coefficients in the individual calibrations, i.e. the variance of  $\hat{\boldsymbol{\theta}}_{i,k}$ . This can be reduced by using more data (and more informative data) in the calibrations, to better inform coefficients. Alternatively, approximating each  $\hat{\boldsymbol{\theta}}_{i,k}$  by its maximum *a posteriori* (MAP) estimate, would reduce the computational cost of evaluating  $\mathbb{E}(\hat{\Delta}_{i,k})$ , and set the blue region in the above plots to zero. Given that the blue region dominates the variance however, this is an approximation of questionable validity.

Finally, to demonstrate the generality and robustness of these results, the above analysis is performed for prediction of each one of the 14 boundary-layer cases, each time using data from the remaining 13 cases only. The results are summarized in Table 3. There we show the set of  $\mathbf{S}_k$  which make up 75 % or more of  $\mathbb{P}(\mathbf{S}_k)$ , the relative error of the BMSA prediction (19) defined as  $\epsilon'_\infty := \|\mathbb{E}[\Delta|\mathbf{z}] - \mathbf{z}_v\|_\infty / \|\mathbf{z}_v\|_\infty$  and the coefficient of variation (CoV), given by  $C_{cov} := \sigma_\Delta / \mathbb{E}[\Delta|\mathbf{z}]$ . The CoV is a measure of the solution variability with respect to the mean. Note that most validation cases favour a  $\mathbf{S}_k$  with a pressure-gradient classification similar to their own, only the mild- and moderately adverse  $\mathbf{S}'$  tend to mix different type of scenarios. Also, almost all BMSA predictions are satisfactory, which can be inferred from the fact that they have both small  $\epsilon'_\infty$  and max  $C_{cov}$ . The only real exceptions are cases 2133 and 2134. These are cases with an extremely adverse pressure gradients, for which all turbulence models in our set  $\mathcal{M}$  struggle to make accurate predictions.

#### 5.4. $C_f$ prediction

Up this point the QoI has always been  $u^+$  profiles, and all posterior  $\hat{\boldsymbol{\theta}}_{i,k}$  have been informed using only experimental  $u^+$  data. The question now arises whether these distributions are also effective when we try to predict a QoI of a different nature. To that end we make predictions for a variety of different skin-friction profiles, using the same procedure, distributions and posterior model probabilities as in the preceding sec-

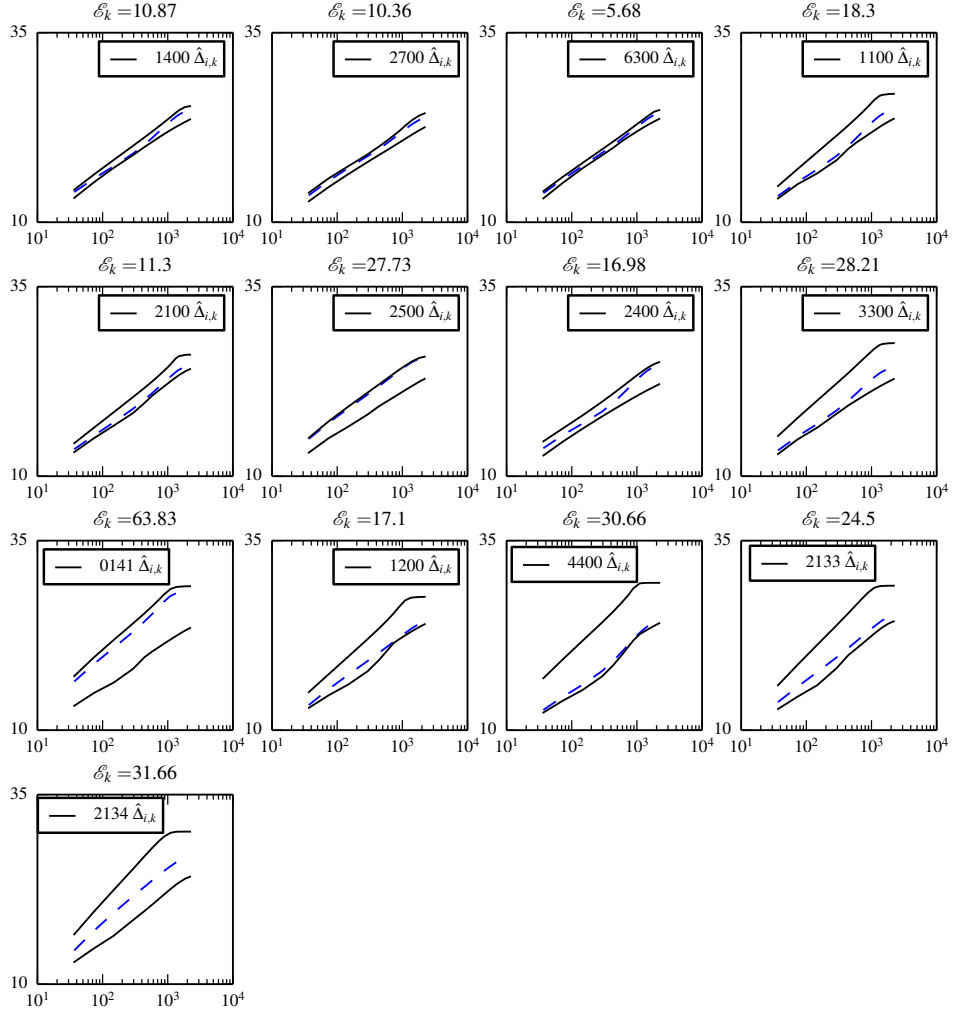


Figure 12: The model prediction spread for validation case 1300. Each subplot shows the envelope (solid lines) formed by the 5  $\hat{\Delta}_{i,k}$ ,  $i = 1, \dots, 5$ , as well as the BMA prediction (dotted line) for a given  $\mathbf{S}_k$ .  $\mathcal{E}_k$  is a measure for the spread as defined in (21). The unit of all vertical axes is  $u^+$ , and the horizontal axes display  $y^+$  units.

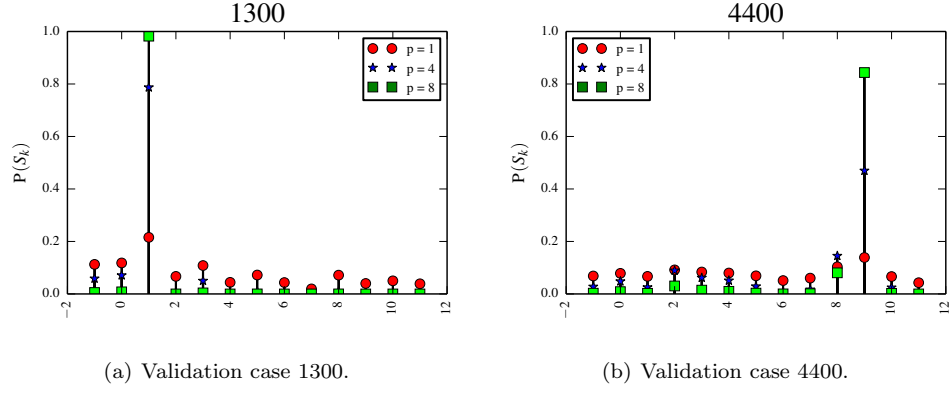


Figure 13: Two  $\mathcal{E}_k$ -based  $\mathbb{P}(\mathbf{S}_k)$  distributions.

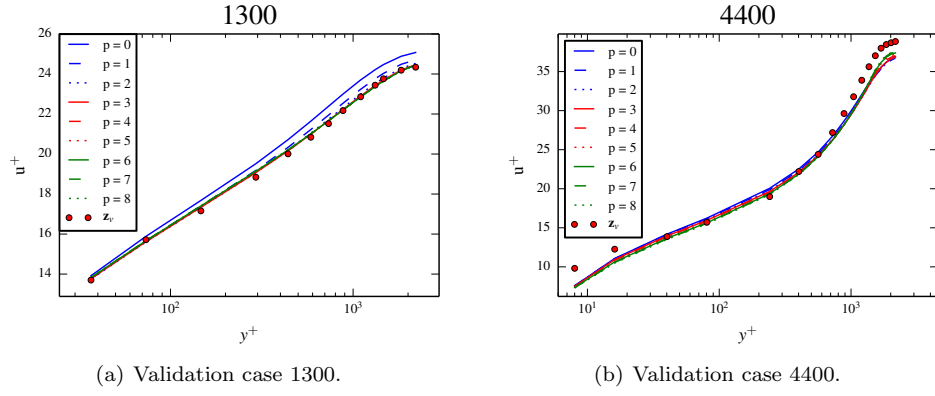
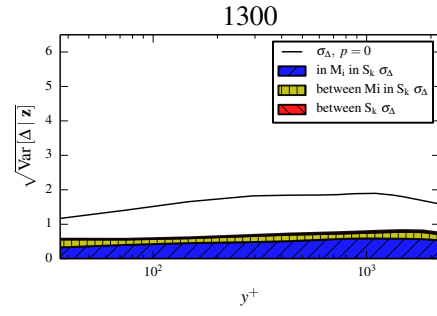
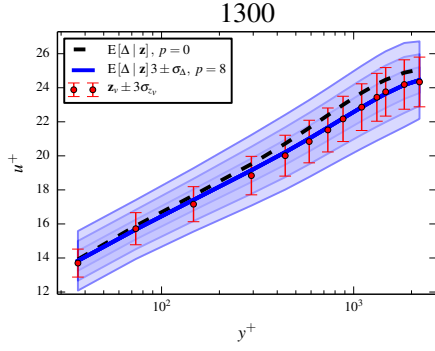
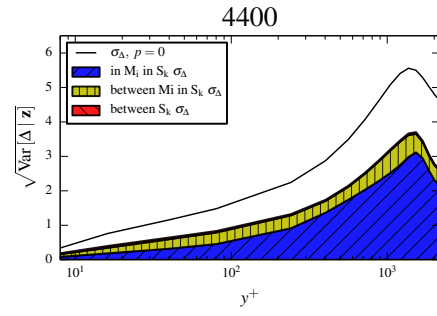
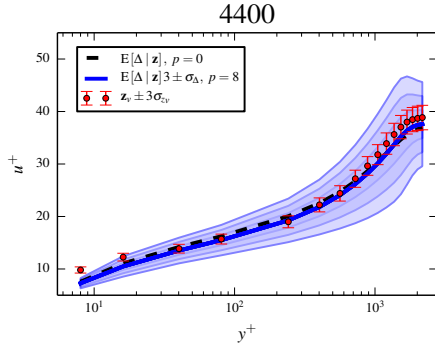


Figure 14:  $\mathbb{E}[\Delta|\mathbf{z}]$  vs  $p$  for two validation flow cases.



(a) The new and old BMSA prediction for validation case 1300. (b) The new BMSA standard deviation for validation case 1300.



(c) The new and old BMSA prediction for validation case 4400. (d) The new BMSA standard deviation for validation case 4400.

Figure 15: Selected BMSA predictions with a  $\mathcal{E}_k$ -based  $\mathbb{P}(\mathbf{S}_k)$ .

Table 3: Results validation flow cases using (21).

Val. case	$\{\mathbf{S}_k   \sum_k \mathbb{P}(\mathbf{S}_k) \geq 0.75\}$	$\epsilon'_\infty$	$\max C_{cov}$
1400 (zero)	6300	0.030	0.042
1300 (fav)	6300	0.013	0.042
2700 (fav)	6300, 1300	0.030	0.058
6300 (fav)	1300	0.011	0.045
1100 (mild adv)	1200, 2100, 6300	0.062	0.061
2100 (mild adv)	6300, 1400	0.028	0.055
2500 (mild adv)	1100, 1400, 6300	0.039	0.057
2400 (mod adv)	2500, 1100, 2700, 1400	0.013	0.067
3300 (mod adv)	1100	0.046	0.065
0141 (strong adv)	4400	0.051	0.081
1200 (strong adv)	1100, 2133, 2100	0.043	0.091
4400 (strong adv)	1200	0.050	0.108
2133 (very strong adv)	0141, 4400, 1200	0.14	0.099
2134 (very strong adv)	0141, 4400, 1200	0.29	0.099

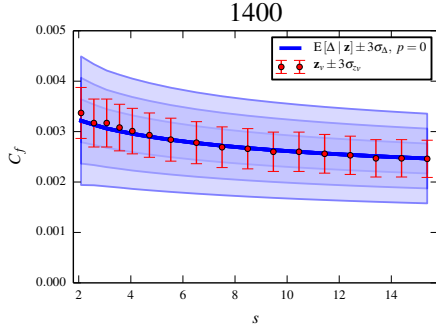
tions. Thus, the only change is our new QoI, the dimensionless skin-friction coefficient  $C_f := \tau_w / (\rho u_\infty^2 / 2)$ . Here  $\tau_w$  is the wall-shear stress and  $u_\infty$  is the freestream velocity.

The results for a uniform  $\mathbb{P}(\mathbf{S}_k)$  distribution and a large range of different scenarios are shown in Figure 16. We see that all predictions fall within the range of experimental uncertainty, and in most cases are quite well centered on the data points. Only the 3300 prediction is slightly off center. The amount of uncertainty in the predictions is reasonable,  $\mathbb{E}[\Delta|\mathbf{z}] \pm \sigma_\Delta$  being in most cases contained within the region of experimental uncertainty.

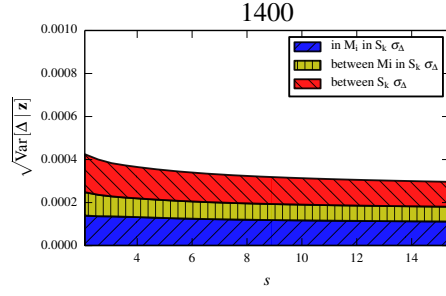
As in the case of the  $u^+$  predictions, we now examine the effect of our scenario-weighting procedure, starting with the influence of  $p$  on  $\mathbb{E}[\Delta|\mathbf{z}]$ . In Figure 17a we plot  $\mathbb{E}[\Delta|\mathbf{z}]$  vs  $p$  on the same vertical scale as used in Figure 16. It is again clear that the BMSA predictions are robust, as they are relatively invariant considering the large range of the exponent  $p$ . When zooming in on  $\mathbb{E}[\Delta|\mathbf{z}]$  (see Figure 17b), we see that unlike the  $u^+$  case, the predictions move slightly away from the data with increasing  $p$ , but basically become insensitive to  $p$  starting from  $p = 3$ . When we examine the model prediction spread for 0141 in Figure 18, it becomes clear that predictions can be more in agreement with each other for flow cases not similar to 0141, which underlines the importance of considering different calibration scenarios. Also, including posterior distributions calibrated for QoIs other than  $u^+$  could possibly be beneficial. The prediction with quantified uncertainty for the same validation case and  $p = 3$  is shown in Figure 19. We can again see that the variance is affected more by  $p$  than the prediction.

### 5.5. Reduction of computational effort - Scenario-Averaged Posteriors

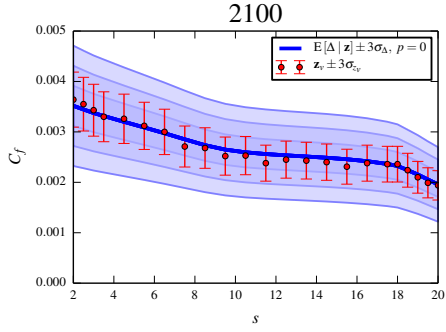
The full BMSA approach, with  $I$  models and posterior distributions coming from  $K$  scenarios requires us to propagate  $I \times K$   $p(\hat{\boldsymbol{\theta}}_{i,k})$  through the code for  $\Delta$ . In order to use BMSA, only convergence in mean and variance is required for each propagation. This poses no problem for our current computational setup. Full propagation can be



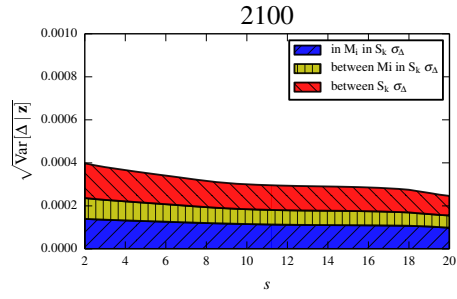
(a) The BMSA prediction for case 1400.



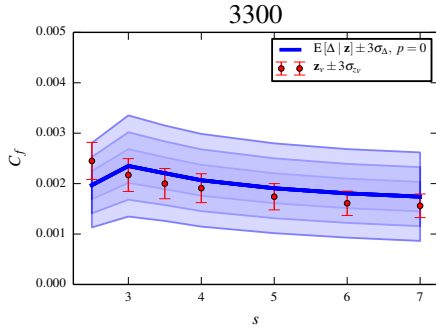
(b) The BMSA std. dev. for case 1400.



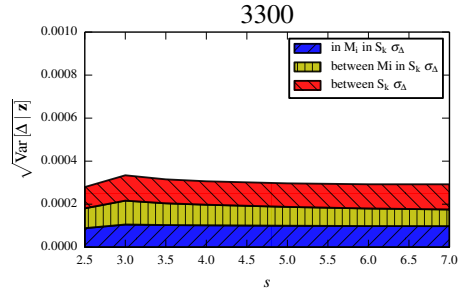
(c) The BMSA prediction for case 2100.



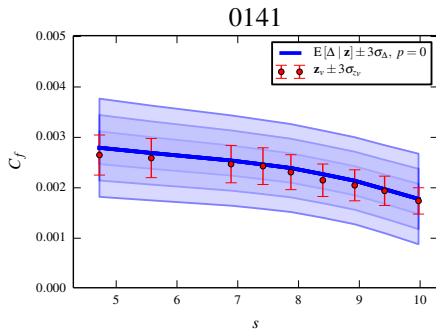
(d) The BMSA std. dev. for case 2100.



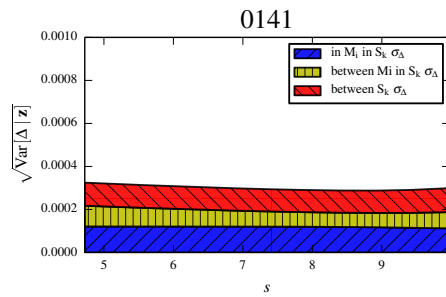
(e) The BMSA prediction for case 3300.



(f) The BMSA std. dev. for case 3300.



(g) The BMSA prediction for case 0141.



(h) The BMSA std. dev. for case 0141.

Figure 16: Selected BMSA predictions for  $C_f$  with uniform  $\mathbb{P}(\mathbf{S}_k)$ .

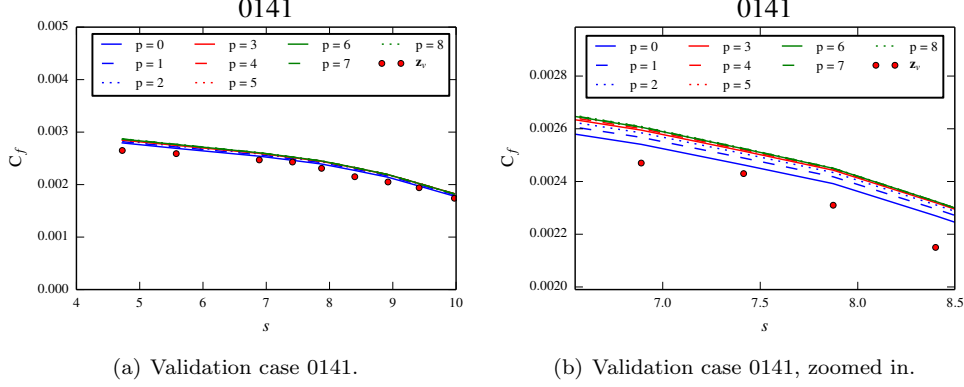


Figure 17:  $\mathbb{E}[\Delta|\mathbf{z}]$  vs  $p$ . Here,  $\Delta = C_f$ .

achieved within a single day with moderate computational resources, in our case parallel propagation using 14 cores. Calibration takes longer, but is still easily achieved. A single calibration can take anywhere from 1 hour up to (at the very most) 12 hours, depending on the turbulence model, the streamwise location of the data, and the number of specified steps in the Markov chain. With the mentioned computational resources, all flow-case model combinations can be calibrated in roughly a week time. This needs to be performed only once, after which the posteriors can be re-used (also for complex flow topologies) as much as needed in the predictive phase of the method.

However, the required propagation by BMSA using the same number of models and scenarios is unlikely to be computationally tractable when considering more complex flow topologies. This could be alleviated, e.g. through the use of surrogate models [37, 21, 23], and will be explored in future work. But even with the use of surrogate models, performing  $I \times K$  propagations could prove to be computationally expensive. Therefore we propose to replace the  $I \times K$   $p(\tilde{\theta}_{i,k})$  by  $I$  scenario-averaged  $p(\tilde{\theta}_i)$  defined as

$$p(\tilde{\theta}_i) := p(\theta|M_i, \mathbf{z}) = \sum_{k=1}^K p(\theta|M_i, \mathbf{S}_k, \mathbf{z}) \mathbb{P}(\mathbf{S}_k). \quad (23)$$

Instead of using (19), the prediction for  $\Delta$  is now made with

$$\mathbb{E}[\Delta|\mathbf{z}] = \sum_{i=1}^I \mathbb{E}[\tilde{\Delta}_i] \mathbb{P}(M_i|\mathbf{z}), \quad (24)$$

where  $\mathbb{E}[\tilde{\Delta}_i] := \mathbb{E}[\Delta|M_i, \mathbf{z}]$  is obtained by propagating  $p(\tilde{\theta}_{i,k})$  through  $\Delta$  and

$$\mathbb{P}(M_i|\mathbf{z}) = \sum_{k=1}^K \mathbb{P}(M_i|\mathbf{S}_k, \mathbf{z}) \mathbb{P}(\mathbf{S}_k). \quad (25)$$



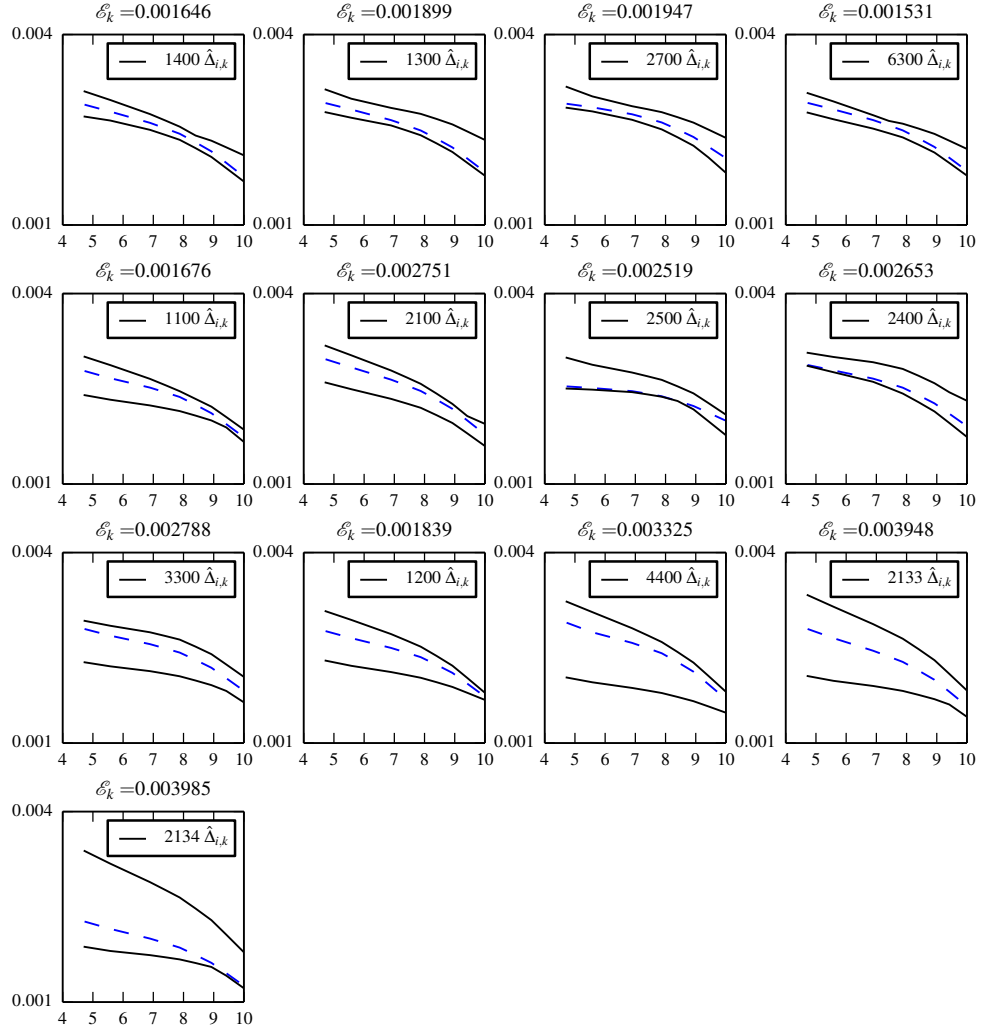


Figure 18: The model prediction spread of  $C_f$  for validation case 0141. Each subplot shows the envelope (solid lines) formed by the 5  $\hat{\Delta}_{i,k}$ ,  $i = 1, \dots, 5$ , as well as the BMA prediction (dotted line) for a given  $\mathbf{S}_k$ .  $\mathcal{E}_k$  is a measure for the spread as defined in (21). The vertical axes all display the skin-friction coefficient  $C_f$ , while the horizontal axes show the distance  $s$ .

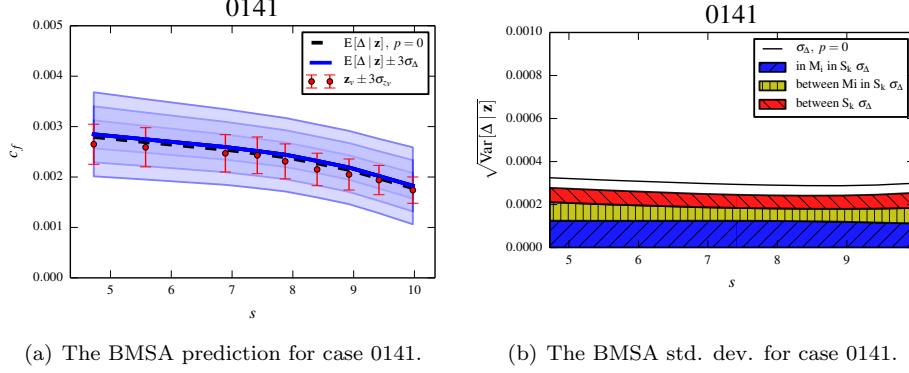


Figure 19: BMSA prediction for  $C_f$  with  $\mathcal{E}_k$ -based  $\mathbb{P}(\mathbf{S}_k)$  and  $p = 3$ .

Furthermore, the posterior variance (20) now reduces to

$$\text{Var}[\Delta | \mathbf{z}] = \sum_{i=1}^I \text{Var}[\tilde{\Delta}_i] \mathbb{P}(M_i | \mathbf{z}) + \sum_{i=1}^I \left( E[\tilde{\Delta}_i] - \mathbb{E}[\Delta | \mathbf{z}] \right)^2 \mathbb{P}(M_i | \mathbf{z}). \quad (26)$$

The first of two terms in (26) is denoted as the *in-model variance*, and the second one is called the *between-model variance*. Note that (24)-(26) is not a traditional BMA, as the influence of the different  $\mathbf{S}_k$  is still implicitly present. Of course, propagating scenario-weighted posteriors is likely to provide different results with respect to a full BMSA using the same scenario weights, if significant nonlinear interaction effects are present. However, the computational effort is reduced significantly, since now only  $I$  propagations are required to compute the prediction and its variance. Scenario probabilities  $\mathbb{P}(\mathbf{S}_k)$  now have to be specified *before* we propagate, and thus we lose the ability of using an automated sensor such as (21).

Instead, we must rely on expert opinion to specify the distribution of the  $\mathbf{S}_k$ . As proof of concept, let us imagine an expert RANS user, interested in predicting  $C_f$  for a favourable pressure gradient case. Furthermore suppose this user selects from the closure-coefficient database two favourable (1300, 2700), one zero (1400) and one mildly-adverse (1100)  $\mathbf{S}_k$ , and sets for each of these scenarios  $\mathbb{P}(\mathbf{S}_k) = 0.25$ . Next, for each model one  $p(\tilde{\theta}_i)$  is created via (23). As an example we show the marginals of  $p(\tilde{\theta}_i)$  for the  $k - \varepsilon$  model in Figure 20. The prediction (24) and standard deviation of (26) are shown in Figure 21. To contrast, the full BMSA prediction and standard deviation are plotted as well. Although the full BMSA prediction lies closer to the validation data, the  $\mathbb{E}[\Delta | \mathbf{z}]$  computed with (24) is not far off. This is especially true for the two  $\sigma_\Delta$ , which almost lie on top of each other. Finally, when examining Figure 21b we see that through  $\text{Var}[\tilde{\Delta}_i]$ , most of the between-scenario uncertainty has now been incorporated into the in-model standard deviation, which is significantly larger than the between-model  $\sigma_\Delta$ .

### 5.6. Discussion - Closure Coefficient Database

The full BMSA approach described would be clearly extremely computationally expensive if applied to complex flow problems. We have deliberately made no simplifications

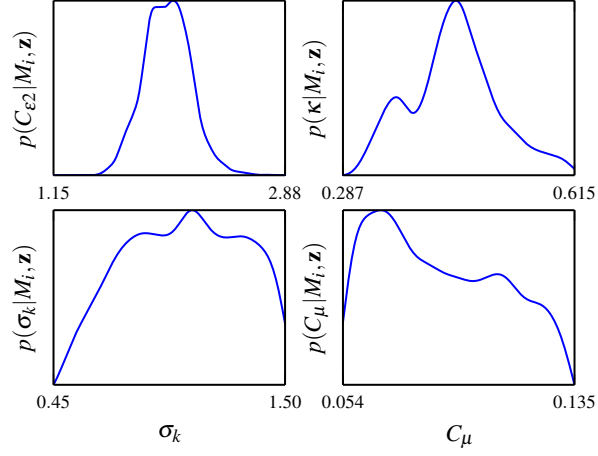
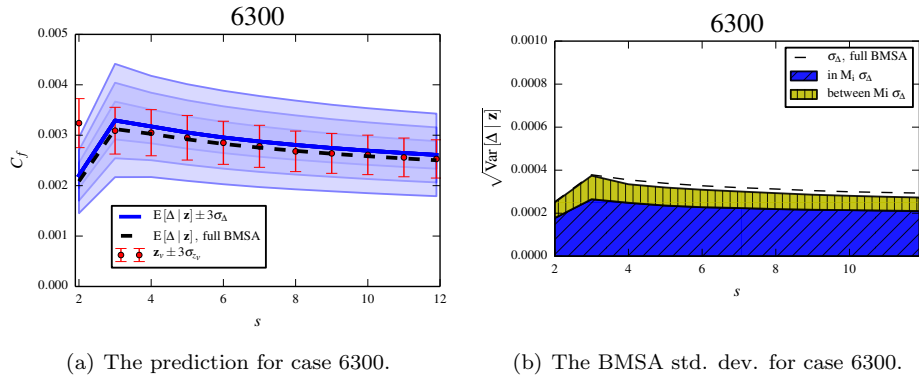


Figure 20: The marginals of  $p(\tilde{\theta}_i)$  in case of the  $k - \epsilon$  model.



(a) The prediction for case 6300.

(b) The BMSA std. dev. for case 6300.

Figure 21: Prediction for  $C_f$  obtained using 5 scenario-averaged posterior distributions. The full BMSA results are depicted by the dotted lines.

– such as using MAP estimators, moment methods, or surrogate models – in order to assess the performance of the approach in the best-case scenario. Clearly, application to complex flow topologies requires some performance optimization. For instance, the fact that (21) usually picks  $\mathbf{S}_k$  similar to  $\mathbf{S}'$  suggests that we might set certain  $\mathbb{P}(\mathbf{S}_k)$  to 0 by hand based upon physical reasoning, saving code runs. More specifically, this approach would deduct  $I$  from the total number of required propagations. Another option with a more significant reduction of computational effort was outlined in Section 5.5, i.e. the use of scenario-averaged posterior distributions. This particular method requires the user to *a priori* specify  $\mathbb{P}(\mathbf{S}_k)$ , which could also be done based upon physical reasoning. This  $\mathbb{P}(\mathbf{S}_k)$  can then be used to create one  $\mathbf{S}_k$ -averaged posterior coefficient distribution per model, reducing the required number of propagations to  $I$ . Initial results look promising, showing only a small deviation from the full BMSA prediction. Such a strategy could allow  $\mathcal{S}$  to contain many hundreds of scenarios, as it is the user who decides which of these to incorporate into  $p(\tilde{\boldsymbol{\theta}}_i)$ . In addition to different scenarios, we envision the closure coefficient database to contain posteriors calibrated for different QoIs as well. So far we have only used  $u^+$  data to inform our posteriors. Our results indicate that we can still use these distributions to make robust predictions of  $C_f$  profiles, but some of the optimality is lost compared to the case where  $u^+$  is the QoI. Enriching the database with distributions calibrated, amongst others, for  $C_f$  profiles could possibly remedy this. Also it would also provide a user with a more natural choice of which  $\mathbf{S}_k$  to include, plus it could inform certain coefficient distributions that are just not informed by  $u^+$  data. Posterior distributions that were calibrated using, e.g., both  $u^+$  and  $C_f$  data could be included as well.

For many propagation techniques, e.g. [37], the computational cost increases with the number of uncertain coefficients. Therefore, further gains could come by reducing the dimensionality of the problem via a Sobol analysis. For instance our results of Section 3.3 show that for boundary-layer flows there are only 2 influential parameters for each model under consideration. If in other flow topologies a similar situation occurs, we might fix the non-influential parameters.

The hope is that this framework could be equally successful when applied to more general and complex classes of flows. There are two approaches here. As described earlier, since closure coefficient posteriors are topology independent, we can apply the “coefficient-database” from our limited class of flows to general flow problems. Then only uncertainty propagation needs to be performed for the complex problem to obtain predictions. Many techniques are available for efficiently propagating pdfs through a computer code, notably sparse-grid stochastic-collocation methods [37, 21, 23]. In the worst-case, MAP estimators can be used instead of full pdfs. The success of this approach will depend on to what extent the features in the complex flow are represented in the database.

The second approach is to build a new database  $\mathcal{S}$ , tailored to the class of flows to be predicted. This would entail performing new calibrations. If possible, these calibrations should be done using cheap codes, and subsequently again use the topology-independent nature of  $\boldsymbol{\theta}$ . If this is not possible, Bayesian calibration could also be performed using a surrogate model as a replacement for the expensive computer code, see e.g. [17]. The scenarios in the new set should be (a) representative of this class, and (b) have rich experimental data available. This need not include only (or even mainly) flows of the same type – for wing aerodynamics, flat-plate boundary-layers are likely to be relevant.

Models performing poorly – as judged by  $\mathbb{P}(M_i|\mathbf{S}_k, \mathbf{z}_k)$  – could be culled from the set to reduce the effort needed in the coming predictions. The calibration step must only be performed once to set up the database. After which predictions are again a matter of uncertainty propagation only.

To summarize, we remain somewhat within the classical paradigm of RANS calibration, in that we calibrate on simple flow problems, the result of which we will apply to more complex problems. However, the classical paradigm: i) uses one point estimate of  $\boldsymbol{\theta}$  and subsequently assumes generality of this estimate, taking no form of parameter error into account; ii) does not account for model-form error. To stochastically model these errors, we build a database of posterior distributions for several alternative turbulence models, which could be calibrated under a wide range of different scenarios and/or QoIs. If computationally allowable, a full BMSA could be performed to obtain predictions with quantified uncertainty, where the results coming from certain posteriors can be automatically favoured through a sensor like (21). If not, posteriors could be averaged over a weighted set of user-selected scenarios. This approach requires the construction of only one surrogate model for each turbulence model in  $\mathcal{M}$ , increasing its range of applicability compared to the full BMSA.

## 6. Conclusion

We performed Bayesian calibrations of the closure coefficients of 5 turbulence models: the Launder-Sharma  $k - \varepsilon$ , Wilcox  $k - \omega$ , Spalart-Allmaras and Baldwin-Lomax model and Wilcox stress- $\omega$ . Each model was calibrated for 14 scenarios consisting of flat-plate boundary-layers at varying pressure-gradients, using experimental velocity measurements. Substantial variation in closure coefficients for all models was observed across this large range of simple flow cases.

In order to synthesize these results, and make predictions for unmeasured flows, we utilized Bayesian Model Averaging, with an emphasis on scenarios, i.e. Bayesian Scenario Averaging. The framework requires the computation of the posterior model probabilities, which can be thought of as a measure of consistency that a model has with the experimental data. We found that, like the closure coefficients, the posterior model probabilities vary greatly with the applied pressure-gradient scenario. These results suggest that there is no single best choice of turbulence model or closure coefficients, and no obvious way to choose an appropriate model *a priori*.

Instead we used BMSA to make stochastic predictions of unmeasured velocity profiles. Closing BMSA with uniform scenario weighting produced predictions which matched reference data, but with excessively large variance. We therefore developed a smart scenario sensor, to automatically preferentially weight those scenarios in the calibration set that are similar to the prediction case. This resulted in substantially improved predictors, both in terms of mean and variance. For almost all of 14 prediction cases, the predictions lie within one standard deviation of the experimental validation data, and the variance was of the same order as the experimental measurement error.

Despite the fact that only experimental  $u^+$  data was used to inform our posterior distributions, we also used their topology independent nature to make predictions for skin-friction profiles. Although these posterior distributions were not optimal for this particular quantity of interest, we were still able to make consistently more robust predictions than can be expected from a single turbulence model.

Finally, an approach was suggested to significantly reduce the computational effort required by the predictive phase of our methodology. By averaging the posterior distributions over different scenarios, the number of pdf propagations drops from  $I \times K$  to just  $I$ . Here,  $I$  is the number of turbulence models and  $K$  the number of scenarios. Even though expert opinion is now needed to specify the scenario weighting, initial results look promising.

Future work will involve the application of our method to more complex flow topologies. We see the methods developed here as a step towards providing estimates of the error due to turbulence closure models, and thereby lending a robust predictive power to RANS.

**Acknowledgements:** The present work was supported by the French 'Agence Nationale de la Recherche' (ANR) under contract ANR-11-MONU-008-002.

- [1] B. S. Baldwin and H. Lomax. *Thin layer approximation and algebraic model for separated turbulent flows*, volume 257. American Institute of Aeronautics and Astronautics, 1978.
- [2] M.H. Chen and Q.M. Shao. Monte carlo estimation of bayesian credible and hpd intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92, 1999.
- [3] S.H. Cheung, T.A. Oliver, E.E. Prudencio, S. Prudhomme, and R.D. Moser. Bayesian uncertainty analysis with applications to turbulence modeling. *Reliability engineering and systems safety*, 96(9):1137–1149, 2011.
- [4] F.H. Clauser. Turbulent boundary layers in adverse pressure gradients. *J. Aeron. l Sci.*, pages 91–108, 1954.
- [5] D.E. Coles and E.A. Hirst. Computation of turbulent boundary layers. In *Proceedings of AFOSR-IFP Stanford Conference*, volume 2, 1968.
- [6] D. Draper. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 45–97, 1995.
- [7] R. Dwight and Z.H. Han. Efficient uncertainty quantification using gradient-enhanced Kriging. In *Proceedings of 11th AIAA Conference on Non-Deterministic Approaches, Palm Springs CA. AIAA-2009-2276*, pages 1–21. AIAA, 2009.
- [8] W.N. Edeling, P. Cinnella, R.P. Dwight, and H. Bijl. Bayesian estimates of parameter variability in the  $k-\epsilon$  turbulence model. *Journal of Computational Physics*, 258C:73–94, 2014.
- [9] M. Emory, J. Larsson, and G. Iaccarino. Modeling of structural uncertainties in reynolds-averaged navier-stokes closures. *Physics of Fluids (1994-present)*, 25(11):–, 2013.
- [10] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. Bayesian data analysis. 2004. *Chapman&Hall, London*, 2, 2004.
- [11] W.K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [12] J.A. Hoeting, D. Madigan, A.E. Raftery, and C.T. Volinsky. Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401, 1999.
- [13] M.C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
- [14] SJ Kline. The 1980-81 afosr-htm-stanford conference on complex turbulent flows: Comparison of computation and experiment. In *Objectives, evaluation of data, specifications of test cases, discussion and position papers*, volume 1, 1981.
- [15] B.E. Launder and B.I. Sharma. Application of the energy-dissipation model of turbulence to the calculation of flow near a spinning disc. *Letters Heat Mass Transfer*, 1:131–137, 1974.
- [16] G.J.A. Loeven, J.A.S. Witteveen, and H. Bijl. Probabilistic collocation: an efficient non-intrusive approach for arbitrarily distributed parametric uncertainties. In *Proceedings of the 45th AIAA Aerospace Sciences Meeting*, volume 6, pages 3845–3858, 2007.
- [17] Youssef Marzouk and Dongbin Xiu. A stochastic collocation approach to bayesian inference in inverse problems. *Communications in Computational Physics*, 6:826–847, 2009.
- [18] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, et al. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087, 1953.

- [19] P.D. Meyer, M. Ye, M.L. Rockhold, S.P. Neuman, and K.J. Cantrell. Combined estimation of hydrogeologic conceptual model, parameter, and scenario uncertainty with application to uranium transport at the hanford site 300 area. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2007.
- [20] A. Mira, R. Solgi, and D. Imparato. Zero variance markov chain monte carlo for bayesian estimators. *Statistics and Computing*, 23(5):653–662, 2013.
- [21] F. Nobile, R. Tempone, and C.G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.
- [22] T.A. Oliver and R.D. Moser. Bayesian uncertainty quantification applied to rans turbulence models. In *Journal of Physics: Conference Series*, volume 318, page 042032. IOP Publishing, 2011.
- [23] R. Pecnik, J.A.S. Witteveen, and G. Iaccarino. Uncertainty quantification for laminar-turbulent transition prediction in rans turbomachinery applications. *AIAA Paper 2011*, pages 1–14, 2011.
- [24] P.D.A. Platteeuw, G.J.A. Loeven, and H. Bijl. Uncertainty quantification applied to the k-e model of turbulence using the probabilistic collocation method. In *AHS Adaptive Structures Conference; 10 th AIAA Non-Deterministic Approaches Conference; 9 th AIAA Gossamer Spacecraft Forum and the 4 th AIAA Multidisciplinary Design Optimization Specialist Conference*. American Institute of Aeronautics and Astronautics, 1801 Alexander Bell Drive, Suite 500, Reston, VA, 20191-4344, USA,, 2008.
- [25] S.B. Pope. *Turbulent flows*. Cambridge Univ Pr, 2000.
- [26] M.E. Riley and R.V. Grandhi. Quantification of model-form and predictive uncertainty for multi-physics simulation. *Computers & Structures*, 89(11):1206–1213, 2011.
- [27] R. Rojas, Sa. Kahunde, L. Peeters, O. Batelaan, L. Feyen, and A. Dassargues. Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling. *Journal of Hydrology*, 394(3):416–435, 2010.
- [28] François G Schmitt. About boussinesq’s turbulent viscosity hypothesis: historical remarks and a direct evaluation of its validity. *Comptes Rendus Mécanique*, 335(9):617–627, 2007.
- [29] I.M. Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.
- [30] P.R. Spalart and S.R. Allmaras. A one-equation turbulence model for aerodynamic flows. *La recherche aérospatiale*, 1(1):5–21, 1994.
- [31] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008.
- [32] G. Tang, M.S. Eldred, and L.P. Swiler. Global sensitivity analysis for stochastic collocation expansion. *CSRI Summer Proceedings 2009*, page 100, 2010.
- [33] S. Tavoularis and U. Karnik. Further experiments on the evolution of turbulent stresses and scales in uniformly sheared turbulence. *Journal of Fluid Mechanics*, 204(1):457–478, 1989.
- [34] D.C. Wilcox. Comparison of two-equation turbulence models for boundary layers with pressure gradient. *AIAA Journal (American Institute of Aeronautics and Astronautics);(United States)*, 31(8), 1993.
- [35] D.C. Wilcox. *Companion Software:Turbulence Modeling for CFD Third Edition*, 2006.
- [36] D.C. Wilcox, American Institute of Aeronautics, and Astronautics. *Turbulence modeling for CFD*, volume 3. DCW industries La Canada, CA, 2006.
- [37] J.A.S. Witteveen and G. Iaccarino. Simplex stochastic collocation with eno-type stencil selection for robust uncertainty quantification. *Journal of Computational Physics*, 239:1–21, 2013.
- [38] Z. Zhang, W. Zhang, Z.J. Zhai, and Q.Y. Chen. 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*, 13(6):871–886, 2007.

## Appendix A. Flowchart

A flowchart of our predictive BMSA approach can be found in Figure A.22.

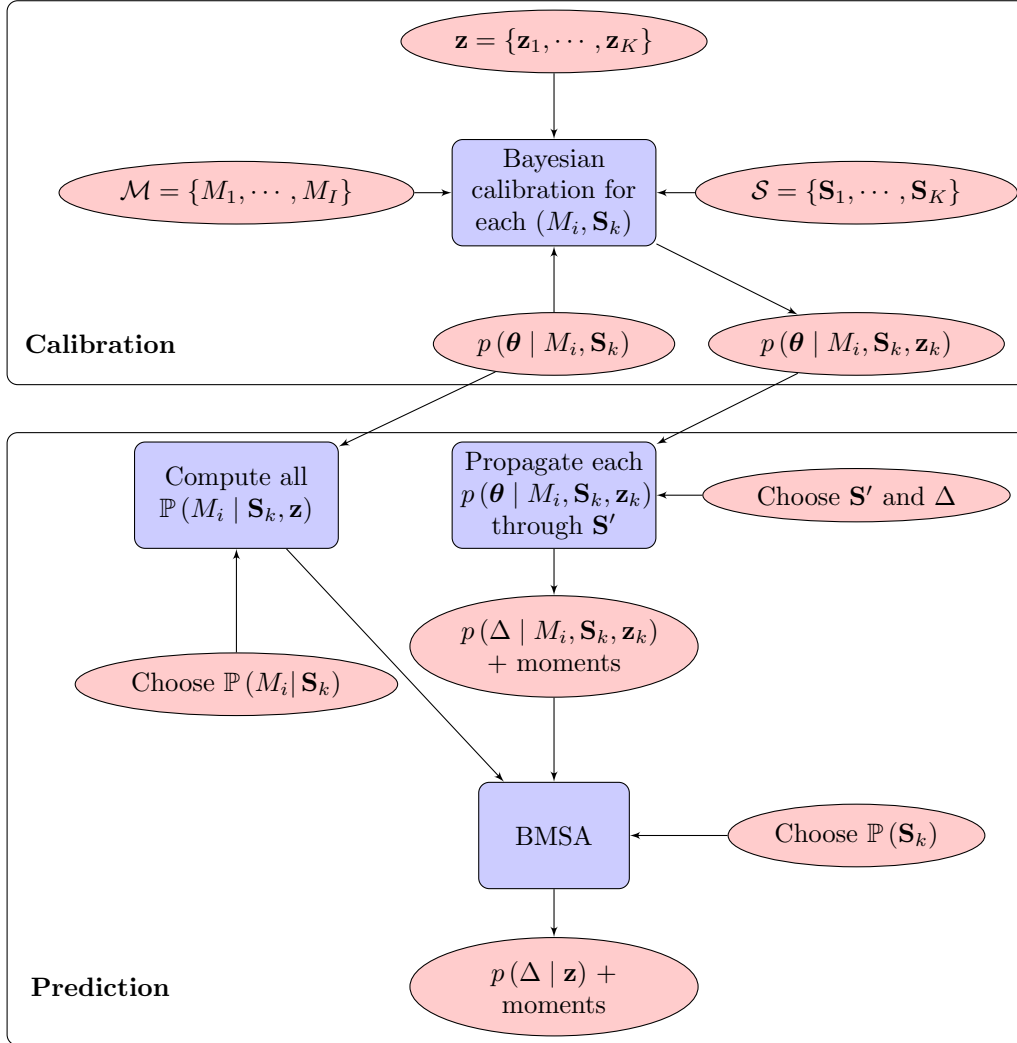


Figure A.22: Flowchart of the BMSA procedure.