Chapter 7

Parameter estimation and uncertainty quantification in turbulence modeling

Richard P. Dwight[a] **and Heng Xiao**[b]
[a]*Aerodynamics, TU Delft, Delft, Netherlands,* [b]*Aerospace Engineering, University of Stuttgart, Stuttgart, Germany*

7.1 Introduction and motivation

We introduce parameter estimation and uncertainty quantification (UQ) in the modeling of turbulent flows. The practice of "modeling" is characterized by the acceptance of imperfection, as summarized by the famous quote:

> *All models are wrong, some are useful.*
>
> George Box (1976)

making the point that we can not expect from a model that it predicts *exactly* a real physical system in all circumstances – that would necessitate a complete and unsimplified representation of reality[1] – rather that it encapsulates the dominant effects of the real system, under the circumstances we care about, and ultimately allows us to reason about the world, and design complex devices with confidence. In Computational Science and Engineering (CSE) our models general consist of PDEs, with associated boundary- and initial-conditions, constitutive laws, and physical- and model parameters. No element of this complex mathematical system is a perfect representation of reality. In UQ our goal is to characterize the difference between the model, and the real system it models. This difference can never be known exactly; we aim to express it using probability.

7.1.1 Turbulence closure modeling

In this chapter we are concerned with the uncertainty due to the modeling of turbulence – so we will first say a few words about the character of turbulence

[1] To underline the point: Navier-Stokes itself is a continuum model of particle dynamics, and it is an increasing poor model in the high Knudsen number limit.

Data Driven Analysis and Modeling of Turbulent Flows. https://doi.org/10.1016/B978-0-32-395043-5.00013-9

models: Turbulence is a multi-scale phenomenon, whereby generally only the *statistics* (temporal or ensemble averages) of turbulence are relevant for engineering applications. Reynolds-averaged Navier-Stokes (RANS) is a framework that predicts the desired statistics *directly*, without resolving *any* turbulent eddies. It does this by solving an averaged Navier-Stokes equation for the averaged velocity field U, and deriving additional transport PDEs for higher-order turbulence statistics (often turbulence kinetic energy k, and dissipation rate ε). These additional PDEs contain terms which are calibrated to match empirical observations. As such the models represent a combination of physical law (mass-, momentum, energy-conservation) which should be respected exactly, and approximate empirical modeling. By solving only for flow-statistics RANS is computationally tractable, at high-Reynolds numbers, for complex geometries, and within design loops; making it the dominant method for predicting turbulent fluids in engineering. However due to the aggressive nature of the approximations made, RANS can make very inaccurate predictions under the wrong circumstances.

Most RANS (a.k.a. *closure*) models involve several free parameters, ranging from a few to a dozen, which need to be calibrated using experimental data or scale-resolving simulations,[2] in order to match the model to reality. Classically, such parameters have been calibrated via by-hand tuning or optimization, against *canonical* flows, representing simple, fundamental turbulence behavior. The procedure results in a set of deterministic parameter values, given which the canonical "training" flows are predicted almost perfectly; similar flows are predicted well; and others predicted only poorly. The flows for which the models *generally* perform well (e.g. wall-bounded attached flow) and poorly (e.g. separated flow), are well-known to experienced practitioners, but the models themselves provide no information about model accuracy to the user.

In fact, no matter what values are chosen for model parameters, there are flows that can not be predicted well by the RANS closure. This inability of a model to perform usefully at any parameter values is called *model inadequacy*, and is the result of modeling choices made in the closure term in the momentum equations (the Reynolds stress divergence), and the particular choice of transport PDEs for k, etc. The choice of the RANS approximation itself (restricting the model to only use turbulence statistics) is the ultimate source of model inadequacy in RANS.

7.1.2 Probability in computational science and engineering

Given the fact that all models are wrong, we wish to characterize the difference between the model and reality. Informally we can ask: What is our confidence in the prediction of the model? Or equivalently: What is our uncertainty?

[2] Since N-S is usually a functionally-exact equation-of-motion for our purposes, Direct Numerical Simulation (DNS) (solving N-S directly) is an alternative source of data for RANS modeling.

Given specific answers to these questions, we can proceed with our decision/design/analysis, or discard our model as not fit for this purpose, and select a better (and likely more expensive) one.

To make these questions more formal, we need to first *represent* our uncertainty – indicating a range of possible values for parameters, state, and predictions. A natural choice is to use random variables and associated probability distributions. Legitimate alternatives are to use an interval to describe a range of values, and then interval arithmetic to relate an input-interval to an output-interval using our model. Other more sophisticated approaches are Dempster-Shaffer evidence theory [7,53], or one of the many fuzzy logics [70]. However we choose probability for a few reasons:

- Multi-variate probability distributions can capture arbitrary statistical dependence between variables in a problem; our physical system and computational model will induce meaningful dependencies, and these we can exploit using *conditional probability*.
- Probability (with Bayes) is the only theory of wagering on events which is "coherent", in the sense that it does not expose the wagerer to certain loss against an optimal opponent [15].
- Probability can be regarded as a generalization of classical logical reasoning, incorporating the possibility of error into reasoning Chapter 6.9 [59].

By representing inputs, parameters, states and outputs of our CSE simulations as random variables, with associated probability densities, we can: determine uncertainty in an output due to uncertainty in an input or parameter ("uncertainty propagation"); determine the distribution on a parameter given measurements of outputs ("calibration"); determine state given measurements ("data-assimilation"); and identify statistical dependence between any pair of variables ("sensitivity analysis"). The specifics of how these can be achieved are detailed in the following sections.

Probability can represent two fundamental kinds of uncertainty:

- **Aleatoric uncertainty** is real and irreducible randomness: e.g. the outcome of tossing a coin. It is objective in the sense that all observers can agree on e.g. the probability of "heads".
- **Epistemic uncertainty** is due to a lack-of-knowledge of the observer, e.g. a coin is already tossed, but is lying under a table. This uncertainty is subjective, and reducible in the sense that learning new information may reduce uncertainty. E.g. an observer peeking under the table will be now certain of the outcome, whereas a blind observer should still assess the probability of "heads" as 50:50.

Both kinds of uncertainty are present in the study of physics, but for the purposes of this chapter we are primarily concerned with (nominally) deterministic physical systems and our imperfect models of them. As such epistemic uncertainty in the accuracy of our models is our main concern, and can potentially be reduced with the use of data.

Regarding again the RANS closure modeling problem, this time from a probabilistic point-of-view: the use of empirical information in the modeling indicates the application of statistics, which provides a rigorous mathematical framework for estimating parameters given data. The fact that RANS models are combination physical laws expressed as nonlinear PDEs, and empirical models (also expressed as PDEs), which can be applied on arbitrary flows, make the use of statistical methods more interesting and challenging than the classical case of curve fitting with e.g. linear models or Gaussian processes.

The remainder of the chapter is structured as follows: in Section 7.2 we consider the application of Bayesian probability to parameter estimation; we develop statistical models, and apply them to the model-problem of linear regression. We pay particular attention to the issue of *model inadequacy*. The reader needing a refresher in (conditional) probability theory may read Appendix 7.A at this point. In Section 7.3 we apply the same techniques to the problem of parameter calibration of a RANS models against experimental data. We will observe directly the model inadequacy present in simple RANS models. Thus motivated in Section 7.4 we address model inadequacy by means of *model-form* uncertainty.

7.2 Bayesian parameter estimation

We want to address the following problem statement:

Problem 1. (Parameter estimation) Consider a known *model* $f : \mathcal{X} \times \mathcal{Q} \to \mathbb{R}$ representing some physical process, taking two arguments: $x \in \mathcal{X} \subset \mathbb{R}$ an independent variable (often representing space), and $\theta \in \mathcal{Q} \subset \mathbb{R}$ a *parameter* of the model. Assume in addition we have $N \in \mathbb{N}$ *observations* $\boldsymbol{y} \in \mathbb{R}^N$ of the same physical process, observed at known locations $\boldsymbol{x} \in \mathbb{R}^N$. What value of $\theta \in \mathcal{Q}$, when used in f best represents the observations?

This problem is often called a *calibration problem*, and belongs to the larger class of *inverse problems* ("inverse" as the objective is to estimate a model *input*, rather than an output). Attempting to solve this problem, the first issue is its vague formulation: the precise meaning of "best" is not specified. We might already consider a least-squares fit of the observations, but it's not clear what consequences that will have for the accuracy of the estimated θ, nor the ability of the resulting model to make predictions. The problem may be underdetermined (with infinitely many solutions), or over-determined (more common in parameter calibration). Finally, the model itself is necessarily imperfect (see Section 7.1), and a least-squares fit provides no insight into the errors that result from model inadequacy versus those from the choice of θ. These issues we attempt to address by formulating the solution of Problem 1 as a probability distribution on θ, and identifying this distribution with the Bayesian framework.

7.2.1 Bayes' theorem

Bayes theorem refers to the trivial consequence of the definition of conditional probability (7.34):

$$\rho_{\Theta,Y}(\theta, y) = \rho_{\Theta|y}(\theta|y) \cdot \rho_Y(y) = \rho_{Y|\theta}(y|\theta) \cdot \rho_\Theta(\theta)$$

often formulated

$$\rho_{\Theta|y}(\theta \mid y) = \frac{\rho_{Y|\theta}(y \mid \theta) \cdot \rho_\Theta(\theta)}{\rho_Y(y)}, \text{ where } \rho_Y(y) > 0, \tag{7.1}$$

relating the conditional distribution of Θ given Y to that of Y given Θ. In the terminology of Bayesian probability $\rho_{Y|\theta}(y \mid \theta)$ is the *likelihood* of y given θ; $\rho_\Theta(\theta)$ is the *prior* on θ; $\rho_Y(y)$ is the *evidence* (a.k.a. the *marginal likelihood*); and $\rho_{\Theta|y}(\theta \mid y)$ is the *posterior* on θ given y.

In the following example we neglect the independent variable x from Problem 1 for simplicity. This is the case when our model $f(\cdot)$ predicts a single value, rather than a function.

Example 1. (Approximate function inverse) Consider a function f taking a single input $\theta \in \mathbb{R}$, returning a single output $y \in \mathbb{R}$, denoted $y = f(\theta)$. We assume that there exists some true but unknown $\tilde{\theta} \in \mathbb{R}$, and we represent our *empirical uncertainty* in this value with the r.v. $\Theta : \Omega \to \mathcal{Q}$, with pdf $\rho_\Theta(\theta)$. Given Θ, our corresponding empirical uncertainty in the true value $\tilde{y}$ is $Y = f \circ \Theta$, with pdf $\rho_Y(y)$.[3]

To apply Bayes, we first specify our empirical uncertainty in the value of $\tilde{\theta}$ (in the absence of observations) as the r.v. Θ. This as our *prior* in (7.1), and this should incorporate everything we know about $\tilde{\theta}$ from all existing sources. Now new information arrives in the form of a noisy measurement $\hat{y}$ of $\tilde{y}$. To make this statement precise we characterize the measurement using a *statistical model* – for instance a common model is:

$$\hat{y} = \tilde{y} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \hat{\sigma}^2),$$

i.e. the measurement contains additive Gaussian noise with standard-deviation $\hat{\sigma}$, but is an un-biased observation of $\tilde{y}$ (since $\mathbb{E}\epsilon = 0$, so $\mathbb{E}\hat{y} = \tilde{y}$). Since $\tilde{y} = f(\tilde{\theta})$ we have $\hat{y} = f(\tilde{\theta}) + \epsilon$ and

$$\hat{y} - f(\tilde{\theta}) = \epsilon \sim \mathcal{N}(0, \hat{\sigma}^2), \implies \hat{Y} \sim \mathcal{N}(f(\tilde{\theta}), \hat{\sigma}^2). \tag{7.2}$$

---

[3] In the case of strictly increasing, differentiable f the explicit expression for $\rho_Y(y)$ is:

$$\rho_Y(y) = \rho_\Theta(f^{-1}(y)) \left. \frac{\mathrm{d}f^{-1}}{\mathrm{d}y} \right|_y,$$

however in general we have to resort to numerical approximations of $\rho_Y(y)$.

Now $\tilde{\theta}$ is unknown, but for any given $\theta \in \mathbb{R}$:

$$\hat{Y} \mid \theta \sim \mathcal{N}(f(\theta), \hat{\sigma}^2),$$

which gives us the likelihood $\rho_{Y|\theta}(y|\theta)$ in (7.1). The denominator $\rho_Y(y)$ we can regard as a normalization factor, as it's independent of θ. So the posterior is:

$$\rho_{\Theta|y}(\theta \mid y) \propto \rho_{Y|\theta}(y \mid \theta) \cdot \rho_\Theta(\theta),$$

which is our answer to the question: "What do we know about $\tilde{\theta}$?". For an affine function f, and a Gaussian prior, a short calculation shows that the posterior is also a Gaussian, whose mean and standard-deviation are a compromise between the prior and the observed data.

It is important to note that the statistical model $\hat{y} = \tilde{y} + \epsilon$ chosen above is in fact a *model* for an objective physical process (the noise of measurement), and as such is "wrong" (in the sense of Box). The r.v. ϵ represents *aleatoric* uncertainty, and errors in this model result in errors in the posterior. This is in contrast to the prior, which represents purely *epistemic* uncertainty, and cannot be "wrong" in the same sense. More concerning, by replacing $\tilde{y}$ with $f(\tilde{\theta})$ in (7.2) we are assuming that the *only* difference between the model predictions and $\hat{y}$ is the measurement noise; i.e. we are assuming f is perfect, which is never the case in reality. This is the issue of model inadequacy, which we return to later.

7.2.2 Parameter identification for a CFD problem

To motivate the Bayesian framework consider a sketch of a simple CFD problem. Consider a lifting aerofoil measured in a wind-tunnel, and simultaneously simulated in a CFD code where three uncertain parameters have been identified to be the dominant source of discrepancy between the two modes of analysis: angle-of-attack α, Mach number M_∞, and Reynolds number Re, see Fig. 7.1, collectively $\boldsymbol{\theta} = (\alpha, M_\infty, \mathrm{Re})$. After discussion with the experimentalists and studying the properties of the tunnel, we impose prior r.v.'s on these parameters representing the experimental uncertainty. These are independent since – *a priori* – we have no reason to expect the errors in α and Re to be correlated.[4] This gives us our prior Θ_0 with pdf $\rho_0(\boldsymbol{\theta}) = \rho_0(\alpha, M_\infty, \mathrm{Re}) = \rho_0(\alpha) \cdot \rho_0(M_\infty) \cdot \rho_0(\mathrm{Re})$ where the subscript 0 hints that these are prior quantities. These priors are indicated by the black pdfs on the left-hand side of Fig. 7.1.

With these priors and our simulation code $f : \mathbb{R}^3 \to \mathbb{R}^2$ we predict r.v.'s $C = (C_L, C_D)$ on the lift c_L and drag c_D, respectively as: $C_0 = f(\Theta_0)$. This is the *prior predictive distribution* whose marginals are sketched as black pdfs on the right-hand side of Fig. 7.1, but which are in fact dependent (since higher lift generally indicates higher drag). This step is not strictly necessary to calibrate

[4] Although arguably Re and M_∞ both depend on U_∞, so these are potentially correlated in a good prior.

parameters in the next step, but it is a useful sanity check on the simulation and data: if the observation $\hat{y}$ in the next step is extremely unlikely (or impossible) under the prior predictive distribution, then the analysis is suspect, and modeling assumptions should be checked, see Chapter 6 of [17].

Next a noisy measurement $\hat{c}_L$ is made of $\tilde{c}_L$, and the noise is characterized as $\epsilon \sim \mathcal{N}(0, \hat{\sigma}^2)$, where $\hat{\sigma}$ is estimated based on the measurement setup – this is the red pdf for c_L on the right-hand side of Fig. 7.1. The statistical model is then $\hat{c}_L = c_L + \epsilon = f_1(\theta) + \epsilon$, and the likelihood has pdf:

$$\rho(c_L \mid \boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\left[\frac{(\hat{c}_L - f_1(\boldsymbol{\theta}))^2}{\hat{\sigma}^2}\right]\right\},$$

so the posterior on the parameters is given as:

$$\rho(\boldsymbol{\theta} \mid C_L = \hat{c}_L) \propto \rho(\hat{c}_L \mid \boldsymbol{\theta}) \cdot \rho_0(\boldsymbol{\theta}),$$

under which the parameters $\boldsymbol{\theta}$ are dependent (unlike under the prior), and their marginals are plotted as red pdfs on the left-hand side of Fig. 7.1. These we have sketched to indicate the fact that:

- The angle-of-attack α would be strongly informed by lift c_L – which can be seen since the standard-deviation of the distribution has been reduced dramatically with respect to the prior.
- Whereas the Reynolds number is hardly informed (assuming attached flow), and so its posterior is very similar to the prior (i.e. little information gain has been realized).

This quantification of the standard-deviation of the parameters is key benefit over e.g. a least-squares fit of the parameters to the data, which will return best guesses of all $\boldsymbol{\theta}$ without indicating that Re is irrelevant.

Finally we can propagate the posteriors on the parameters to the forces, the *posterior predictive distribution*, giving us uncertain estimates of the *unmeasured* quantity c_D, informed by measurements of c_L. This distribution is given by:

$$\rho(c_D \mid C_L = \hat{c}_L) \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \rho(c_D \mid \boldsymbol{\theta}) \cdot \rho(\boldsymbol{\theta} \mid \hat{c}_L)\, \mathrm{d}\boldsymbol{\theta},$$

and is the red pdf on the drag in Fig. 7.1. Similarly the distribution $\rho(c_L \mid C_L = \hat{c}_L)$ can be compared to the observation $\hat{c}_L$. If the observation is implausible under this distribution then the model is suspect, and may be inconsistent, see Section 6.3 [17].

7.2.3 Application to regression: model inadequacy

In the examples so far our statistical model has been of the form $y = f(\theta) + \epsilon$ for $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The noise ϵ captures the effect of measurement error, but the

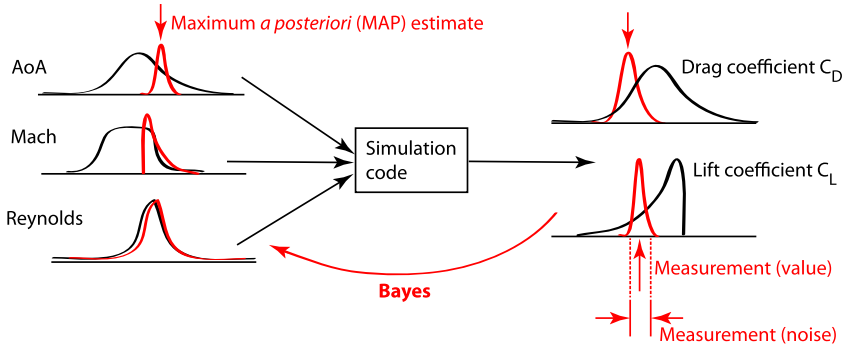


FIGURE 7.1 Cartoon of Bayesian updating for an aerofoil. 1) Black pdfs at the left represent priors on the simulation inputs; when propagated through the code they result in the black pdfs in the forces on the right (the "prior predictive distribution"). 2) A noisy measurement is made of C_L represented by the red pdf; applying Bayes gives the posterior $\rho(\theta|y)$ on the inputs (red pdfs on the left). 3) Propagating this posterior through the code, gives new information about the unmeasured C_D (the "posterior predictive distribution").

implicit assumption is that the model f is perfect, in the sense that: if the true x is specified, the model will reproduce the *noiseless truth*. This is unlikely in reality, where models are always wrong, and has consequences for our posterior [26].

To illustrate, consider the problem of regressing data $\mathcal{D} := \{(x_i, y_i) \mid i = 1, \ldots, N\}$ with an affine function $f(x; a, b) = ax + b$; where the two parameters $(a, b) =: \boldsymbol{\theta} \in \mathbb{R}^2$ must be estimated. Assume non-informative priors on the parameters: $\rho(a, b) \propto 1$, and the statistical model

$$y = f(x; a, b) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2), \tag{7.3}$$

which we call the "denial" model,[5] and the posterior is $\rho(a, b \mid \mathcal{D})$. Consider now two possible ground-truths:

A: $X \sim \mathcal{U}(0, 1), \quad Y = a^\star X + b^\star + \mathcal{N}(0, \sigma^2),$
B: $X \sim \mathcal{U}(0, 1), \quad Y = a^\star X + b^\star + c^\star X^2 + \mathcal{N}(0, \sigma^2),$

from which data $\mathcal{D}_A$ and $\mathcal{D}_B$ are generated. If A is true, the model (7.3) is perfect (given $a = a^\star$, $b = b^\star$), and we expect to recover estimates of $a^\star$, $b^\star$ from the posterior. If B is true, the model is inadequate, and estimates of $a^\star$, $b^\star$ will be contaminated by the presence of $c^\star X^2$.

Numerically the effect is visible in Figs. 7.2a and 7.2b. In the former, as the number of samples increases, the posterior estimate (visualized by 50 samples from the posterior predictive distribution) approaches the truth. Not only do the estimates of a, b converge to the exact values, but also the estimate of uncertainty in these parameters (and therefore the regression model) is reliable. On

[5] As it's "in denial" about the possibility of model inadequacy.

the other hand, in Fig. 7.2b, the inadequacy of the model does not prevent convergence of a, b, but they converge to values which do not correspond to $a^\star$, $b^\star$. As such, if the parameter values themselves are of primary interest, the estimate has completely failed. Furthermore the posterior predicts very high confidence (low variance) of a regression model which does not correspond to the truth. This is an example of an error in the statistical model for the likelihood leading to false conclusions, even in the case of unlimited data. Contrast this situation with poorly chosen priors, which become increasingly ignored/irrelevant as data increases.[6]

To address this we can explicitly add the possibility of model inadequacy into the statistical model. Since the form of the inadequacy is *a priori* unknown, we require a very general parameterization. Following Kennedy and O'Hagan [26] we could choose:

$$y = f(x; a, b) + \delta(x) + \epsilon, \quad \delta \sim \mathcal{GP}(\mu_\delta(\cdot), r(\cdot, \cdot)), \quad \epsilon \sim \mathcal{N}(0, \sigma^2), \quad (7.4)$$

where δ is a Gaussian process (GP), intended to account for systemic inadequacy in f. Briefly a random process is a generalization of random variables to functions. A Gaussian process is a random process with the properties that:

1. The process evaluated at any point x_i is a Gaussian r.v.: $\delta(x_i) \sim \mathcal{N}(\mu_\delta(x_i), r(x_i, x_i))$;
2. The process evaluated at any two points x_i, x_j is a joint Gaussian random vector: $[\delta(x_i), \delta(x_j)] \sim \mathcal{N}([\mu_\delta(x_i), \mu_\delta(x_j)], \Sigma)$, with $[\Sigma]_{ij} = r(x_i, x_j)$.

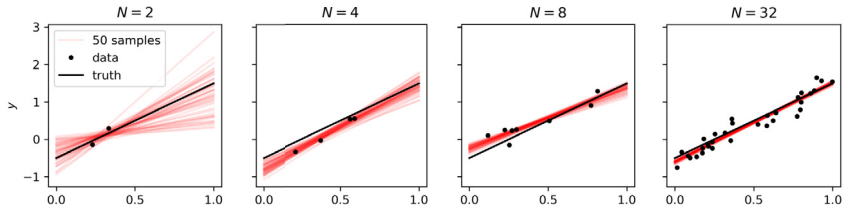
As such the function $\mu_\delta : \mathbb{R} \to \mathbb{R}$ defines the process mean, and the *covariance function* $r : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defines the covariance, including how rapidly the process is allowed to vary.

Since the Gaussian process, evaluated on a grid $x_0, \ldots, x_M$ is a multivariate Gaussian, the expression for the likelihood is very similar to the case of the denial model. Notably the posterior is now $\rho(a, b, \delta \mid \mathcal{D})$, i.e. the data is used to estimate both the parameters a, b and δ. Applying this model to $\mathcal{D}_B$ leads to Fig. 7.2c. This time we have convergence to the true process as $N \to \infty$, and reliable uncertainty estimates on our regression model. However, it is still not the case that a, b will converge to their exact values $a^\star$, $b^\star$ in B above; as the model has no means of distinguishing between effects that are part of f versus part of δ. By choosing the prior mean of δ to be zero, and the prior variance to be small, we can specify an prior of *small* model inadequacy. Then, broadly speaking, most of the variability observed in the data will be ascribed to f, and only that variability not represented by f will be ascribed to δ.

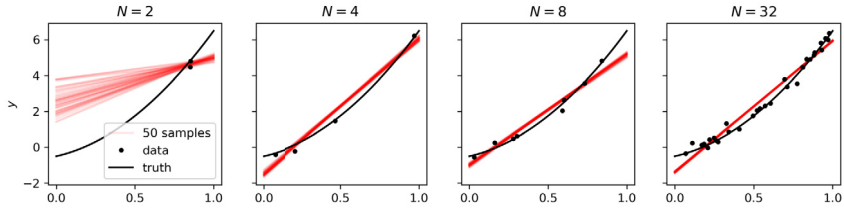
7.2.4 Computational considerations

We've discussed so far only the statistical aspects of the Bayesian updating procedure, but computational aspects of the required steps of the are very chal-

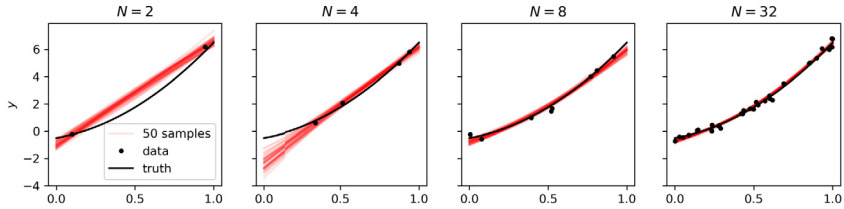
[6] Provided they don't specify zero probability of the truth.



(a) Linear data (A), denial model $y = ax + b + \epsilon$.



(b) Nonlinear data (B), denial model $y = ax + b + \epsilon$.



(c) Nonlinear data (B), model inadequacy $y = ax + b + \delta(x) + \epsilon$.

FIGURE 7.2 Regression of data with a straight-line, with and without model inadequacy.

lenging. Bayesian methods were developed in the context of models with trivial computational requirements (e.g. linear regression) [17], where analytic or Monte-Carlo approaches are tractable – when f is a CSE code this is no longer the case. Furthermore unlike linear regression (cf. Section 7.2.3), the CSE code introduces a nonlinear dependence of the likelihood on the model parameters. As such the likelihood will not in general be a Gaussian in the parameters, neither will the posterior, for which an analytic expression will not exist in general. Finally the posterior can not be entirely characterized by a mean and covariance matrix.

Another feature of the posterior that makes evaluating statistics challenging is the following: though Bayes gives an explicit expression for the posterior pdf, the coefficient space is usually relatively high-dimensional (6-d in the case of calibrating the $k - \varepsilon$ model in the next section), and the posterior pdf is likely to be very close to zero in a large region of this space. Wherever coefficient values result in models that match poorly to data, the posterior probability will be small; and if the data is very informative with respect to the coefficients (a desirable

situation for calibration), then the region of significantly nonzero probability may be very localized indeed. *Finding* this region may be a challenge, and is a prerequisite to estimating statistics there. This issue is either addressed by sampling from the posterior in Section 7.2.4.1, or solving optimization problems in Section 7.2.4.3.

The three operations needed above are (for inputs θ, outputs y):

1. Estimating the pdf or moments of $Y = f(\Theta_0)$, for a given prior Θ_0 ("uncertainty propagation");
2. Obtaining samples/moments of the posterior $\rho(\theta \mid Y = \hat{y})$;
3. Estimating the posterior predictive distribution $\rho(y \mid Y = \hat{y})$.

Often an approximation of the statistics is necessary to make the approach tractable – see e.g. [13]. Here we will only give an impression of the computational approaches that can be used, which at a high level are all either: Monte-Carlo-like methods, surrogate modeling, or variational methods.

7.2.4.1 Monte-Carlo methods

These methods all rely on generating random samples from a distribution to estimate statistics. First consider the problem of estimating the mean of $Y = f(\Theta_0)$. Given samples $\theta_1, \ldots, \theta_N \sim \Theta_0$ this is straightforward:

$$\mathbb{E}_\Theta Y := \int_{-\infty}^{\infty} f(\theta)\rho_0(\theta)\,\mathrm{d}\theta \simeq \frac{1}{N}\sum_{i=1}^{N} f(\theta_i) + \mathcal{O}\left(\frac{\sigma_y}{\sqrt{N}}\right), \tag{7.5}$$

where the error in the estimate is proportional to the standard-deviation of Y, σ_y [47]. These methods are characterized by a slow ($\mathcal{O}N^{-1/2}$), but dimension-independent convergence rate. For moderate-dimensional inputs θ, Quasi-Monte-Carlo methods [31], which use identical estimator, but with low-discrepancy (a.k.a. quasi-random) sequences, achieve errors of $\mathcal{O}((\log N)^M/N)$, where M is the dimension of θ. Generating low-discrepancy sequences for dependent r.v.'s can be challenging [21], and may not be possible for e.g. the posterior. Another acceleration approach is Multi-level Monte-Carlo [19], used in CFD in [28], which exploits a sequence of increasingly coarser grids (the "levels"), and the fact that the variance of the *difference* between y on successive levels is small – giving a small error in (7.5).

The problem of estimating the pdf of Y can be solved with kernel-density estimation (KDE) [52]. Begin with random samples from Θ: $\theta_1, \ldots, \theta_N$, and compute:

$$\rho(y) \propto \frac{1}{N}\sum_{i=1}^{N} \frac{1}{h} K\left(\frac{|y - f(\theta_i)|}{h}\right),$$

where $h \in \mathbb{R}$, $h > 0$ is a scale in the space of Y (estimated by e.g. Scott's rule [52]; and K is a positive kernel function, e.g. a Gaussian. This can be com-

pared to a generalized histogram, and again this estimate requires N evaluations of the CFD code, where N is likely large.

Finally, the problem of obtaining samples from a distribution with known pdf – such as the posterior $\rho(\theta \mid Y = \hat{y})$, can be addressed with Markov-Chain Monte-Carlo (MCMC) methods [5]. Direct sampling from the posterior is difficult, since the cdf is not known, and most of the probability mass is usually located in a small region of the θ-space. MCMC obtains samples from $\rho(\theta \mid Y = \hat{y})$ (or any other computable pdf), by constructing a continuous-state Markov-chain, whose stationary distribution is exactly the target density. Sampling from the chain sufficiently long will thereby yield samples of the target. Again this process is costly, and each subsequent sample requires evaluation of the posterior density (and so evaluation of the simulation code for CSE models). Acceleration methods such as Hamiltonian Monte-Carlo (Chapter 5 of [5]), and especially the "No U-turn sampler" (NUTS) [23] exist, but typically thousands of evaluations are still required. Note that MCMC methods do not share the dimension-independence of Monte-Carlo estimators, and in fact generally scale poorly with dimension.

In general, all the Monte-Carlo methods are not practical for Bayesian problems in expensive CFD codes – they are rather applied on cheap surrogates.

7.2.4.2 Surrogate modeling methods

For relatively low-dimensional spaces – up to about 20 – surrogate models of the CSE code response can be constructed based on evaluations of f, and then statistics computed using Monte-Carlo, and samples generated using MCMC methods on this surrogate.

Methods first introduced were for uncertainty propagation based on polynomials: named variously polynomial chaos (PC) [18], probabilistic collocation [35], etc. The response of the model in the stochastic space is represented by a polynomial, which is then fit with either Galerkin projection (in PC), or collocation. The polynomial basis is often chosen optimally with respect to the input distribution – e.g. a Hermite basis for Gaussian inputs, and the statistics of the output distribution can then be written explicitly in terms of the polynomial coefficients. These methods enjoy spectral convergence thanks to the polynomial bases. Applying them to dependent distributions is not trivial however. Alternatively any quadrature rule can be applied in the stochastic space to evaluate expectations (which are essentially just weighted integrals), e.g. quadrature on simplices [66,11], or sparse grids [9].

Alternatively Gaussian process regression (a.k.a. Kriging) models can be used as a surrogate [16]. These models are inherently statistical – the underlying approximant is a random-process – and are therefore a good fit in the Bayesian framework (and more widely used in the statistics community than polynomial methods). Their variance can be used to incorporate the surrogate modeling error into predictions. Gradient-information, if available, can be incorporated [8], leading to improved dimensional scaling [2].

However all these methods suffer from the curse-of-dimensionality, sparse-grids and gradient-enhanced surrogates only delay it. Therefore dimensionality reduction is a prerequisite in most practical applications [1]. Furthermore, note that if surrogate models are applied to generate samples or estimate statistics of the *posterior*, ideally the surrogate should be generated using many code samples in regions of high-posterior probability. Since these regions are not known *a priori*, adaptive sampling methods are needed.

7.2.4.3 Variational methods

Variational methods circumvent the curse-of-dimensionality by reformulating the problem statement as an optimization problem, and subsequently using gradient-based methods to solve it. The twin facts that (a) gradients of a loss-function can be evaluated independently of the parameter-space dimension (with adjoint methods [43] or back-propagation), and (b) steepest-descent-like methods for optimization have convergence which is independent of dimension [39]; together lead to dimension independent methods for estimating statistics of the posterior. In the case of PDE CSE models, this results in a *PDE constrained optimization* problem [22].

The simplest example of a variational method is the *Maximum A-Posteriori* (MAP) estimate. Given a posterior $\rho(\theta \mid Y = \hat{y}) \propto \rho(\hat{y} \mid \theta) \cdot \rho_0(\theta)$, we can approximate it by the single value of $\theta_{\text{MAP}} \in \mathcal{Q}$ which gives maximum probability:

$$\theta_{\text{MAP}} := \arg\max_{\theta \in \mathcal{Q}} \rho(\theta \mid \hat{y}) = \arg\max_{\theta \in \mathcal{Q}} \left[\log \rho(\hat{y} \mid \theta) + \log \rho_0(\theta) \right], \tag{7.6}$$

where working with log-probabilities is necessary to avoid floating-point underflow for extremely low values of the likelihood that can arise as the result of (e.g.) an uninformed initial guess for θ. Note that once again the *evidence* $\rho(y)$ is not needed, as it is not a function of θ, and therefore only leads to an additive constant in the objective of (7.6). Derivatives $\nabla_\theta \log \rho(\hat{y} \mid \theta)$ and $\nabla_\theta \log \rho_0(\theta)$ are required. The former necessitates derivatives of the CSE code (e.g. an adjointed version of the code), whereas the latter is a matter of analytic differentiation (for analytically defined priors).

The main limitation of this formulation is that the vast majority of statistical information in the posterior is lost. There is no information retained about the variance or co-variance of the estimated quantities, and no estimate is available for the mean – which may be far from the MAP value. One approach to regain some statistical information is to compute the Hessian of the posterior at the MAP estimate (the gradient of the posterior is zero at this point) in a post-processing step, and use the implied quadratic response to fit a Gaussian, this is known as *Laplace's approximation* [25].

A more general approach – not restricted to Gaussians – is to approximate the posterior $\rho(\boldsymbol{\theta} \mid \hat{y})$ by a *variational distribution* $q(\boldsymbol{\theta})$ with unknown hyper-parameters. If q were a Gaussian, this distribution could be $q(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, with μ and Σ the hyper-parameters. The hyper-parameters are then

estimated by minimizing a measure of distance between $q(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma)$ and the target distribution $\rho(\boldsymbol{\theta} \mid \hat{y})$. Most commonly the Kullback-Leibler divergence is used as a convenient measure of distance:

$$D_{\mathrm{KL}}\left(q(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma) \,||\, \rho(\boldsymbol{\theta} \mid \hat{y})\right) := \int q(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma) \log\left(\frac{q(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma)}{\rho(\boldsymbol{\theta} \mid \hat{y})}\right) \mathrm{d}\boldsymbol{\theta} \tag{7.7}$$

$$= \mathbb{E}_{q(x|\mu,\Sigma)} \log\left(\frac{q(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma)}{\rho(\boldsymbol{\theta} \mid \hat{y})}\right). \tag{7.8}$$

The optimization problem is formulated as

$$\boldsymbol{\mu}_{\mathrm{opt}}, \Sigma_{\mathrm{opt}} := \underset{\boldsymbol{\mu}, \Sigma}{\arg\min}\, D_{\mathrm{KL}}\left(q(x \mid \boldsymbol{\mu}, \Sigma) \,||\, \rho(x \mid \hat{y})\right),$$

given which $q(\boldsymbol{\theta} \mid \boldsymbol{\mu}_{\mathrm{opt}}, \Sigma_{\mathrm{opt}})$ is taken as an estimate of the posterior.

The challenge of this formulation is estimating the integral in (7.8), which must be performed at every iteration of the optimizer. In principle a Monte-Carlo estimate could be used, which retains the dimension-independence of the method as a whole. However this introduces significant noise into the objective function, as well as multiplying the cost per iteration. In practice it is preferable to maximize the *evidence lower-bound* (ELBO), which minimizes an upper-bound on the KL-divergence [27]. The former can be maximized with stochastic gradient ascent. A particularly elegant algorithm for this problem is the *Stein Variational Gradient Descent* [34], which performs gradient descent on a set of particles, which are used to approximate the expectation in (7.8), and simultaneously represent independent samples of the posterior.

In the case that gradients are not available, these same objectives can be optimized with gradient-free optimizers for sufficiently low parameter dimensions. E.g. the Nelder-Mead simplex method [46] is recommended in input spaces up to about 10 dimensions, and is extremely robust to the noise present in estimates of the integrals in (7.8). In this gradient-free scenario however, MCMC methods are likely to be preferable.

7.3 RANS closure model coefficient calibration

The Bayesian updating procedure of the previous chapter is applied to the problem of calibrating closure coefficients for RANS closure models. Classical calibration methods are first exemplified in Section 7.3.1. These methods identify coefficients 1- or 2-at-a-time using fundamental flows. Next an application of the Bayesian approach to the $k - \varepsilon$ model is given in Section 7.3.2. Finally in Section 7.3.3 we introduce the concepts of Bayesian model selection and averaging, which answer the question: Given multiple candidate models, of different forms, and various coefficients, how can we choose a single (or average), best model for a specific problem? As always we formulate the answers to these questions in a probabilistic sense.

7.3.1 Classical identification of closure coefficients

In order to draw a contrast to the Bayesian approach, we give two examples of the manner in which turbulence models are classically calibrated. In general, this procedure consists of (i) identifying a "canonical" flow for which the PDE model simplifies significantly, (ii) solving the simplified system analytically, and (iii) fitting the solution to experimental or DNS data by tuning the closure coefficients. If the model is well-constructed, it is possible to consistently fit multiple fundamental flows; the anticipation then being that the essential properties of more complex flows will also be reproduced.

7.3.1.1 Canonical flow I: decay of isotropic turbulence

Consider a flow whose mean $\bar{u}$ is homogeneous (invariant to translations) and isotropic (invariant to rotations), i.e. constant everywhere. Due to the lack of velocity gradients, production of turbulence is zero – which can be seen from the exact T.K.E. equation [32] and as such the kinetic energy contained in the turbulence scales ($k^\star := \frac{1}{2}\overline{u_i' u_i'}$) decays. In the following the subscript $\star$ indicates the exact turbulence quantity, rather than a model. It is an empirical observation (not a theoretical fact), supported by experiment and DNS, that this decay follows a power-law, with an exponent close to 1.3 [37], i.e. $k^\star \propto t^{-1.3}$.

Now consider the $k-\varepsilon$ model [62] applied to this flow. The model consists of two PDEs nominally approximating the turbulence kinetic energy $k \simeq k^\star$, and dissipation rate $\varepsilon + \varepsilon_0 \simeq \varepsilon^\star := \nu \overline{\frac{\partial u_i'}{\partial x_j}\frac{\partial u_i'}{\partial x_j}}$:

$$\frac{\partial k}{\partial t} + \bar{u_i}\frac{\partial k}{\partial x_i} = P_k - \varepsilon + \frac{\partial}{\partial x_i}\left[\left(\nu + \frac{\nu_T}{\sigma_k}\right)\frac{\partial k}{\partial x_i}\right], \tag{7.9a}$$

$$\frac{\partial \varepsilon}{\partial t} + \bar{u_i}\frac{\partial \varepsilon}{\partial x_i} = C_{\varepsilon 1} f_1 \frac{\varepsilon}{k} P_k - C_{\varepsilon 2} f_2 \frac{\varepsilon^2}{k} + E + \frac{\partial}{\partial x_i}\left[\left(\nu + \frac{\nu_T}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_i}\right], \tag{7.9b}$$

where the production term $P_k := \tau_{ij}\frac{\partial \bar{u_i}}{\partial x_j} \simeq \tau_{ij}^\star \frac{\partial \bar{u_i}}{\partial x_j} =: P_k^\star$, with a Boussinesq model for the Reynolds-stress tensor τ_{ij}:

$$\tau_{ij} := 2\nu_T S_{ij} - \frac{2}{3}k\delta_{ij}, \quad S_{ij} := \frac{1}{2}\left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i}\right),$$

and the eddy viscosity is defined as

$$\nu_T := C_\mu f_\mu \frac{k^2}{\varepsilon},$$

where f_μ, f_1 and f_2 are blending functions, and five closure coefficients are C_μ, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, σ_k and σ_ε.

Assuming isotropic homogeneous flow, all spatial derivatives disappear: $\frac{\partial \bar{u}_i}{\partial x_j} = 0$, $\frac{\partial k}{\partial x_j} = 0$ and $\frac{\partial \varepsilon}{\partial x_j} = 0$, as does turbulence production P_k. In this setting

the equations simplify dramatically, to a system of ODEs with one remaining closure coefficient:

$$\frac{dk}{dt} = -\varepsilon, \tag{7.10}$$

$$\frac{d\varepsilon}{dt} = -C_{\varepsilon 2}\frac{\varepsilon^2}{k}, \tag{7.11}$$

with exact solution

$$k(t) = k_0 \left(\frac{t}{t_0}\right)^{-n}, \tag{7.12}$$

$$\varepsilon(t) = \varepsilon_0 \left(\frac{t}{t_0}\right)^{-(n+1)}, \tag{7.13}$$

for an arbitrary timescale t_0 and the constraints $\varepsilon_0 t_0 = nk_0$ and $nC_{\varepsilon 2} = n + 1$.

Since there is only one remaining parameter ($C_{\varepsilon 2}$) and the power-law solution of the $k - \varepsilon$ model corresponds to the power-law seen in experiment, it is straightforward to fit the two against each other (e.g. with least-squares). For a rate of $n \simeq 1.3$, we have $C_{\varepsilon 2} \simeq 1.77$; however values in the range $C_{\varepsilon 2} \in [1.68, 1.92]$ are commonly used in practice [12].

7.3.1.2 Canonical flow II: free shear flows

In free-shear flows (i.e. shearing flows away from walls, including shear-layers, jets and wakes) it is an empirical observation that the production and dissipation of T.K.E. approximately balance $P_k^\star \simeq \varepsilon^\star$ – at least in some region of the flow. This is also approximately true in the law-of-the-wall region of boundary-layers. The $k - \varepsilon$ model, explicitly models both production and dissipation, so the balance can be formulated:

$$P_k = \nu_T \left(\frac{\partial \bar{u}}{\partial y}\right)^2 = C_\mu \frac{k^2}{\varepsilon}\left(\frac{\partial \bar{u}}{\partial y}\right)^2 \simeq \varepsilon. \tag{7.14}$$

In doing so we have assumed a coordinate system with y tangential to the shear layer, and that the layer is thin enough that components of the velocity gradient other than $\partial u/\partial y$ can be neglected (a modeling approximation). Using once more the Boussinesq model for τ_{ij}, we have $\tau_{xy}^\star := \overline{u'v'} \simeq \nu_T \frac{\partial \bar{u}}{\partial y}$, which together with (7.14) yields

$$C_\mu = \left(\frac{\overline{u'v'}}{k}\right)^2.$$

In DNS data [48] a plateau of the ratio $\overline{u'v'}/k$ can be seen, see Fig. 7.3 taking a value close to -0.3 [45]. Although the clarity and spatial extent of the plateau,

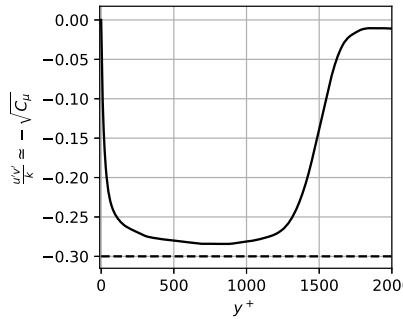


FIGURE 7.3 Ratio $u'v'/k$ as a function of wall-distance for a zero-pressure gradient turbulent boundary layer from DNS at $Re_\tau = 1272$, data from [48].

as well as the particular value varies based on the flow. On this basis $C_\mu \simeq 0.09$ is often considered a reasonable value.

Also in this case the value of the coefficient used in practice can vary dramatically. Notably when RANS is used wind-energy (in which the atmospheric boundary-layer is modeled), C_μ is widely chosen to be 0.03 [29], irrespective of the considerations here.

7.3.1.3 Discussion

While the technique of calibrating against canonical flows has been enormously successful in generating useful models; the calibration procedure does not generalize well to complex flows – where many or all coefficients may be relevant, nor to calibration against multiple (potentially incompatible) datasets.

In the above calibrations – after performing simplifications – the final step is parameter estimation. This is essentially a *statistical* step. In the case of isotropic decay (Section 7.3.1.1) this is comparatively straight-forward, as data shows a clean, unambiguous power-law [37]; so a very good estimate of n can be made with a simple application of least-squares. Even for this problem however, looking more closely we are faced with complications:

- Using the experimental data: noise, measurement error and uncertain boundary conditions (notably most experiments are of grid-turbulence, an experimental *model* of isotropic decay);
- Using DNS data: averaging error, finite domain size;
- An initial transient before the power-law is established, which must be excluded by some criteria;
- The sensitivity of the asymptotic decay rate to Reynolds number and the initial spectrum [69].

However these issues are relatively easily handled, and statistical modeling choices – including choice of data-sets, transient removal, etc. – made here will have a relatively minor effect the estimate of n. On the other hand, in Section 7.3.1.2 the variety of behavior observed in the free-shear flows, and the lack

of a clear plateau of $u'v'/k$ in some cases, makes the existence of a universal value for C_μ less clear. Here statistical modeling choices are more likely to influence the result. More critically the data indicates a range of true C_μ in reality, whereas the model specifies a single, true C_μ. This is another example of model inadequacy, since the model is unable to represent reality (at any values of the parameters).

Finally, and most relevant for practical engineering: the calibrations performed above depend on the ability to simplify the equations to the point where an analytic solution is available. As such re-calibration of a model against cases of practical interest is out of reach. Neither is it clear how calibration may be achieved if the required quantities are not available in the calibration data set (e.g. because Reynolds-stresses are not measured). Bayesian calibration approaches will address all these limitations, as well as giving uncertainty information on the parameters.

7.3.2 Bayesian calibration of closure model coefficient

In this section will perform Bayesian calibration of the $k-\epsilon$ model against experimental data from flat-plate boundary-layers at various pressure gradients, following Edeling et al. [12], see that reference for more detail. The work by Oliver and Moser [40] was the first work of this kind.

7.3.2.1 RANS with $k-\varepsilon$ closure model

The $k-\epsilon$ model for a boundary-layer reduces to [64]:

$$\frac{\partial k}{\partial t} + \bar{u}\frac{\partial k}{\partial x_1} + \bar{v}\frac{\partial k}{\partial x_2} = \nu_T \left(\frac{\partial \bar{u}}{\partial x_2}\right)^2 - \varepsilon + \frac{\partial}{\partial x_2}\left[\left(\nu + \frac{\nu_T}{\sigma_k}\right)\frac{\partial k}{\partial x_2}\right], \tag{7.15a}$$

$$\begin{aligned}\frac{\partial \varepsilon}{\partial t} + \bar{u}\frac{\partial \varepsilon}{\partial x_1} + \bar{v}\frac{\partial \varepsilon}{\partial x_2} = C_{\varepsilon 1} f_1 \frac{\varepsilon}{k}\nu_T \left(\frac{\partial \bar{u}}{\partial x_2}\right)^2 - C_{\varepsilon 2} f_2 \frac{\varepsilon^2}{k} + E \\ + \frac{\partial}{\partial x_2}\left[\left(\nu + \frac{\nu_T}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_2}\right],\end{aligned} \tag{7.15b}$$

where the eddy-viscosity is

$$\nu_T = C_\mu f_\mu \frac{k^2}{\varepsilon}.$$

Note that true turbulence dissipation $\varepsilon^\star$ takes a nonzero, approximately constant value ε_0 at solid walls – the $k-\varepsilon$ model, solves for ε, the offset from this value: $\varepsilon^\star \simeq \varepsilon_0 + \varepsilon$. The system (7.15a)-(7.15b) contains several closure coefficients and empirical damping functions, which act directly on these coefficients. Without the damping functions the $k-\varepsilon$ model would not give accurate predictions in the viscous near-wall region [64]. The Launder-Sharma $k-\varepsilon$ model [30] is

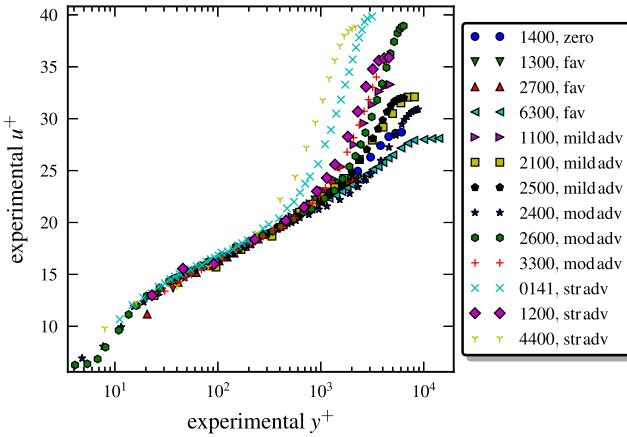


FIGURE 7.4 The experimental data from [6]. Reproduced from [12].

obtained by specifying the damping functions:

$$f_\mu = \exp\left[-3.4/(1+Re_T/50)\right], \quad f_1 = 1,$$
$$f_2 = 1 - 0.3\exp\left[-Re_T^2\right], \quad \varepsilon_0 = 2\nu\left(\frac{\partial\sqrt{k}}{\partial x_2}\right)^2, \quad E = 2\nu\nu_T\left(\frac{\partial^2\bar{u}}{\partial x_2^2}\right)^2, \tag{7.16}$$

where $Re_T \equiv k^2/\varepsilon\nu$. In the case of the Launder-Sharma $k-\varepsilon$ model, the closure coefficients have the following values

$$C_\mu = 0.09, \quad C_{\varepsilon 1} = 1.44, \quad C_{\varepsilon 2} = 1.92, \quad \sigma_k = 1.0, \quad \sigma_\varepsilon = 1.3, \tag{7.17}$$

and we will obtain joint posterior distribution on all 5 parameters.

7.3.2.2 *Experimental data for flat-plate boundary-layers*

The data set consists of 13 selected experiments from the 1968 AFOSR-IFP-Stanford conference proceedings [6], which includes a number of flat-plate turbulent boundary layer flows with various pressure gradients artificially applied by adjusting the shape of the upper-wall of the test-section. The mean velocity profiles are measured via hot-wire anemometry. In [63], the flows are qualitatively identified as being "favorable", "mildly adverse", "moderately adverse" and "strongly adverse", based on the velocity profile shape compared to the zero-pressure gradient case. Most cases are in- or close to- equilibrium conditions, with the exception of some of the strongly adverse pressure gradients. The experimental velocity profiles of all cases are shown in Fig. 7.4.

In the follow we calibrate the parameters of the $k-\varepsilon$ model against each of the 13 cases individually. Despite the fact that dataset represents a very limited

spectrum of turbulent flows – we will observe significant variability in calibrated parameter values.

7.3.2.3 Likelihood and model inadequacy

We do not expect the $k-\varepsilon$ model to be capable of reproducing all these BL profiles exact, even given optimal parameter values per case. As such, analogously to Section 7.2.3, we use a model-inadequacy term in the statistical model. Since velocity at the wall is zero, it's convenient to use a *multiplicative* term, which preserves this property – compare with the additive model inadequacy of Section 7.2.3. Given velocity-profile data for a single flow $\boldsymbol{z}$ we model the data-to-simulation relationship as [12]:

$$\boldsymbol{z} = \eta(\boldsymbol{y}^+) \cdot u^+(\boldsymbol{y}^+, \boldsymbol{t}; \boldsymbol{\theta}) + \boldsymbol{\epsilon} \tag{7.18}$$

where $\boldsymbol{y}^+$ are the wall-distances of the velocity measurements, and all quantities are formulated in wall-units (without loss of generality). The term $\boldsymbol{t}$ refers to boundary-conditions (including the imposed pressure-gradient), which vary from case to case. We model the measurement error as:

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_z^2 I),$$

assuming that any quantifiable bias has already been removed from the measurements in post-processing, and that error at all measurement locations is identical and uncorrelated.[7]

Model-inadequacy is modeled with the Gaussian process is in Section 7.2.3 [26]:

$$\eta(\boldsymbol{y}^+) \sim \mathcal{GP}(1, c_\eta)$$

i.e. with unit mean and covariance function chosen as:

$$c_\eta(y^+, y^{+\prime} \mid \gamma) := \sigma^2 \exp\left[-\left(\frac{y^+ - y^{+\prime}}{10^\alpha \ell}\right)^2\right],$$

where y^+ and $y^{+\prime}$ represent two measurement points along the velocity profile, and $\ell = 5.0$ is a user-specified length scale in wall-units. The smoothness of the model-inadequacy term is controlled by the correlation-length parameter α, which controls how rapidly (in y^+ units) the discrepancy between the model predictions and reality can vary. Finally σ represents the magnitude of the model inadequacy (in velocity units). Defining the hyper-parameters $\boldsymbol{\gamma} := (\sigma, \alpha)$ directly would require *a priori* knowledge of the model inadequacy, which is generally not available. As such we treat them as stochastic *hyper-parameters*

[7] A situation in which the "uncorrelated" assumption is clearly inaccurate would be Particle-Image Velocimetry (PIV) measurements with overlapping interrogation windows. In that case a correlated error can be specified with $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \Sigma_z)$, see [51].

(parameters of random terms in the statistical model), and identify them simultaneously with the turbulence model parameters $\boldsymbol{\theta}$. The Bayesian update equation becomes:

$$\rho(\boldsymbol{\theta}, \boldsymbol{\gamma} \mid z) \propto \rho(z \mid \boldsymbol{\theta}, \boldsymbol{\gamma}) \cdot \rho_0(\boldsymbol{\theta}) \cdot \rho_0(\boldsymbol{\gamma}), \tag{7.19}$$

and so priors on the hyper-parameters are required, and two additional quantities must be estimated from the data – but otherwise nothing changes.

The expression for the likelihood follows directly from (7.18). First note that since $\eta(\cdot)$ is a Gaussian process, $\zeta(\boldsymbol{y}^+) := \eta(\boldsymbol{y}^+) \cdot u^+(\boldsymbol{y}^+)$ is also a Gaussian process with modified mean and covariance functions:

$$\begin{aligned} \zeta \mid \boldsymbol{\theta}, \boldsymbol{\gamma} &\sim \mathcal{GP}(\mu_\zeta, c_\zeta) \\ \mu_\zeta(y^+ \mid \boldsymbol{\theta}) &= u^+(y^+, \boldsymbol{t}; \boldsymbol{\theta}) \\ c_\zeta(y^+, y^{+\prime} \mid \boldsymbol{\theta}, \boldsymbol{\gamma}) &= u^+(y^+, \boldsymbol{t}; \boldsymbol{\theta}) \cdot c_\eta(y^+, y^{+\prime} \mid \boldsymbol{\gamma}) \cdot u^+(y^{+\prime}, \boldsymbol{t}; \boldsymbol{\theta}). \end{aligned} \tag{7.20}$$

Evaluating the Gaussian process at the observation locations $\boldsymbol{y}^+$ results in a multivariate Gaussian random variable $\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \boldsymbol{\gamma}$:

$$\boldsymbol{\zeta} \mid \boldsymbol{\theta}, \boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{u}^+(\boldsymbol{\theta}), K_\zeta), \quad \left[K_\zeta\right]_{ij} := u^+(y_i^+; \boldsymbol{\theta}) \cdot c_\eta(y^{+,i}, y^{+,j} | \boldsymbol{\gamma}) \cdot u^+(y_j^+; \boldsymbol{\theta}),$$

i.e. all terms remain Gaussian. As a consequence the expression for the likelihood is still simple:

$$\begin{aligned} p(z \mid \boldsymbol{\theta}, \boldsymbol{\gamma}) &\propto \exp\left[-\frac{1}{2}\boldsymbol{d}^T K^{-1} \boldsymbol{d}\right], \\ \boldsymbol{d} &:= z - u^+(\boldsymbol{y}^+; \boldsymbol{\theta}) \\ K &:= \sigma_z^2 I + K_\zeta, \end{aligned} \tag{7.21}$$

where $\boldsymbol{d}$ is the data-misfit.

7.3.2.4 Priors on model coefficients and hyper-parameters

Priors – for both the closure coefficients $\boldsymbol{\theta}$ and hyper-parameters $\boldsymbol{\gamma}$ – are chosen as independent uniform distributions. The choice of interval end-points was made based on three factors: (i) the spread of coefficients recommended in the literature, (ii) the range of coefficients for which the solver was stable, and (iii) avoidance of apparent truncation of the posterior at the edge of the prior domain. To explain the latter: a consequence of Bayes rule (7.19) is that values that have zero probability under the prior, will have zero probability under the posterior, and this is true regardless of the preponderance of evidence from the data. As such zero-priors should only be chosen where a *certainty of impossibility* exists – a common instance is the impossibility of negative values for provably positive quantities. Here we violate this principle in the aid of maintaining stability of the solver. As a check that the prior is not unduly constraining the poste-

rior, we examine the posteriors later to check truncation at the edge of the prior domain.

As an implementation note, we constrain the value of $C_{\varepsilon 1}$ using:

$$C_{\varepsilon 1} = \frac{C_{\varepsilon 2}}{\mathcal{P}/\varepsilon} + \frac{\mathcal{P}/\varepsilon - 1}{\mathcal{P}/\varepsilon}$$

with fixed $\mathcal{P}/\varepsilon = 2.09$ following [44]. Also we don't use σ_ε as a parameter, rather the Kármán constant κ, and the relation

$$\kappa^2 = \sigma_\varepsilon \sqrt{C_\mu}(C_{\varepsilon 2} - C_{\varepsilon 1}).$$

To be clear: this is not equivalent to specifying a uniform prior on σ_ε; indeed in doing this we are imposing dependence between σ_ε and other coefficients in the prior.

The distributions we use for the remaining parameters are:

$$C_{\varepsilon 2} \sim \mathcal{U}(1.15, 2.88), \quad C_\mu \sim \mathcal{U}(0.054, 0.135), \quad \sigma_k \sim \mathcal{U}(0.450, 1.15), \tag{7.22}$$

$$\kappa \sim \mathcal{U}(0.287, 0.615), \quad \sigma \sim \mathcal{U}(0, 0.1), \quad \log\alpha \sim \mathcal{U}(0, 4), \tag{7.23}$$

all independently.

7.3.2.5 Calibration results

The simulation code used is Wilcox's EDDYBL [64]. Sampling from the posterior is performed directly with MCMC (see Section 7.2.4.1) implemented in `PyMC` [41]. MCMC is computationally tractable in this case, since the simulation reduces to one spatial dimension. For more complex flows, a surrogate modeling approach would be tractable, since the dimension of the parameter space is low.

The posterior $\rho(\boldsymbol{\theta}, \boldsymbol{\gamma} \mid z)$ is a six-dimensional pdf for each of the 13 datasets. Visualization of a single pdf can be achieved by plotting all 1d and 2d marginals:

$$\rho(\boldsymbol{\theta}_i \mid z) = \int \rho(\boldsymbol{\theta}, \boldsymbol{\gamma} \mid z)\, \mathrm{d}\boldsymbol{\gamma}\, \mathrm{d}\theta_{\sim i}, \quad \rho(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j \mid z) = \int \rho(\boldsymbol{\theta}, \boldsymbol{\gamma} \mid z)\, \mathrm{d}\boldsymbol{\gamma}\, \mathrm{d}\theta_{\sim i,j},$$

where the integrals are over indices other than i, and i, j respectively: $\sim i := \{k : k \neq i\}$, etc. Given samples of the posterior from MCMC the marginal pdfs can be estimated by histograms (or KDEs) where the values of the marginalized parameters are ignored. A representative instance of the result can be seen in Fig. 7.5 for a moderately adverse pressure-gradient case. The plot shows most likely values and variances of all parameters individually, as well as dependencies between parameters – e.g. a weak negative correlation between the values of $C_{\varepsilon 2}$ and C_μ is visible – that are the result of interactions of the parameters

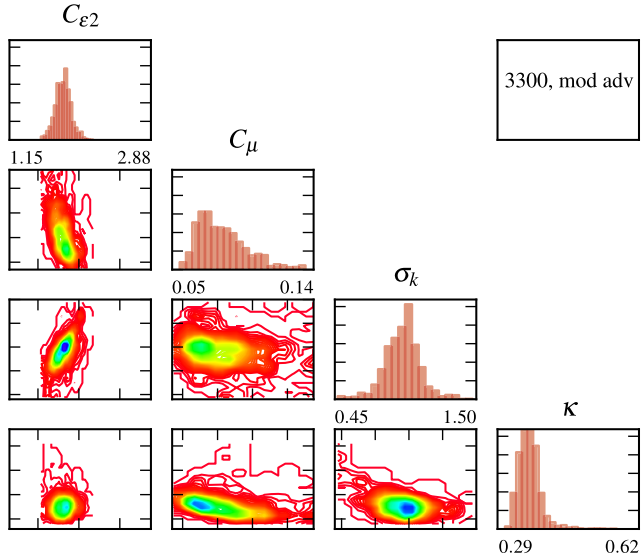


FIGURE 7.5 Joint posterior density on model coefficients, visualized as all pairs of 2d-marginal distributions – for a moderately adverse pressure-gradient (flow 3300). Reproduced from [12].

within the model. The relatively weak dependency structure seen here is partly a consequence of the model parameters being careful designed to represent different aspects of turbulence.

From the 1d marginals it can be seen that $C_{\varepsilon 2}$ and κ are relatively well-identified from the data (small posterior variance), and C_μ and σ_k are less well-identified. This is only partially consistent with our intuition: the Kármán constant κ refers directly to the gradient of velocity in the log-region of boundary-layers, so should be identified well. On the other hand from the discussion of canonical flows above, we would expect C_μ to be more important for the BL than $C_{\varepsilon 2}$. The discrepancy might be related to higher sensitivity of the model with respect to some parameters than others, or the different prior ranges chosen for the parameters. Nonetheless the posterior gives specific information over the range of parameter values that are consistent with the data.

Comparing posteriors between cases, we look now exclusively at 1d marginals in Fig. 7.6. Far from each case giving consistent estimates of parameters, we see significant case dependence. Baring in mind that all cases are flat-plates, this is surprising. Of all parameters $C_{\varepsilon 1}$ is best identified, and also most consistent, with all posteriors sharply peaked within a range of $1.3 - 1.5$; followed by $C_{\varepsilon 2}$. In comparison C_μ is hardly informed at all beyond the prior. Most unsettling is that the κ is clearly identified in most cases, but with widely varying peak values. This suggests that the parameter (at least as it applies to the $k - \varepsilon$ model) lacks universality. The hyper-parameter σ is hardly identified in any case, indicating a lack of evidence about the magnitude of the model inad-

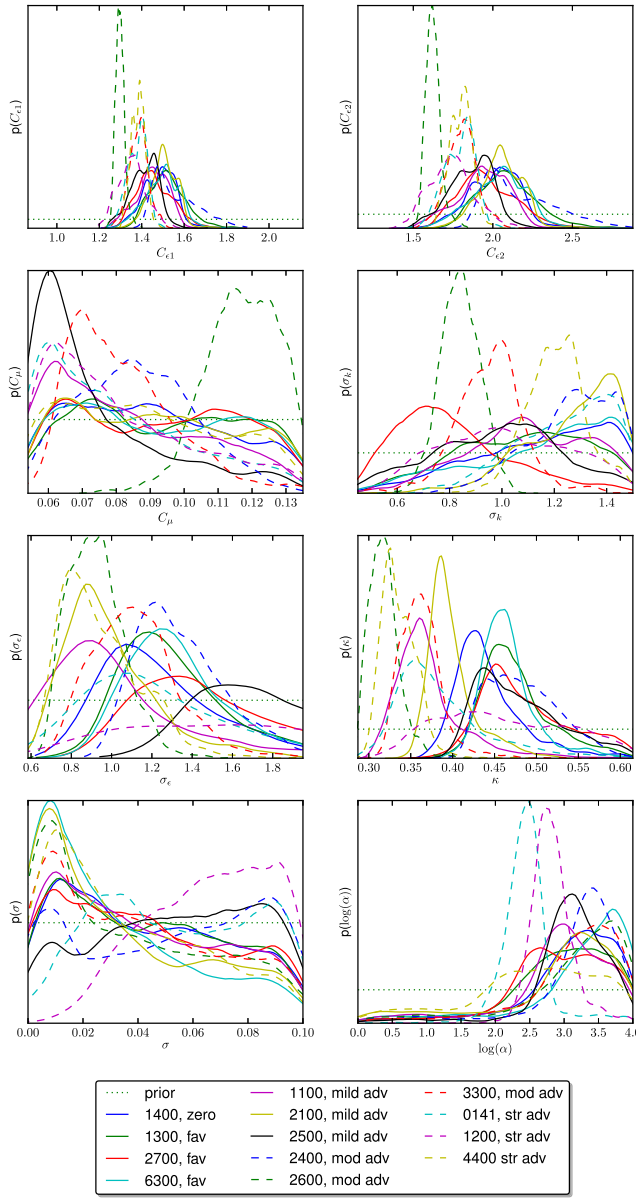


FIGURE 7.6 The marginal posterior distributions of the model coefficients and hyper-parameters for the 13 boundary-layer cases.

equacy – perhaps a result of the large number of parameters and limited data in each case. On the other hand the length-scale of the inadequacy is consistently estimated to be towards the upper bound.

Further analysis of these results can be found in [12]. It is clear though that identifying universal parameters for $k-\varepsilon$ for the full range of flows considered is unlikely. The following section considers Bayesian model selection averaging, which is one means of addressing this.

7.3.3 Bayesian model selection and model averaging

As with other complex physical systems, there exist many available models for RANS closure. Notable for their widespread adoption are: $k-\varepsilon$, $k-\omega$ [65], $k-\omega$ SST [36], and Spalart-Allmaras [57], which are all linear eddy-viscosity models (LEVMs), resting on the Boussinesq approximation. Each model has a unique set of coefficients, and though coefficients with the same *name* may appear in multiple models, the meaning of the coefficient (in the sense of its relation to the model predictions) will almost always be distinct.

This – combined with the observations of the previous section (that optimal calibration coefficients vary significantly from test-case to test-case) – leads us to the natural question: What model with what set of coefficients should we use for a particular test-case? There are two cases of this question: 1. Where some (limited) reference data is available for this particular, or a similar, case; and 2. Where the flow considered is *unseen*, i.e. we have no available reference data.

Once again we attack this problem with Bayesian statistics, and obtain *Bayesian Model Selection* (BMS) and *Bayesian Model Averaging* (BMA), which give probabilistic answers to the question.

7.3.3.1 Theory of BMS and BMA

In essence both BMS and BMA are straightforward applications of the Bayesian calibration methodology of the previous sections. Consider a set of $Q \in \mathbb{N}$ candidate models $\boldsymbol{M} = \{M_1, M_2, \ldots, M_Q\}$, each with its own vector of closure coefficients $\boldsymbol{\theta}^{(q)} \in \mathbb{R}^{N_q}$ for $q \in \{1, \ldots, Q\}$, potentially with differing numbers of coefficients per model. Given data z, the posterior probability of each model conditioned on the data can be obtained by using Bayes' theorem as:

$$\mathbb{P}(M_q|z) = \frac{\rho(z|M_q)\,\mathbb{P}_0(M_q)}{\sum_{i=1}^{Q}\rho(z|M_i)\,\mathbb{P}_0(M_i)} \tag{7.24}$$

where we use $\mathbb{P}(\cdot)$ to indicate the probability mass of a *discrete* r.v., as opposed to $\rho(\cdot)$ for the probability density of a continuous r.v. In (7.24) $\mathbb{P}_0(M_q)$ is the prior, representing the subjective confidence in model M_q (without considering the data); $\rho(z|M_q)$ is the likelihood of data z conditioned on the given model M_q, indicating the probability of observing data z with model M_q at any value of the parameters. The likelihood $\rho(z|M_q)$ is defined in terms of $\rho(z|\boldsymbol{\theta}^{(q)}, M_q)$ as:

$$\rho(z|M_q) = \int \cdots \int \rho(z|\boldsymbol{\theta}^{(q)}, M_q)\,\rho(\boldsymbol{\theta}^{(q)}|M_q)\,d\boldsymbol{\theta}^{(q)} \tag{7.25}$$

whereby the parameters of the model M_q are integrated out; $\rho(\boldsymbol{z}|\boldsymbol{\theta}_i, M_q)$ is the likelihood of data $\boldsymbol{z}$ conditioned on both the specific model M_q and its parameters. This again being the result of a statistical model, for instance:

$$\boldsymbol{z} = u^+(y^+; M_q, \theta^{(q)}) + \epsilon$$

by analogy with (7.18), where now the CSE code u^+ depends on the choice of model and corresponding coefficients. The term $\rho(\boldsymbol{\theta}^{(q)}|M_q)$ in (7.25) is the prior probability density of the parameters $\boldsymbol{\theta}^{(q)}$ conditioned on the given model M_q. This may be chosen uniformly within some interval, as in Section 7.3.2.4, or as the posterior of some previous calibration exercise against a different flow and data-set.

The only difference between this formulation, and that of single-model calibration in Section 7.3.2 is that there are now more variables to identify, M and $\theta^{(q)}$ for all $q \in \{1, \ldots, Q\}$, and that one of these variables is discrete rather than continuous. Note that by adding so many extra variables, the demands on the data are increased, and typically we learn less about any given parameter (given the same amount of data).

Given these preliminaries, **Bayesian model selection (BMS)** involves computing $\mathbb{P}(M_q|\boldsymbol{z})$, and choosing a single model $\hat{M} \in \boldsymbol{M}$ satisfying

$$\mathbb{P}(\hat{M}|\boldsymbol{z}) \geq \mathbb{P}(M_q|\boldsymbol{z}) \qquad \forall q \in \{1, \ldots, Q\},$$

i.e. a maximum *a posteriori* estimate. The coefficients $\hat{\boldsymbol{\theta}}$ chosen based on the posterior $\rho(\hat{\boldsymbol{\theta}} \mid \boldsymbol{z}, \hat{M})$ as before, which is readily computed, given the existing computation of (7.25).

Bayesian model averaging (BMA) goes a step further, using a weighted linear combination of all models to make predictions about unmeasured quantities. Let $\hat{z}$ be an unmeasured quantity related to $\boldsymbol{z}$. For instance $\boldsymbol{z}$ might be flow velocities measured within a limited window, and $\hat{z}$ the velocity at another unmeasured coordinate in the same flow. This is a common task in *data-assimilation*. BMA models $\hat{z}$ as:

$$\rho(\hat{z} \mid \boldsymbol{z}) = \sum_{q=1}^{Q} \left[\int \cdots \int \rho(\hat{z} \mid M_q, \boldsymbol{\theta}^{(q)}) \rho(\boldsymbol{\theta}^{(q)} \mid M_q, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{\theta}^{(q)} \right] \cdot \mathbb{P}(M_q \mid \boldsymbol{z}), \tag{7.26}$$

which is the posterior predictive distribution of $\hat{z}$ given $\boldsymbol{z}$. In this expression *all* models at *all* values of their coefficients contribute to the result, which necessitates a great many evaluations of the CSE code. The expression can be simplified by replacing $\rho(\boldsymbol{\theta}^{(q)} \mid M_q, \boldsymbol{z})$ by a Dirac-δ function at the MAP estimate, reducing the computational cost to Q CSE solves, at the expense of neglecting the role of parameter variability in the uncertainty in $\hat{z}$ [13].

Regarding prediction of unmeasured flows (with no $\boldsymbol{z}$ available for the new flow), (7.26) can be applied with model probabilities $\mathbb{P}(M_q \mid \boldsymbol{z})$, and parameter

distributions $\rho(\boldsymbol{\theta}^{(q)} \mid M_q, \boldsymbol{z})$ taken from a similar flow for which $\boldsymbol{z}$ is available. In this way genuine predictions with uncertainty estimates can be achieved [10].

7.3.3.2 Application to turbulence modeling

As an example of BMS/BMA applied to turbulence modeling we turn to work by Edeling et al. [10], who applied it to the same set of flat-plate boundary-layer (BL) flows seen in Section 7.3.2.2. This study considers five turbulence models ($k-\omega$, $k-\varepsilon$, the stress–ω model, Baldwin-Lomax, and Spalart-Allmaras) applied to each of the 14 BL cases. The model probabilities are computed according to (7.24), where once again MCMC is used to sample from the posteriors. To clarify the role of the 14 cases, each is termed a "scenario" who boundary-conditions and material properties are summarized by S_j for $j \in \{1, \ldots, 14\}$. As such we compute model probabilities $\mathbb{P}(M_i|S_j, \boldsymbol{z}_j)$ conditioned on both data and scenario. These probabilities are plotted in Fig. 7.7. In the plot the scenarios are clustered into favorable/adverse cases, yet there is still no detectable pattern in the best choice of model per case. In many scenarios a single model dominates, reducing most other model probabilities to (close to) zero. This noisy result is likely a consequence of all models performing a very good fit of a BL, and the model probabilities consequently depend on small differences.

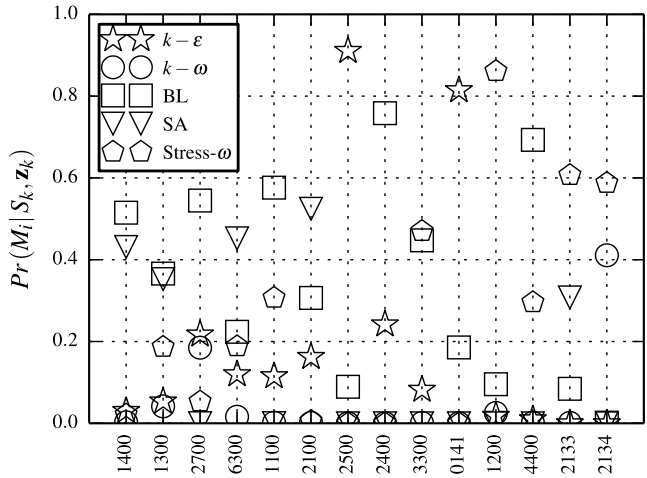
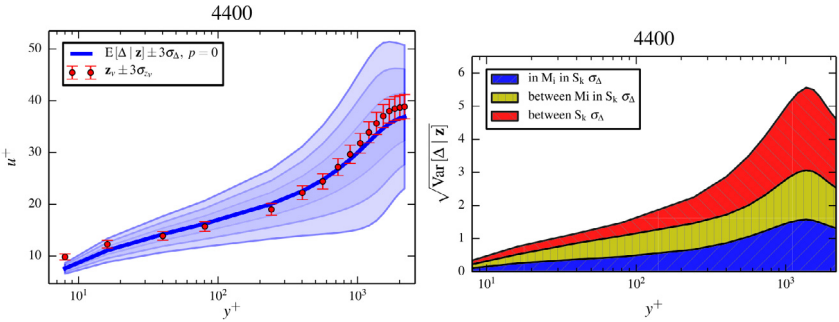


FIGURE 7.7 Posterior probabilities of five turbulence models for 14 calibration datasets. Figure reproduced with permission from [10].

These results do not allow confident selection of a single model – though that may well be the case for a more complex collection of flows. However we can still make predictions under uncertainty using multiple models. In particulars, since we have five models, each calibrated for 14 scenarios, we can make predictions accounting for the full implied range of models, see [10] for details. Fig. 7.8(a) shows the BMA prediction for an adverse pressure-gradient



(a) Prediction with standard-deviation. (b) Uncertainty breakdown.

FIGURE 7.8 BMA posterior predictive distribution for case 4400 (strong adverse pressure-gradient). Figure reproduced with permission from [10].

case (4400) which was excluded from the training. While predictions of the mean correspond to the experimental results (within experimental uncertainty), the variance is unjustifiably large. Fig. 7.8(b) give some insight into the source of this large variance by decomposing it into three parts: (a) due to coefficient variability in a single model, for a single scenario (blue), (b) due to variability between model predictions for a single scenario (yellow), and (c) due to the variability in prediction between scenarios (red). All contribute roughly equally to the total variance in this example. The assumption contributing to this is that all of the training scenarios are equally relevant to the prediction scenario. By choosing only those training scenarios that correlate well with the prediction, the latter source of uncertainty can be dramatically reduced [10]. Nevertheless, under the current set of statistical and modeling assumptions, the level of confidence in RANS models is empirically quite low.

7.3.4 Discussion

In the turbulence model calibration activities in this section, we have demonstrated both the effectiveness of the classical approach of canonical flows selection and fitting, and the possibilities offered by the Bayesian approach of explicit statistical modeling and evaluation of uncertainties. In the latter we have incorporated information about the accuracy of the data, and make explicit statistical choices regarding the relationship of the (CFD) model to that data (including model inadequacy). The result of calibration (after applying Bayes' rule) has been a joint posterior distribution on the model coefficients, C_μ, $C_{\varepsilon 2}$, etc., which characterizes our knowledge about the values of these parameters given the data, under our priors. This joint density has been plotted via marginal distributions, and has been propagated through the model to get the *posterior predictive distribution*, which has given uncertainties on model predictions for unseen cases.

The basic procedure outlined can be applied to flows of any complexity, and delivers a probabilistic characterization of the model accuracy. The meaningfulness of this result depends to a large extent on the validity of the statistical modeling, both in the prior and likelihood, and in particular the effect of neglecting model inadequacy (when present) can lead to serious misrepresentations of accuracy. The downside is the high computational price to be paid: the CFD solver must be run a great many times, both in the calibration phase, and in the posterior predictive phase (see Section 7.2.4).

For a given turbulence closure model – or model*s* (in the case of BMA and BMS) – this represents the best that we can do in terms of calibration. However there remain fundamental structural assumptions in RANS models that are not addressed by the above procedure: notably the Boussinesq approximation, and the particular model form. This will be the focus of the next section.

7.4 Bayesian estimation of Reynolds stress anisotropy

So far we have considered only adjusting parameters of models ("calibration"), while acknowledging that the model-forms themselves are inadequate to fully describe the physics of turbulence. Notably the Boussinesq hypothesis is the foundation of all the closure models discussed so far – and turbulence anisotropy is neglected. As a consequence, even when performing model selection with BMS, and averaging with BMA, we have selected or averaging amongst a set of models that have a common source of modeling inadequacy. Furthermore this inadequacy is not well represented by the η term in (7.18), since that treats the inadequacy as a post-hoc correction to the model's prediction of the observed quantities, whereas the origin of the inadequacy is *internal* to the closure model, and occurs at the point where the RST is modeled. The mis-predictions of observed quantities are strictly a consequence of this RST modeling, mediated by the transport equations.

Therefore we would like to (i) construct a statistical model which represents the inadequacy at its source, and (ii) generalize the parameterization of the model to allow identification of corrections to Boussinesq. In this section we will infer information about *local* turbulence anisotropy from surface pressure measurements, and then predict unmeasured quantities (velocity profiles) for the same flow. The main challenge is the fact that we are now attempting to infer *local* quantities (spatially-varying stress corrections), instead of global coefficients. I.e. we infer *fields* in the sense of field-inversion [42], which makes extracting information from the posterior much more challenging. Additional challenges are the difficultly of constructing a prior for the tensor-valued correction to the anisotropy, and the limited information surface pressure provides regarding the turbulence.

7.4.1 Learning Reynolds-stress anisotropy from surface pressure

In this section we follow closely the work of Belligoli et al. [4], in which our goal is *data assimilation*. I.e. we do not aim to create of a new or calibrated closure model, but rather: for a specific steady-state flow, with limited measurement observations, estimate model error and unmeasured flow quantities. As such we are only concerned with the errors the baseline turbulence closure is making in this particular flow, rather than in general.

As before we assume knowledge of $\boldsymbol{z} \in \mathbb{R}^P$ experimentally obtained observations of true quantities $\boldsymbol{z}^\star \in \mathbb{R}^P$, with independent Gaussian noise of known variance, and zero bias, i.e.

$$\boldsymbol{z} = \boldsymbol{z}^\star + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_{\text{exp}}^2 \boldsymbol{I}),$$

where it is assumed that any known bias has already been removed from the data. Generalization of this noise model to correlated and spatially varying noise is straightforward [50]. Further assume that the full (compressible) flow state is characterized the variables $\boldsymbol{w} = [\boldsymbol{\rho}, \boldsymbol{\rho u}, \boldsymbol{\rho E}, \boldsymbol{k}, \boldsymbol{\omega}] \in \hat{\mathcal{W}}$, defined on a computational mesh, where $N \in \mathbb{N}$ is the size of the mesh. Given these, the observation operator $H : \hat{\mathcal{W}} \to \mathbb{R}^P$ extracts the observed quantities by interpolation on the mesh, and therefore H is usually a simple linear operator which introduces negligible additional error. We can formulate our simulation's prediction of $\boldsymbol{z}^\star$ as:

$$\boldsymbol{z} = H\boldsymbol{w} + \boldsymbol{\epsilon}, \quad \text{s.t.} \quad \mathcal{R}(\boldsymbol{w}; \boldsymbol{\theta}) = 0, \tag{7.27}$$

where $\mathcal{R}(\cdot;\cdot) = 0$ represents the PDE constraint on the state, given the parameters of the model $\boldsymbol{\theta}$. These parameters may include boundary-conditions, material properties, etc., but as in the previous sections, we treat these as given and deterministic, and use $\boldsymbol{\theta}$ to describe only the parameterization of the turbulence model.

In contrast to previous sections however, now we consider $\boldsymbol{\theta}$ to be defined at every mesh-point, and as such it can be a much more general parameterization than the closure coefficients. For instance we could define θ to describe the tensor-valued Reynolds-stress at each mesh point – thereby circumventing the Boussinesq assumption in the turbulence model. In a continuous setting then the r.v. describing uncertainty in $\boldsymbol{\theta}$ is a random process $\{\boldsymbol{\Theta}(\boldsymbol{x})\}_{\boldsymbol{x}}$, perhaps scalar-, vector-, or tensor-valued, which can be discretized on the computational mesh to give a high-dimensional r.v. $\boldsymbol{\Theta} : \Omega \to \mathbb{R}^N$ (for a scalar process). Three particular model parameterizations are considered in the following subsections.

If θ represents Reynolds stresses, then the posterior distribution $\boldsymbol{\Theta} \mid \boldsymbol{z}$ itself is of interest. However usually, our primary interest is in other unmeasured flow quantities $\hat{z} \in \mathbb{R}$. Our predictive model for these is:

$$\hat{z} = \hat{H}\boldsymbol{w}, \quad \text{s.t.} \quad \mathcal{R}(\boldsymbol{w};\ \boldsymbol{\Theta} \mid \boldsymbol{z}) = 0, \tag{7.28}$$

where $\hat{H}$ extract the desired unmeasured quantity from the state. A usual this is a stochastic prediction, with the distribution of $\hat{z}$ reflecting the uncertainty

represented by the posterior $\boldsymbol{\Theta} \mid \boldsymbol{z}$, with in turn represents the state of knowledge about the turbulence model.

The assumption implicit in the statistical model (7.27) is that: with correct choice of $\boldsymbol{\theta} \in \mathbb{R}^{M \times N}$, $\boldsymbol{z}^\star$ will be recovered exactly. As such sufficient generality in the parameterization of the closure model is needed, for this to be a plausible assumption. Note that we do not use a Gaussian-process $\eta(\cdot)$ to represent model inadequacy as in (7.18), and the above construction in (7.27) makes clear why: such a η term refers directly to the error made by the observation operator H, which – by assumption – is negligible. Furthermore it does not speak to the error in prediction (7.28), whereas the uncertainty in $\boldsymbol{\theta}$ does.

As we can see, in terms of formulation this problem is not very different from the parameter calibration problem of Section 7.3.2. The major practical difference is in the number of parameters to be identified: five in the case of calibration of $k - \varepsilon$, versus $\mathcal{O}(N)$ here. In we wish to find the MAP estimate, we are faced with a very high-dimensional PDE-constrained optimization problem. The only effective methods for solving these problems are based on adjoint methods, which have been extensively developed in CFD for the purposes of aerodynamic design [43], and such a method will be used in the following. By restricting ourselves to the MAP estimate, we lose the uncertainty information contained in the posterior. This may be recovered using more general methods of *variational Bayes* [38].

It remains to consider the particular choice of model parameterization. We consider three different particular parameterizations of Menter's $k - \omega$ SST model in the following sections.

7.4.1.1 Turbulence production perturbation (TPP)

Following Parish and Duraisamy [42] and Singh and Duraisamy [54], for a general $k - \omega$ closure, the structure of the transport equations in a schematic form is:

$$\frac{\mathrm{D}k}{\mathrm{D}t} = \mathcal{P}_k - \mathcal{D}_k + \mathcal{T}_k, \tag{7.29}$$

$$\frac{\mathrm{D}\omega}{\mathrm{D}t} = \theta_{\mathrm{TPP}}(\boldsymbol{x})\mathcal{P}_\omega - \mathcal{D}_\omega + \mathcal{T}_\omega, \tag{7.30}$$

where $\mathcal{P}$, $\mathcal{D}$ and $\mathcal{T}$ are the production, destruction, and cross-production/diffusion terms respectively, each of which is modeled as a function of the available quantities, k, ω and mean-velocity gradients, and where the spatial field $\theta_{\mathrm{TPP}} : \mathbb{R}^3 \to \mathbb{R}$ has been introduced to modify the balance between production and destruction, which can be the source of error in e.g. incipiently separating flow. The rationale for modifying the ω-equation in this way (rather than the k-equation), is the greater theoretical basis for the PDE (7.29), which mimics the exact equation for T.K.E.; compared to (7.30) which is largely empirical, even before the addition of θ_{TPP}.

As a prior we take:

$$\theta_{\text{TPP}} \sim \mathcal{N}(1, \sigma_1^2 I),$$

so that in the absence of data, the original SST model is the model of maximum probability, and no correlation between mesh points is assumed *a priori*. The MAP estimate is found by solving:

$$\theta_{\text{TPP}}^{\text{MAP}} := \arg\max_{\theta} \rho(\theta \mid \boldsymbol{z}) = \arg\min_{\theta} \left[\frac{1}{2\sigma_{\text{exp}}^2 P} \| H\boldsymbol{w}(\theta) - \boldsymbol{z} \|^2 + \frac{1}{2\sigma_1^2 N} \| \boldsymbol{\theta} - 1 \|^2 \right],$$

where the likelihood is responsible for the data-misfit term, and the prior for the regularization. The relative sizes of σ_1^2 and σ_{exp}^2 are seen to control the level of Tikhonov regularization. The second equality above follows by working with log-likelihood and log-prior pdfs – this is a practical necessity to avoid values of the objective function (and its gradients) that are extremely close to zero.

7.4.1.2 Anisotropy tensor perturbation (ATP)

The previous correction (TPP) is still limited by the Boussinesq hypothesis. To achieve greater generality it is necessary to correct the model of the Reynolds Stress Tensor (RST) $\boldsymbol{\tau} = \tau_{ij} := \overline{u_i' u_j'}$ in the Reynolds-averaged Navier-Stokes equations:

$$\frac{\mathrm{D}U_i}{\mathrm{D}t} = -\frac{1}{\rho} \nabla_i p + \nu \nabla^2 U_i - \nabla_j \tau_{ij}, \tag{7.31}$$

where U_i is the mean-velocity, and u_i' the fluctuating part. It is useful to decompose τ_{ij} into factors separating its amplitude, shape, and orientation [14]. The Reynolds stress tensor is a symmetric positive-semi-definite tensor[8] whose anisotropic and component is

$$\boldsymbol{b} := \frac{\boldsymbol{\tau}}{2k} - \frac{1}{3}\boldsymbol{I},$$

where k characterizes the magnitude of $\boldsymbol{\tau}$, and the symmetric tensor $\boldsymbol{b}$, is known as the *normalized anisotropy tensor*. A turbulence model is called *realizable* if its model for $\boldsymbol{\tau}$ is always positive semi-definite.

We aim to parameterize $\boldsymbol{b}$ with $\boldsymbol{\theta}$, and then use a k-equation to compute $\boldsymbol{\tau}$ in (7.31). We start from the eigendecomposition

$$\boldsymbol{b} = X \Lambda X^T,$$

where $X \in \mathbb{R}^{3\times 3}$ is the orthonormal matrix of eigenvectors of $\boldsymbol{b}$ and Λ is the diagonal matrix with nonzero entries the (real) eigenvalues λ_i. By construction

[8] By construction, since $\tau_{ij} := \frac{1}{K} \sum_{k=1}^{K} u_i' u_j'$ is an ensemble average of outer-products $\boldsymbol{u} \otimes \boldsymbol{u}$, each of which is positive semi-definite.

$tr\,(\boldsymbol{b}) = 0$, so $\sum_i \lambda_i = 0$, and there are only two independent invariants of $\boldsymbol{b}$. We represent these invariants using the barycentric map of Banerjee et al. [3]:

$$x_B = C_{1c}x_{1c} + C_{2c}x_{2c} + C_{3c}x_{3c}$$
$$y_B = C_{1c}y_{1c} + C_{2c}y_{2c} + C_{3c}y_{3c}$$

with $C_{1c} = \lambda_1 - \lambda_2$, $C_{2c} = 2(\lambda_2 - \lambda_3)$, and $C_{3c} = 3\lambda_3 + 1$, which describe the character of the turbulence anisotropy (1-component, 2-component, or 3-component) and where x_{ic}, y_{ic} are coordinates of the corners of a triangle in (x, y) space. We then use perturbations of these coordinates $(\Delta x, \Delta y)$ as our first parameters $\boldsymbol{\theta}$. To modify the orientation of $\boldsymbol{b}$ we define a rotation Q via the unit quaternion:

$$\begin{aligned} h &= h_r + h_i\boldsymbol{i} + h_j\boldsymbol{j} + h_k\boldsymbol{k} \\ &= \cos\frac{\phi}{2} + n_1 \sin\frac{\phi}{2}\boldsymbol{i} + n_2 \sin\frac{\phi}{2}\boldsymbol{j} + n_3 \sin\frac{\phi}{2}\boldsymbol{k}, \end{aligned}$$

for a unit vector $\boldsymbol{n}$ and rotation angle ϕ, which make up the remaining components of θ:

$$\theta_{\text{ATP}} := [\Delta x, \Delta y, \phi, \boldsymbol{n}].$$

We assume the priors of each random field of θ_{ATP} to uncorrelated processes with zero mean, and specified variance, e.g. $\Delta x \sim \mathcal{N}(0, \sigma_2^2)$, given which the MAP estimate is:

$$\theta_{\text{ATP}}^{\text{MAP}} = \arg\min_{\theta} \left[\frac{1}{2\sigma_{\text{exp}}^2 P} \| H\boldsymbol{w}(\theta) - \boldsymbol{z} \|^2 + \frac{1}{2\sigma_2^2 N} \|\boldsymbol{\theta}\|^2 \right].$$

7.4.1.3 Random matrix perturbation (RMP)

We consider a final parameterization, again of the Reynolds stress tensor which addresses two deficits of the previous section: 1. that the particular choice of perturbation of $\boldsymbol{b}$ is arbitrary, and 2. that the result is not guaranteed to be realizable. Ideally we would define a probability distribution directly on the space of admissible matrices. This is the case for the random matrix approach [68,56], which models the RST as a random matrix defined on the set $\mathbb{M}^{+0}$ of positive semi-definite matrices.

To define a distribution on this set we use the maximum entropy principle, which gives the most noncommittal distribution (maximizing entropy) while satisfying given constraints.[9] For the set $\mathbb{M}_0^+$ the entropy of a pdf $\rho(\boldsymbol{\tau})$ is

$$S(\rho) = -\int_{\mathbb{M}_0^+} \rho(\boldsymbol{\tau}) \log \rho(\boldsymbol{\tau})\, \mathrm{d}\boldsymbol{\tau}$$

[9] For example, the normal distribution is the maximum entropy distribution on $\mathbb{R}$ for specified mean and variance.

which we maximize subject to a given mean $\mathbb{E}(\boldsymbol{\tau}) = \bar{\boldsymbol{\tau}}$, and a specified *dispersion* δ satisfying $0 < \delta < \sqrt{2}/2$, which is analogous to a measure of variance [56].

In practice, it is convenient to work with a normalized positive definite random matrix whose mean is the identity, i.e. $\mathbb{E}\boldsymbol{G} = I$, so we write $\boldsymbol{\tau} = L_\tau^T \boldsymbol{G} L_\tau$ where $\bar{\boldsymbol{\tau}} = L_\tau^T L_\tau$ is the Cholesky factorization of $\bar{\boldsymbol{\tau}}$. Enforcing the maximum entropy principle on $\boldsymbol{G}$ gives $\boldsymbol{G} = \boldsymbol{L}^T \boldsymbol{L}$ with $\boldsymbol{L}$ upper-triangular and:

1. The off-diagonal elements of $\boldsymbol{L}$ are:

$$\boldsymbol{L}_{ij} = \frac{\delta}{2} w_{ij}, \quad w_{ij} \sim \mathcal{N}(0,1), \quad \forall i \neq j.$$

2. The diagonal elements of $\boldsymbol{L}$ are:

$$\boldsymbol{L}_{ii} = \frac{\delta}{2}\sqrt{2u_i}, \quad u_i \sim \Gamma(\alpha_i, 1),$$

where u_i are gamma-distributed, with pdfs:

$$\rho(u_i) = \mathbb{1}_{\mathbb{R}^+}(u_i) \frac{1}{\Gamma(\alpha_i)} u^{\alpha_i - 1} e^{-u}, \text{ for } i \in \{1,2,3\},$$

where $\alpha_i := \frac{d+1}{\delta^2} + \frac{1-i}{2}$, and $\mathbb{1}_{\mathbb{R}^+}(u) = 1$ if $u \geq 0$ and 0 otherwise.

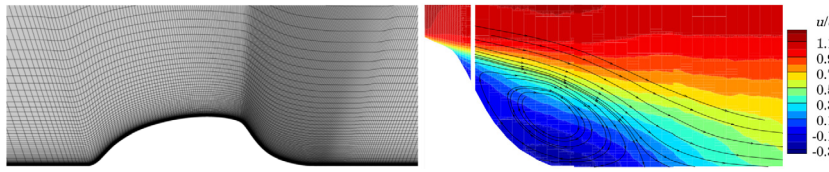
Thus the parameters of our Bayesian calibration problem are $\theta_{\text{RMP}} = [\boldsymbol{L}_{11}, \boldsymbol{L}_{12}, \boldsymbol{L}_{13}, \boldsymbol{L}_{22}, \boldsymbol{L}_{23}, \boldsymbol{L}_{33}]$, the six components of $\boldsymbol{L}$, with prior given by the maximum entropy distribution with mean $\bar{\boldsymbol{\tau}}$ taken as the Boussinesq prediction of $\boldsymbol{\tau}$ at each mesh point, and user specified dispersion δ. As such the MAP estimate is obtained as $\theta_{\text{RMP}}^{\text{MAP}} = \arg\max \mathcal{J}(\theta)$ with

$$\mathcal{J}(\theta) = \frac{1}{2\sigma_{\text{exp}}^2 \boldsymbol{P}} \| H\boldsymbol{w}(\theta) - \boldsymbol{z} \|^2 + \sum_{j=1}^{N} \left[\frac{\boldsymbol{L}_{11}^2 + \boldsymbol{L}_{12}^2 + \boldsymbol{L}_{13}^2 + \boldsymbol{L}_{22}^2 + \boldsymbol{L}_{23}^2 + \boldsymbol{L}_{33}^2}{2\sigma_\delta^2} - \right.$$
$$\left. \log\left(\mathbb{1}_{\mathbb{R}^+}(\boldsymbol{L}_{11,j}) \boldsymbol{L}_{11,j}^{s_1 - 1} \cdot \mathbb{1}_{\mathbb{R}^+}(\boldsymbol{L}_{22,j}) \boldsymbol{L}_{22,j}^{s_2 - 1} \cdot \mathbb{1}_{\mathbb{R}^+}(\boldsymbol{L}_{33,j}) \boldsymbol{L}_{33,j}^{s_3 - 1} \right) \right]$$

where $s_i = \frac{4}{\delta^2} + 1 - i$. Note that the presence of the gamma distributions in the prior results is quite a different form of the regularization term, in particular non-realizable tensors are penalized with $\log 0 = -\infty$.

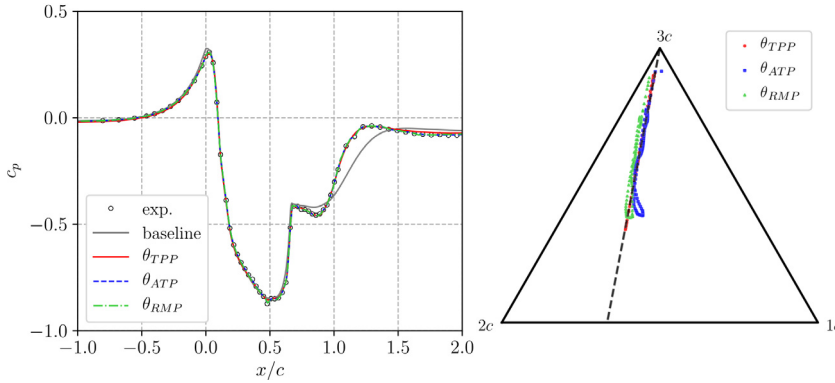
7.4.1.4 Test-case: separated flow behind a hump (2d)

To illustrate the above approach we consider separated flow behind the NASA hump with chord-based Reynolds number of 9.36×10^5 [20]. The wall-resolved mesh is depicted in Fig. 7.9(a), the experimental pressure coefficients on the surface of the hump are used as training data Fig. 7.10(a). The value of σ_{exp} is set to 10^{-3} as specified in the experiment of Greenblatt et al. [20]. PIV measurements



(a) NASA Hump mesh (b) PIV data

FIGURE 7.9 Computation mesh and Particle Image Velocimetry (PIV) data for the NASA hump flow.



(a) Wall pressure coefficient (b) Turbulence anisotropy at $x/c = 0.8$

FIGURE 7.10 Pressure measurements (circles) and RANS predictions, for the baseline (SST) model, and the corrected models (left). Turbulence anisotropy along a line in the wake for the corrected models.

and hot-wire velocity profile measurements are also available, Fig. 7.9(b), and will be held-back test data.

RANS estimates of the surface pressure coefficient over the hump are seen in Fig. 7.10(a). The baseline SST model over-predicts the pressure trough at $x/c \approx 1.3$ (note that $x/c = 0$ corresponds to the hump's leading edge). All three corrected models are remarkable for their ability to reproduce the training data. In particular they capture subtle features such as the pressure kinks at $x/c = 0.5$ and $x/c = 1.2$ that are likely experimental artifacts. This suggests that all three parameterization are sufficiently general – including the simplest TTP, which just adjusts turbulence dissipation locally.

The corrective fields resulting from ATP and RMP are visualized in Fig. 7.11, whereby – for the purposes of comparison – both corrections have been plotted in terms of eigenvalue shifts in the barycentric map. Corrections in the orientation of $\boldsymbol{b}$ were negligible in both ATP and RMP and were not plotting. In this figure – alone in this section – data from some velocity profiles in the wake are used in addition to surface pressure, due to extremely small corrections using surface pressure alone. Corrections to turbulence anisotropy in

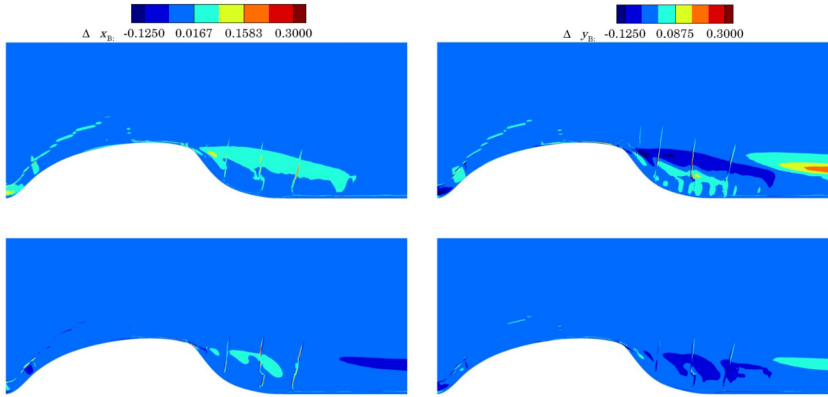


FIGURE 7.11 Perturbations to barycentric coordinates Δx, Δy for the ATP (top) and RMP (bottom) models.

the separated shear layer are most visible, both immediately after separation, as well as during the wake recovery phase. Visualizing the anisotropy correction in the barycentric map, Fig. 7.10(b), this time assimilating pressure-data only, we see that the corrective models give Reynolds stress anisotropy predictions that are extremely close to the baseline SST prediction, with very minor modifications. This attests to the sensitivity of the mean flow to small perturbations of the RST, especially at high Reynolds numbers [61]. As a consequence in none of the scenarios we consider is specific information about the RST obtained, and the calibrated parameters $\boldsymbol{\theta}$ have role, only as model coefficients, not physically meaningful quantities.

The value of the method can nonetheless be demonstrated in the predictions of unmeasured quantities. In Fig. 7.12 we examine predicted velocity profiles (no velocity data was used for training). In all cases the velocity profiles are significantly improved with respect to the baseline SST model, and the choice of parameterization – TPP, ATP or RMP – makes little difference. We can infer from these results that the pressure distribution provides relatively little direct information about the turbulence, but does say something about the separation- and reattachment locations, and these are the quantities that have been matched by the calibration with appropriate choices of $\boldsymbol{\theta}$. As a consequence velocity profiles are improved by virtue of having the correct separation bubble extent.

Ultimately this exercise has demonstrated the importance of having access to measurement data relevant to (or in statistical terms: correlated with) the quantities that one wishes to estimate. Estimating turbulence characteristics purely from pressure data is probably not feasible, no matter what the model parameterization framework; however deriving a better model (on the sense of better predicting mean-flow quantities) may be more feasible.

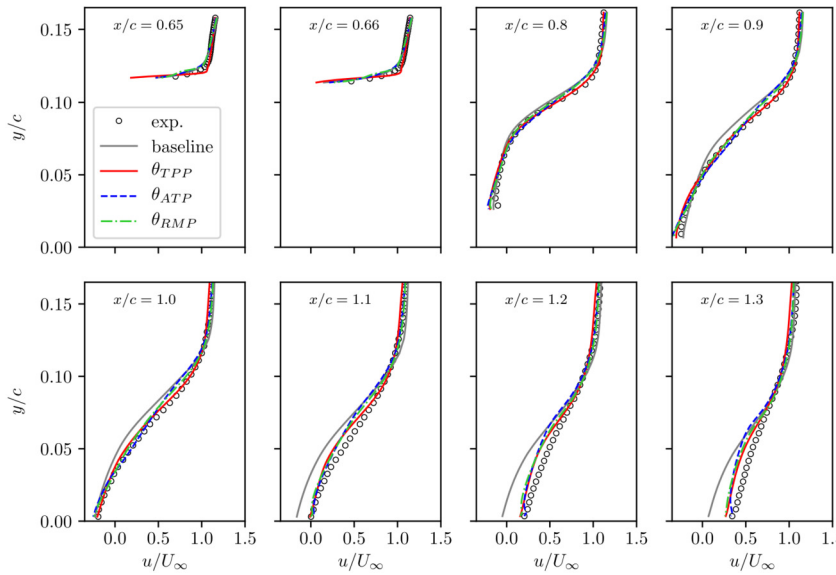


FIGURE 7.12 Velocity profiles at various x-locations for the corrected models compared with experimental data.

7.5 Conclusions

In this chapter we have described the Bayesian framework as it relates to calibration- and inverse problems in the context of CSE codes, and RANS turbulence modeling in particular. In our view the major value of the framework is its ability to precisely characterize relationships between arbitrary inputs, outputs, and state-variables, of arbitrarily complex computational models. We have seen for example how the framework immediately generalizes to the case of multiple models (in Bayesian Model Selection/Averaging in Section 7.3.3), and it generalizes similarly effortlessly to time-dependent problems. We've warned about the risks of model inadequacy, how they can lead to fallacious conclusions, and illustrated how these risks can be mitigated using statistical modeling (Section 7.2.3).

The Bayesian framework's power comes from representing relationships between variables in terms of joint pdfs, but this generality comes at the cost of significant computational cost. The major challenge in applying Bayesian methods is managing this cost – especially for expensive CSE codes and high-dimensional parameter spaces. We've discussed some means of handling this, notably the large class of *variational approaches* (Section 7.2.4.3), but often a statistical simplification is necessary at the cost of discarding some – or all – uncertainty information. Even when this is necessary, we see Bayes as a unusually effective means of reasoning about CSE systems, making otherwise arbitrary modeling choices explicit and motivated.

The particular application to RANS modeling is timely, given the great current interest in data-driven modeling [33,67,42,49,58,24]. We believe the best approach to these questions is necessarily statistical.

Appendix 7.A Probability refresher

We summarize in this section the basic aspects of probability that are specifically relevant for the coming sections. This is intended as a brief refresher only - for thorough treatments we recommend the books by Skilling & Sivia, for basic Bayesian statistics [55]; Gelman et al., for effective approaches to Bayesian inference, from a statistical point-of-view [17]; and Tarantola, for Bayesian methods in the context of inverse problems [60].

The fundamental object we need for probabilistic modeling is the *random variable* (r.v.). To define an r.v. we first need a *probability space*:

Definition 1. (Probability space) A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

- A *sample space* Ω is the set of all possible outcomes.
- An *event space* $\mathcal{F}$, is a set of all subsets of Ω; e.g. defining $\mathcal{F}$ as a σ-algebra on Ω, allows for the definition of a measurable space $(\Omega, \mathcal{F})$.
- A *probability function* $\mathbb{P} : \mathcal{F} \to [0, 1]$ with
  - **(i)** $\mathbb{P}(\Omega) = 1$,
  - **(ii)** $\mathbb{P}(\emptyset) = 0$,
  - **(iii)** If $A, B \in \mathcal{F}$ and $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

This is very general formulation, but in the case of continuous probability spaces, it is convenient to let $\Omega = [0, 1]$ (or any other interval), $\mathcal{F}$ the σ-algebra of Ω, given which

$$\mathbb{P}(A) := \frac{1}{|\Omega|} \int_A \mathrm{d}\omega, \tag{7.32}$$

is well-defined, and satisfies the three criteria for the probability function. Informally, we can imagine we have a random-number generator that gives us a sample $\omega \in \Omega$. By the construction of $\mathcal{F}$ we can assign a probability to any subset of Ω, with probability given by (7.32).

To represent uncertainties in parameters in our models as uncertain, we could – in principle – construct a sample space for each parameter, choosing a weighted version of (7.32) to represent different distributions. But this is unwieldy. Instead we use:

Definition 2. (Random variables) A real-valued random variable is a measurable *function* $X : \Omega \to \mathbb{R}$.

The caveat "measurable" is always satisfied in engineering, the important point is that these are just functions from the sample-space to the uncertain quantity

(dimensional or non-dimensional). The function allows us to take samples (by generating random ω):

Definition 3. (Realization) $x := X(\omega)$ for $\omega \in \Omega$ is a realization of X.

Critically the definition of a r.v. allows function composition. E.g. consider a (deterministic) function $f : \mathbb{R} \to \mathbb{R}$ giving $f(x)$ for $x \in \mathbb{R}$. Replacing x with X gives $f(X) = f \circ X(\omega)$ for $\omega \in \Omega$. Thus $f(X) : \Omega \to \mathbb{R}$ is *also* a r.v. defined on the same probability space. It is on this basis that inputs, parameters and outputs of CSE models can be replaced with r.v.'s in a meaningful way.

The specific function X is hardly used in practice (and Ω is rarely specified), rather r.v.'s are characterized in one of five main ways[10]:

(a) Cumulative density function (cdf): $F_X(x) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) < x\})$,
(b) Probability density function (pdf): $\rho_X(x) := \mathrm{d}F_x/\mathrm{d}x|_x$,
(c) Samples: $X_1 = X(\omega_1), X_2 = X(\omega_2), \ldots, X_M = X(\omega_M)$,
(d) (Centered) moments: $\mu_X := \mathbb{E}_X(x)$, $\sigma_X^2 := \mathbb{E}_X(x - \mu_X)^2$, $m_3 := \mathbb{E}_X(x - \mu_X)^3, \ldots$
(e) Characteristic function: $\phi_X(t) := \mathbb{E}_X(e^{itx})$, i.e. the Fourier transform of the pdf,

where the expectation is defined as:

Definition 4. (Expectation operator) $\mathbb{E}_X[f(x)] := \int_{-\infty}^{\infty} f(x)\rho_X(x)\,\mathrm{d}x$.

The cdf, pdf and characteristic function are exact representations of the r.v., while samples and moments are approximate (for finite M). However some representations are more convenient for certain tasks (as we shall see), and much of the numerical effort involved in working with r.v.'s involves transforming between different representations. The challenge of uncertainty *propagation* (often known as uncertainty quantification (UQ)), is to obtain a representation of the r.v. $f(X)$ given a known, deterministic f, and the r.v. X. Some examples transformations between representation of r.v.'s, and the associated numerical methods are:

- Given samples $X_1, \ldots, X_M$, estimate moments $\mu_X, \ldots$ (Monte-Carlo)
- Given samples $X_1, \ldots, X_M$, estimate the pdf ρ_X (histogram, kernel density estimate).
- Given the pdf ρ_X obtain samples $X_1, \ldots, X_M$ (Gibbs sampling, Markov-chain Monte-Carlo).
- Given the pdf ρ_X, estimate moments of $f(X)$ (polynomial chaos, probabilistic collocation).

Note that, given M samples from X, obtaining samples of $f(X)$ is straightforward, as is then estimating the moments $\mu_{f(X)}$ (via Monte-Carlo estimators) – but this requires M evaluations of f. Polynomial chaos and associated methods aim to reduce this cost.

[10] A random variable can also be characterized as the transformation of another r.v., e.g. $Y = \exp X$ for $X \sim \mathcal{N}(0, 1)$ is a log-normal distribution.

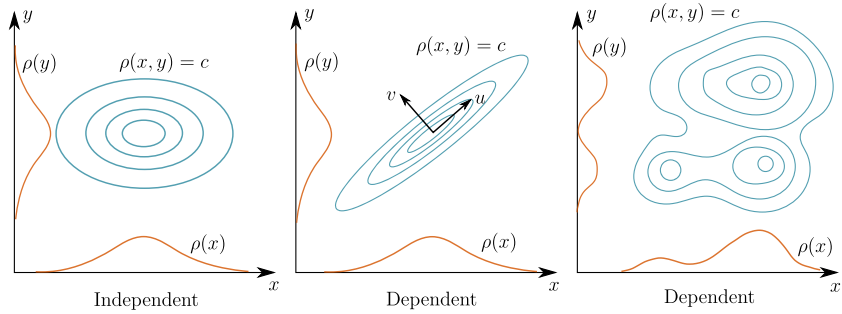


FIGURE 7.13 (Cartoon) Joint-pdfs of (X, Y). Blue lines are contours of the joint densities; orange lines on axes are marginal densities. Left: $X \perp Y$ so $\rho(x, y) = \rho(x)\rho(y)$. Middle: $X \not\perp Y$, larger x implies larger y on average; although in the rotated coordinate system (u, v), $U \perp V$. Right: $X \not\perp Y$ and their relationship is complex.

Multivariate random-variables

Multivariate random-variables have properties additional to the one-dimensional r.v.'s discussed above. They are still just functions from the sample space:

Definition 5. (Random vector) A real-valued random vector is a measurable function $X : \Omega \to \mathbb{R}^N$.

The two-dimensional case is illustrative: consider $\boldsymbol{X} = (X, Y)$, i.e. a random 2-vector. It has a joint cdf defined analogously to the scalar case: $F_{\boldsymbol{X}}(\boldsymbol{x}) := \mathbb{P}(X < x \cap Y < y)$, and joint pdf $\rho_{\boldsymbol{X}}(\boldsymbol{x}) := \partial^2 F_{\boldsymbol{X}} / \partial x \, \partial y |_{\boldsymbol{x}}$, which are both now functions of two variables. Note that $\rho_{\boldsymbol{X}}(x, y)$ refers to the probability that $X = x$ *and* $Y = y$ simultaneously.

However, now there is the possibility that X and Y are either *dependent* or *independent* – see Fig. 7.13. In the case that their joint pdf can be separated into a product of scalar functions:

$$\rho_{\boldsymbol{X}}(\boldsymbol{x}) = \rho_X(x) \cdot \rho_Y(y) \quad \Longleftrightarrow \quad X \perp Y \tag{7.33}$$

X and Y are said to be *independent*, and we write $X \perp Y$. Otherwise they are dependent, $X \not\perp Y$.

Example 2. (Multivariate normal) Let $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} = (\mu_x, \mu_y)$ is the (vector) mean, and $\Sigma \in R^{2\times 2}$ is the positive-definite covariance matrix. If Σ is diagonal, the joint density $\rho(x, y)$ will resemble Fig. 7.13 (left), and $X \perp Y$. If Σ is not diagonal, the joint density $\rho(x, y)$ will resemble Fig. 7.13 (middle). Given the eigen-decomposition $\Sigma = Q \Lambda Q^{-1}$, the transformed variables $(u, v) = \boldsymbol{u} = Q^{-1}\boldsymbol{x}$ are independent.

In both dependent and independent cases we can define *marginal* and *conditional* distributions:

Definition 6. (Marginal distribution) Given jointly distributed (X, Y) with density $\rho_{X,Y}$ the marginal density of X alone is

$$\rho_X(x) := \int_{-\infty}^{\infty} \rho_{X,Y}(x, y)\, \mathrm{d}y,$$

which might be interpreted as is the distribution of x when Y is "ignored" (integrated out). In the case of independence the marginal distribution is exactly the density of X in (7.33). In contrast the conditional distribution tells us something about the value of x given that we know y precisely:

Definition 7. (Conditional distribution) Given (X, Y) with density $\rho_{X,Y}$ the conditional distribution of X given $Y = y$ is defined by the relation

$$\rho_{X|y}(x|Y = y) \cdot \rho_Y(y) = \rho_{X,Y}(x, y) \tag{7.34}$$

where $\rho_Y(y)$ is the marginal density of Y.

Again, in the case of independence $\rho_{X|y}(x|Y = y) = \rho_X(x)$. Note that $\rho_{X|y}(x|Y = y)$ is a probability density in x, but not necessarily in y.

Example 3. (Exact function inverse) Consider a function f taking a single input $x \in \mathbb{R}$, returning a single output $y \in \mathbb{R}$, denoted $y = f(x)$. We assume that there exists some true but unknown $\tilde{x} \in \mathbb{R}$, and we represent our *empirical uncertainty* in this value with the r.v. $X : \Omega \to \mathbb{R}$, with pdf $\rho_X(x)$. Our corresponding empirical uncertainty in the true value $\tilde{y}$ is $Y = f \circ X$, with pdf $\rho_Y(y)$. Since X and Y are related via f, clearly $X \not\perp Y$, and the joint density $\rho_{X,Y}(x, y)$ is not separable. As a consequence learning *additional* information about the output value $\tilde{y}$ gives us a new information about $\tilde{x}$. Specifically if we observe $\tilde{y}$ *exactly*, the conditional r.v. $X \mid Y = \tilde{y}$ represents our updated empirical uncertainty in $\tilde{x}$, with distribution

$$\rho_{X|\tilde{y}}(x|\tilde{y}) = \frac{\rho_{X,Y}(x, \tilde{y})}{\rho_Y(\tilde{y})}.$$

If f is a bijection this distribution will be a single Dirac δ-function centered at $f^{-1}(\tilde{y})$ (implying we now know $\tilde{x}$ exactly), but the distribution is well-defined for non-invertible f as well.[11]

Conditional probability therefore gives a framework for characterizing solutions of inverse problems, that are well-defined in broad circumstances, including for over- and under-determined problems. There are two practical issues with the above construction however: 1) it relies on direct knowledge of the joint-density $\rho(x, y)$, which might not be accessible, and 2) it does not address the case that $\tilde{y}$ is *approximately* observed – which is the default case in measurements of reality. Bayes' theorem resolves both these issues, while maintaining conditional probability as the means of defining solutions. See Section 7.2.1.

[11] Including if f is a function of more than one variable, in which case the distribution will have support over some manifold in the input space.

References

[1] Farzana Anowar, Samira Sadaoui, Bassant Selim, Conceptual and empirical comparison of dimensionality reduction algorithms, Computer Science Review (ISSN 1574-0137) 40 (May 2021) 100378, https://doi.org/10.1016/j.cosrev.2021.100378.

[2] J.H.S. de Baar, T.P. Scholcz, R.P. Dwight, Exploiting adjoint derivatives in high-dimensional metamodels, AIAA Journal (ISSN 1533-385X) 53 (5) (May 2015) 1391–1395, https://doi.org/10.2514/1.J053678.

[3] S. Banerjee, et al., Presentation of anisotropy properties of turbulence, invariants versus eigenvalue approaches, Journal of Turbulence (ISSN 1468-5248) 8 (Jan. 2007) N32, https://doi.org/10.1080/14685240701506896.

[4] Zeno Belligoli, Richard P. Dwight, Georg Eitelberg, Reconstruction of turbulent flows at high Reynolds numbers using data assimilation techniques, AIAA Journal (ISSN 1533-385X) 59 (3) (Mar. 2021) 855–867, https://doi.org/10.2514/1.J059474.

[5] Steve Brooks, et al., Handbook of Markov Chain Monte Carlo, CRC Press, 2011.

[6] D.E. Coles, E.A. Hirst, Computation of turbulent boundary layers, in: Proceedings of AFOSR-IFP Stanford Conference, vol. 2, 1968.

[7] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, The Annals of Mathematical Statistics (ISSN 0003-4851) 38 (2) (Apr. 1967) 325–339, https://doi.org/10.1214/aoms/1177698950.

[8] Richard Dwight, Zhong-Hua Han, Efficient uncertainty quantification using gradient-enhanced Kriging, in: 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, American Institute of Aeronautics and Astronautics, May 2009, https://doi.org/10.2514/6.2009-2276.

[9] Richard P. Dwight, Stijn G.L. Desmedt, Pejman Shoeibi Omrani, Sobol indices for dimension adaptivity in sparse grids, in: Simulation-Driven Modeling and Optimization, Springer International Publishing, 2016, pp. 371–395, https://doi.org/10.1007/978-3-319-27517-8_15.

[10] W.N. Edeling, P. Cinnella, R.P. Dwight, Predictive RANS simulations via Bayesian model-scenario averaging, Journal of Computational Physics (ISSN 0021-9991) 275 (Oct. 2014) 65–91, https://doi.org/10.1016/j.jcp.2014.06.052.

[11] W.N. Edeling, R.P. Dwight, P. Cinnella, Simplex-stochastic collocation method with improved scalability, Journal of Computational Physics (ISSN 0021-9991) 310 (Apr. 2016) 301–328, https://doi.org/10.1016/j.jcp.2015.12.034.

[12] W.N. Edeling, et al., Bayesian estimates of parameter variability in the $k-\varepsilon$ turbulence model, Journal of Computational Physics (ISSN 0021-9991) 258 (Feb. 2014) 73–94, https://doi.org/10.1016/j.jcp.2013.10.027.

[13] Wouter N. Edeling, et al., Bayesian predictions of Reynolds-averaged Navier–Stokes uncertainties using maximum a posteriori estimates, AIAA Journal (ISSN 1533-385X) 56 (5) (May 2018) 2018–2029, https://doi.org/10.2514/1.J056287.

[14] Michael Emory, Johan Larsson, Gianluca Iaccarino, Modeling of structural uncertainties in Reynolds-averaged Navier-Stokes closures, Physics of Fluids (ISSN 1089-7666) 25 (Oct. 2013) 11, https://doi.org/10.1063/1.4824659.

[15] Branden Fitelson, Mind 114 (454) (2005) 394–400, ISSN 00264423, 14602113, http://www.jstor.org/stable/3489113 (visited on 11/13/2023).

[16] Alexander I.J. Forrester, András Sóbester, Andy J. Keane, Engineering Design via Surrogate Modelling: A Practical Guide, Wiley, ISBN 9780470770801, July 2008, https://doi.org/10.1002/9780470770801.

[17] A. Gelman, et al., Bayesian Data Analysis, 2nd ed., CRC Texts in Statistical Science, Chapman & Hall, 2003.

[18] Roger Ghanem, P.D. Spanos, Polynomial chaos in stochastic finite elements, Journal of Applied Mechanics (ISSN 1528-9036) 57 (1) (Mar. 1990) 197–202, https://doi.org/10.1115/1.2888303.

[19] Michael B. Giles, Multilevel Monte Carlo path simulation, Operations Research (ISSN 1526-5463) 56 (3) (June 2008) 607–617, https://doi.org/10.1287/opre.1070.0496.

[20] David Greenblatt, et al., Experimental investigation of separation control part 1: baseline and steady suction, AIAA Journal (ISSN 1533-385X) 44 (12) (Dec. 2006) 2820–2830, https://doi.org/10.2514/1.13817.
[21] S.G. Henderson, B.A. Chiera, R.M. Cooke, Generating "dependent" quasi-random numbers, in: 2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165), WSC-00, IEEE, https://doi.org/10.1109/WSC.2000.899760.
[22] M. Hinze, et al., Optimization with PDE Constraints, Mathematical Modelling: Theory and Applications, vol. 23, Springer, New York, ISBN 978-1-4020-8838-4, 2009, pp. xii+270.
[23] Matthew D. Hoffman, Andrew Gelman, The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo, Journal of Machine Learning Research 15 (2011) 1593–1623, https://api.semanticscholar.org/CorpusID:12948548.
[24] Mikael L.A. Kaandorp, Richard P. Dwight, Data-driven modelling of the Reynolds stress tensor using random forests with invariance, Computers & Fluids (ISSN 0045-7930) 202 (Apr. 2020) 104497, https://doi.org/10.1016/j.compfluid.2020.104497.
[25] Robert Kass, Luke Tierney, Joseph Kadane, Laplace's method in Bayesian analysis, https://doi.org/10.1090/conm/115/07, 1991.
[26] Marc C. Kennedy, Anthony O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society, Series B, Statistical Methodology (ISSN 1467-9868) 63 (3) (Sept. 2001) 425–464, https://doi.org/10.1111/1467-9868.00294.
[27] Diederik P. Kingma, Max Welling, Auto-encoding variational Bayes, arXiv:1312.6114 [stat.ML], 2013.
[28] Prashant Kumar, Martin Schmelzer, Richard P. Dwight, Stochastic turbulence modeling in RANS simulations via multilevel Monte Carlo, Computers & Fluids (ISSN 0045-7930) 201 (Apr. 2020) 104420, https://doi.org/10.1016/j.compfluid.2019.104420.
[29] M. Paul van der Laan, et al., An improved k-ε model applied to a wind turbine wake in atmospheric turbulence, Wind Energy (ISSN 1099-1824) 18 (5) (Apr. 2014) 889–907, https://doi.org/10.1002/we.1736.
[30] B.E. Launder, B.I. Sharma, Application of the energy-dissipation model of turbulence to the calculation of flow near a spinning disc, Letters in Heat and Mass Transfer 1 (1974) 131–137.
[31] C. Lemieux, Monte Carlo and Quasi-Monte Carlo Sampling, Springer Series in Statistics, Springer, New York, ISBN 9780387781655, 2009, https://books.google.nl/books?id=wj5OyydZ5bkC.
[32] Michael Leschziner, Statistical Turbulence Modelling for Fluid Dynamics — Demystified: An Introductory Text for Graduate Engineering Students, Imperial College Press, ISBN 9781783266623, Dec. 2014, https://doi.org/10.1142/p997.
[33] Julia Ling, Andrew Kurzawski, Jeremy Templeton, Reynolds averaged turbulence modelling using deep neural networks with embedded invariance, Journal of Fluid Mechanics (ISSN 1469-7645) 807 (Oct. 2016) 155–166, https://doi.org/10.1017/jfm.2016.615.
[34] Qiang Liu, Dilin Wang, Stein variational gradient descent: a general purpose Bayesian inference algorithm, arXiv:1608.04471 [stat.ML], 2019.
[35] G.J.A. Loeven, J.A.S. Witteveen, H. Bijl, Probabilistic collocation: an efficient non-intrusive approach for arbitrarily distributed parametric uncertainties, in: 45th AIAA Aerospace Sciences Meeting and Exhibit, American Institute of Aeronautics and Astronautics, Jan. 2007, https://doi.org/10.2514/6.2007-317.
[36] F.R. Menter, Two-equation eddy-viscosity turbulence models for engineering applications, AIAA Journal (ISSN 1533-385X) 32 (8) (Aug. 1994) 1598–1605, https://doi.org/10.2514/3.12149.
[37] Mohsen S. Mohamed, John C. Larue, The decay power law in grid-generated turbulence, Journal of Fluid Mechanics (ISSN 1469-7645) 219 (1) (Oct. 1990) 195, https://doi.org/10.1017/S0022112090002919.
[38] Shinichi Nakajima, Kazuho Watanabe, Masashi Sugiyama, Variational Bayesian Learning Theory, Cambridge University Press, ISBN 9781107430761, June 2019, https://doi.org/10.1017/9781139879354.

[39] Jorge Nocedal, Stephen J. Wright, Numerical Optimization, Springer Series in Operations Research and Financial Engineering, 2. ed., Springer, New York, NY, ISBN 978-0-387-30303-1, 2006.

[40] Todd A. Oliver, Robert D. Moser, Bayesian uncertainty quantification applied to RANS turbulence models, Journal of Physics. Conference Series (ISSN 1742-6596) 318 (4) (Dec. 2011) 042032, https://doi.org/10.1088/1742-6596/318/4/042032.

[41] Abril-Pla Oriol, et al., PyMC: a modern and comprehensive probabilistic programming framework in Python, PeerJ Computer Science 9 (2023) e1516, https://doi.org/10.7717/peerj-cs.1516.

[42] Eric J. Parish, Karthik Duraisamy, A paradigm for data-driven predictive modeling using field inversion and machine learning, Journal of Computational Physics (ISSN 0021-9991) 305 (Jan. 2016) 758–774, https://doi.org/10.1016/j.jcp.2015.11.012.

[43] Jacques E.V. Peter, Richard P. Dwight, Numerical sensitivity analysis for aerodynamic optimization: a survey of approaches, Computers & Fluids (ISSN 0045-7930) 39 (3) (Mar. 2010) 373–391, https://doi.org/10.1016/j.compfluid.2009.09.013.

[44] P.D.A. Platteeuw, G.J.A. Loeven, H. Bijl, Uncertainty quantification applied to the k-ε model of turbulence using the probabilistic collocation method, in: 49th AIAA Structures, Structural Dynamics, and Materials Conference, American Institute of Aeronautics and Astronautics, 2008, https://doi.org/10.2514/6.2008-2150.

[45] Stephen B. Pope, Turbulent Flows, Cambridge University Press, Aug. 2000, ISBN 978051184053, https://doi.org/10.1017/CBO9780511840531.

[46] M.J.D. Powell, On search directions for minimization algorithms, Mathematical Programming (ISSN 1436-4646) 4 (1) (Dec. 1973) 193–201, https://doi.org/10.1007/bf01584660.

[47] Christian P. Robert, George Casella, Monte Carlo Statistical Methods, Springer, New York, ISBN 9781475741452, 2004, https://doi.org/10.1007/978-1-4757-4145-2.

[48] Philipp Schlatter, Ramis Örlü, Assessment of direct numerical simulation data of turbulent boundary layers, Journal of Fluid Mechanics 659 (2010) 116–126, https://api.semanticscholar.org/CorpusID:122723804.

[49] Martin Schmelzer, Richard P. Dwight, Paola Cinnella, Discovery of algebraic Reynolds-stress models using sparse symbolic regression, Flow, Turbulence and Combustion (ISSN 1573-1987) 104 (2–3) (Dec. 2019) 579–603, https://doi.org/10.1007/s10494-019-00089-x.

[50] A. Sciacchitano, Uncertainty quantification in particle image velocimetry, Measurement Science & Technology (ISSN 1361-6501) 30 (9) (July 2019) 092001, https://doi.org/10.1088/1361-6501/ab1db8.

[51] Andrea Sciacchitano, Bernhard Wieneke, PIV uncertainty propagation, Measurement Science & Technology (ISSN 1361-6501) 27 (8) (June 2016) 084006, https://doi.org/10.1088/0957-0233/27/8/084006.

[52] David W. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, Wiley, ISBN 9781118575574, Mar. 2015, https://doi.org/10.1002/9781118575574.

[53] Glenn Shafer, A mathematical theory of evidence, https://doi.org/10.2307/j.ctv10vm1qb, June 2020.

[54] Anand Pratap Singh, Karthik Duraisamy, Using field inversion to quantify functional errors in turbulence closures, Physics of Fluids (ISSN 1089-7666) 28 (Apr. 2016) 4, https://doi.org/10.1063/1.4947045.

[55] D.S. Sivia, J. Skilling, Data Analysis - A Bayesian Tutorial, 2nd, Oxford Science Publications, Oxford University Press, 2006.

[56] C. Soize, Random matrix theory for modeling uncertainties in computational mechanics, Computer Methods in Applied Mechanics and Engineering (ISSN 0045-7825) 194 (12–16) (Apr. 2005) 1333–1366, https://doi.org/10.1016/j.cma.2004.06.038.

[57] P. Spalart, S. Allmaras, A one-equation turbulence model for aerodynamic flows, in: 30th Aerospace Sciences Meeting and Exhibit, American Institute of Aeronautics and Astronautics, Jan. 1992, https://doi.org/10.2514/6.1992-439.

[58] Julia Steiner, Axelle Viré, Richard P. Dwight, Classifying regions of high model error within a data-driven RANS closure: application to wind turbine wakes, Flow, Turbulence and Combustion (ISSN 1573-1987) 109 (3) (Aug. 2022) 545–570, https://doi.org/10.1007/s10494-022-00346-6.

[59] Terence Tao, Compactness and contradiction, Miscellaneous Books, Providence, R.I.; American Mathematical Society, Toronto; Fields Institute for Research in Mathematical Sciences https://doi.org/10.1090/mbk/081, 2013.

[60] Albert Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics, 2004.

[61] Roney L. Thompson, et al., A methodology to evaluate statistical errors in DNS data of plane channel flows, Computers & Fluids (ISSN 0045-7930) 130 (May 2016) 1–7, https://doi.org/10.1016/j.compfluid.2016.01.014.

[62] D.C. Wilcox, Turbulence Modelling for CFD, DCW Industries, La Cañada, 1993.

[63] D.C. Wilcox, Comparison of two-equation turbulence models for boundary layers with pressure gradient, AIAA Journal 31 (8) (1993).

[64] D.C. Wilcox, American Institute of Aeronautics, and Astronautics, Turbulence Modeling for CFD, vol. 3, DCW Industries La Canada, CA, 2006.

[65] David C. Wilcox, Formulation of the k-w turbulence model revisited, AIAA Journal (ISSN 1533-385X) 46 (11) (Nov. 2008) 2823–2838, https://doi.org/10.2514/1.36541.

[66] Jeroen A.S. Witteveen, Alex Loeven, Hester Bijl, An adaptive stochastic finite elements approach based on Newton–Cotes quadrature in simplex elements, Computers & Fluids (ISSN 0045-7930) 38 (6) (June 2009) 1270–1288, https://doi.org/10.1016/j.compfluid.2008.12.002.

[67] Jin-Long Wu, Heng Xiao, Eric Paterson, Physics-informed machine learning approach for augmenting turbulence models: a comprehensive framework, Physical Review Fluids (ISSN 2469-990X) 3 (July 2018) 7, https://doi.org/10.1103/PhysRevFluids.3.074602.

[68] Heng Xiao, Jian-Xun Wang, Roger G. Ghanem, A random matrix approach for quantifying model-form uncertainties in turbulence modeling, Computer Methods in Applied Mechanics and Engineering (ISSN 0045-7825) 313 (Jan. 2017) 941–965, https://doi.org/10.1016/j.cma.2016.10.025.

[69] Huidan Yu, Sharath S. Girimaji, Li-Shi Luo, DNS and LES of decaying isotropic turbulence with and without frame rotation using lattice Boltzmann method, Journal of Computational Physics (ISSN 0021-9991) 209 (2) (Nov. 2005) 599–616, https://doi.org/10.1016/j.jcp.2005.03.022.

[70] Lotfi A. Zadeh, George J. Klir, Bo Yuan, Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems, Advances in Fuzzy Systems — Applications and Theory, ISSN 2010-2771, May 1996, https://doi.org/10.1142/2895.