On the use of Bayesian inference for the quantification of turbulence modelling uncertainties

Paola CINNELLA
Università del Salento and Laboratoire DynFluid (EA 92), Paris
paola.cinnella@ensam.eu

Joint work with : W. Edeling, R. Dwight

Forum incertitudes du CEA DAM, October the 1st, 2014

Talk overview

- Introductory thoughts
- Parametric and model-form uncertainties in Reynolds-Averaged Navier-Stokes (RANS) solutions of turbulent flows
- Bayesian calibration of turbulence models
- Quantifying model-form uncertainty
  - Bayesian model-scenario averaging (BMSA)
- Conclusions and perspectives

Talk overview

- Introductory thoughts
- Parametric and model-form uncertainties in Reynolds-Averaged Navier-Stokes (RANS) solutions of turbulent flows
- Bayesian calibration of turbulence models
- Quantifying model-form uncertainty
  - Bayesian model-scenario averaging (BMSA)
- Conclusions and perspectives

Errors and uncertainties

Computer models of physical systems affected by both errors and uncertainties

- Numerical approximation errors, solution errors, round-off errors → can be improved
- Model definition uncertainties (geometry, operating conditions)

  → cannot be improved
- Errors/uncertainties specific to the physical/mathematical model
  - E.G.: fluid properties (density, viscosity, compressibility,...)
  - Submodels describing material behavior (E.G.: EOS, turbulence models, viscosity, ...)

→ Could be improved, but…

Aleatoric and Epistemic uncertainties

- **Uncertainties on geometrical and operating conditions and model tuning parameters are essentially irreductible**
  → aleatoric uncertainties

- **Physical/mathematical models: error or uncertainty?**
  - **Modeling errors : conscious use of a possibly unsuitable/partially suitable model for a given problem**
    - **e.g. use of an inviscid or incompressible flow model, use of turbulence models, use of the ideal polytropic gas model**
  - **Modelling uncertainties : does a model fit a given problem? How close it is to reality? → lack of knowledge that could be improved**
    → epistemic uncertainty

Epistemic uncertainties

- Choice of the appropriate modelling level : essentially "expert judgement"

- For a given level

  - Several possible models, which differ by
    - Their mathematical structure
    - The associated closure parameters

- Common practice

  - Model structure chosen by **expert judgement**
  - Model constants **imperfectly known/adjustable**

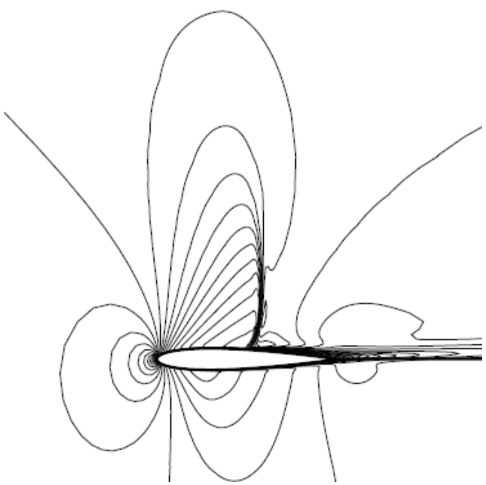
→ Sources of uncertainty

Talk overview

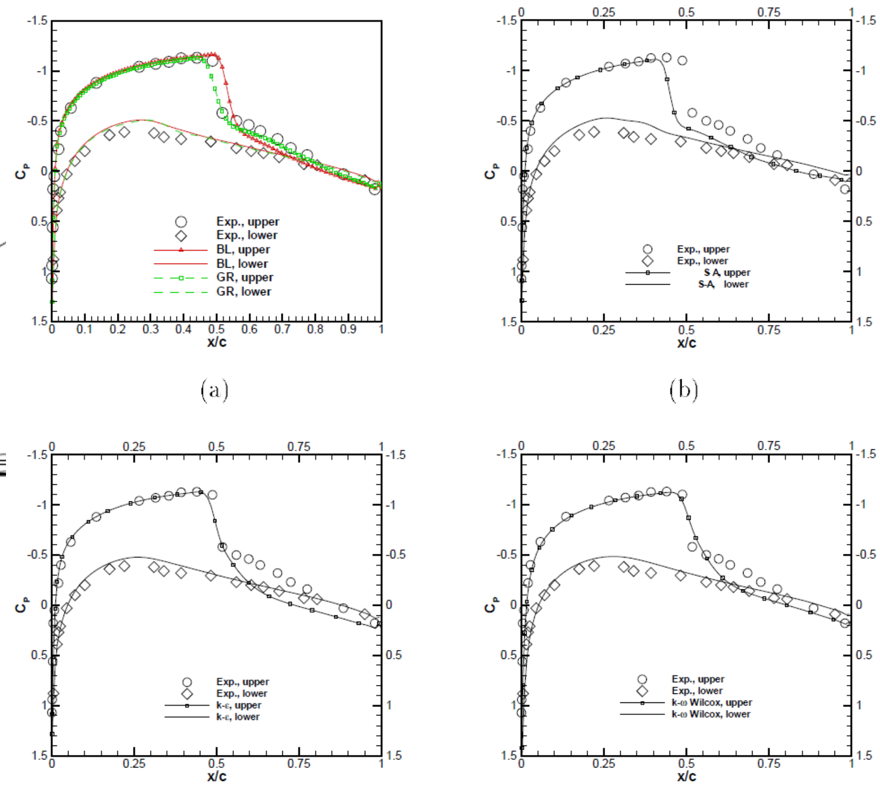
- Introductory thoughts
- **Parametric and model-form uncertainties in Reynolds-Averaged Navier-Stokes (RANS) solutions of turbulent flows**
- Bayesian calibration of turbulence models
- Quantifying model-form uncertainty
  - Bayesian model-scenario averaging (BMSA)
- Conclusions and perspectives

Motivating example

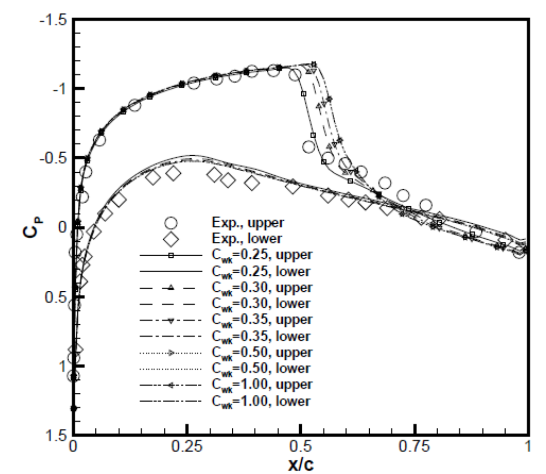
- Turbulent flow past an airfoil, M=0.8, AoA=2°, Re=9e6
- Perfect gas, Newtonian fluid, <u>eddy viscosity turbulence model</u>



IsoMach lines



Wall pressure coefficient, different models



Wall pressure coefficient, BL model, different values of a closure coefficient

Navier-Stokes equations

- Incompressible Navier-Stokes equations:

$$\nabla \cdot \mathbf{u} = 0$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho}\nabla p + \frac{1}{\rho}\nabla \left(\mu \nabla \mathbf{u}\right)$$

- Reynolds averaged:

$$\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$$

$$\nabla \cdot \mathbf{u} = 0$$

$$\frac{\partial \bar{\mathbf{u}}}{\partial t} + \bar{\mathbf{u}} \cdot \nabla \bar{\mathbf{u}} = -\frac{1}{\rho}\nabla \bar{p} + \frac{1}{\rho}\nabla \left(\mu \nabla \bar{\mathbf{u}} \underbrace{- \rho \overline{\mathbf{u}'\mathbf{u}'}}_{\text{Reynolds stress tensor}}\right)$$

Reynolds stresses need additional modelling: a **turbulence model**

Closure models for the Reynolds stresses

- Rich zoology of models proposed through the years:
  - Algebraic: Prandtl's mixing length (1921), Cebeci-Smith (1974), Baldwin-Lomax (1978), …
  - Half-equation: Johnson and King (1984)
  - One equation: Bradshaw (k, 1967), Baldwin-Barth (1990), Spalart-Allmaras (1992), …
  - Two equations: k-ε (several), k-ω (several), k-l, k-t, …
  - Two equations non linear: ….
  - Three-equations …
  - Seven equations (Reynolds-Stress Models): …
  - …

→ No universally accepted and valid model yet

Closure models for the Reynolds stresses

- **Example: *k*-ε models**
  - k-eps models:

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

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

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

  - Launder-Sharma:

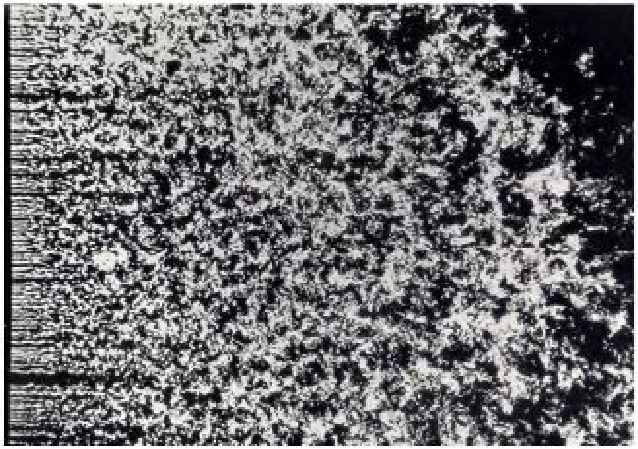
$$C_\mu = 0.09, \quad C_{\varepsilon 1} = 1.44, \quad C_{\varepsilon 2} = 1.92, \quad \sigma_k = 1.0, \quad \sigma_\varepsilon = 1.3.$$

=> θ

  - Jones-Launder:

$$C_\mu = 0.09, \quad C_{\varepsilon 1} = 1.55, \quad C_{\varepsilon 2} = 2.0, \quad \sigma_k = 1.0, \quad \sigma_\varepsilon = 1.3.$$

Model coefficients are not sacred!!



- E.g. Isotropic decaying turbulence
- Equations reduce to

$$\frac{dk}{dt} = -\varepsilon,$$
$$\frac{d\varepsilon}{dt} = -C_{\varepsilon 2}\frac{\varepsilon^2}{k}.$$

- With exact solution

$$k(t) = k_0 \left(\frac{t}{t_0}\right)^{-n} \qquad n = 1/(C_{\varepsilon 2} - 1)$$

- Values for $C_{\varepsilon 2}$ vary a lot:
  - Commonly used **1.92**
  - RNG k-eps **1.68**
  - k-tau **1.83**
  - Best fit to data (n=1.3) **1.77**

Summary of modelling uncertainties in RANS

- Parametric uncertainty
  - Imperfect knowledge of model parameters
    - Use of uncertain experimental data
    - Uncertain choice of calibration scenario

- Model-form uncertainty
  - Imperfect knowledge of model mathematical structure

Talk overview

- Introductory thoughts
- Parametric and model-form uncertainties in Reynolds-Averaged Navier-Stokes (RANS) solutions of turbulent flows
- **Bayesian calibration of turbulence models**
- Quantifying model-form uncertainty
  - Bayesian model-scenario averaging (BMSA)
- Conclusions and perspectives

Bayesian inference

- **Bayesian inference** is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations
  - Represents uncertainty as a probability distribution
  - Uses a set of observational data to infer a PDF of the closure coefficients → estimate + measure of confidence in estimate
  - All uncertainties are treated in terms of probabilities, including model-form uncertainties

Bayesian inference for model calibration: problem statement

- **Problem data**
  - A model $y = M(x, \theta)$
    with $\boldsymbol{\theta}$ the unknown model random inputs and x the explanatory (known) variables
  - An *a priori* probability distribution for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$
  - A sample of observations for y
- **Problem outcome**
  - The *a posteriori* probability distribution for $\boldsymbol{\theta}$
    → results from our *a priori* knowledge on $\boldsymbol{\theta}$, ***plus***
    the observation *likelihood*
  - An estimate of the model/measurement error variance

Remark

→ standard calibration by, e.g. least mean square regression, only provides the « best fit » value for θ, θ^*, *no error estimate*

Bayesian calibration of model parameters

Model calibration results from Bayes theorem on conditional probability:

$$p(\theta|\mathbf{z}) = \frac{p(\mathbf{z}|\theta)p(\theta)}{p(\mathbf{z})} \qquad (1)$$

where $p(\theta)$ is the prior belief about θ and $p(\mathbf{z}|\theta)$ is the likelihood function; $p(\mathbf{z})$ can be treated as a normalization constant

→ Equation (1) is a statistical **calibration** : it infers the posterior pdf of the parameters that fits the model to the observations y.

For most engineering problems, estimations of z often require running a computer code!!

→ The posterior has to be computed numerically

Parameter vector and *a priori* distribution

In Bayesian calibration the parameter vector is the parameter vector from the physical model θ ***PLUS*** parameters associated to the statistical model, e.g. σ:

- The *a priori* distribution on is the same we use for the initial propagation problem

- In practice, when we know almost nothing about the parameter vector, we choose a large, non informative, distribution

→ typically, this will be a wide uniform distribution

It may be subject to constraints to avoid including unphysical values

Likelihood function

- The likelihood function models the dispersion of the observed data around the model output

→Often taken as a multivariate Gaussian function.

For our model problem:

$$p(z\,|\,\theta) \sim N\left(M\left(x, \boldsymbol{\theta}\right) - \mathbf{z}, \Sigma\right)$$

with Σ a covariance matrix that may involve an additional vector of parameters, called hyperparameters, σ

- For a given model and set of observations, *L* depends only on the unknown parameters

A posteriori distribution

Bayes' theorem provides the joint posterior for θ, σ:

$$p(\theta,\sigma\,|\,\mathbf{z}) = \frac{1}{c}\,p(\theta)\,p(\sigma)\,p(\mathbf{z}\,|\,\theta,\sigma)$$

- **In general this distribution does not belong to a known family of PDF**

→ **It has to be approximated numerically**

- **Markov-Chain Monte Carlo (MCMC) sampling is used to obtain a representative sample of the posterior**
- **The sample is used to represent marginal posteriors for each parameter**
- **Hereafter, we use the Python package pymc**

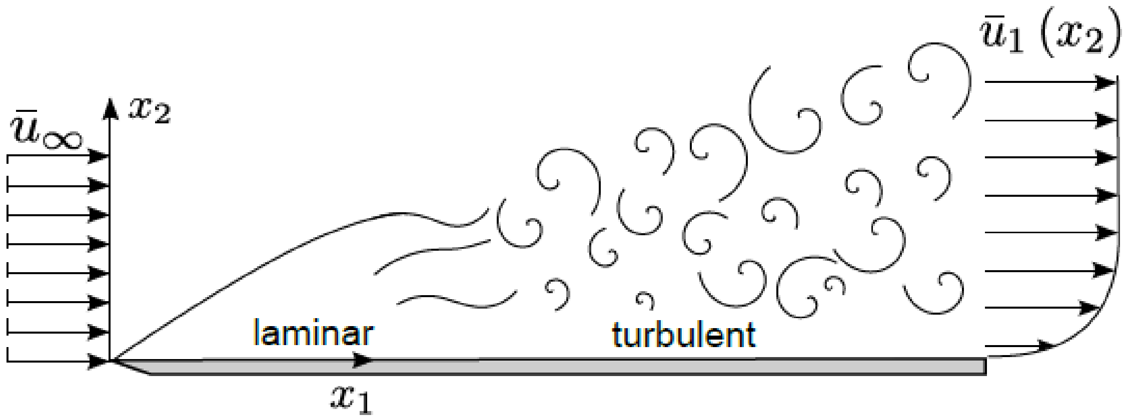
https://pymcmc.readthedocs.org/en/latest/#

Talk overview

- Introductory thoughts
- Parametric and model-form uncertainties in Reynolds-Averaged Navier-Stokes (RANS) solutions of turbulent flows
- **Bayesian calibration of turbulence models**
- Quantifying model-form uncertainty
  - Bayesian model-scenario averaging (BMSA)
- Conclusions and perspectives

Incompressible turbulent flow over a flat plate: model parameter calibration

Objective: predict velocity profiles developing in the turbulent boundary layer close to the wall



- **Governing equations**: Reynolds-Averaged Navier-Stokes equations supplemented by a turbulence model
  - Algebraic Baldwin-Lomax' (1972) model
  - Launder-Jones's (1972) k-ε model
  - Menter's (1992) k-ω SST model
  - Spalart-Allmaras (1992) one-equation model
  - Wilcox' stress-ω model (2006)

Calibration framework

Calibrate from available experimental velocity profiles for several values of the pressure gradient (explanatory variable)

The statistical model incorporates :

- The code output as a function of the uncertain closure coefficients : $\mathbf{y}(\theta_p)$
- A statistical model to represent the structural uncertainty : $\eta \sim N(1, K_{mc})$

$$K_{mc} = \sigma^2 u(\mathbf{x}) u(\mathbf{x}') \exp\left[-\left(\frac{\mathbf{x}-\mathbf{x}'}{\alpha l}\right)^2\right]$$

where σ, α are additional uncertain (hyper) parameters, to be calibrated along with closure coefficients θ_p

- A statistical model to represent the experimental error ***e***

$$z_i = \eta_i y_i(\theta_i) + e_i$$

$e_i \sim N(0, \lambda)$ λ=2% of the mean experimental velocity

Model parameters: priors (1)

Example of the *k*-ε model:

"Standard" coefficients

$$C_{\varepsilon 1} = 1.55, \quad C_{\varepsilon 2} = 2.0, \quad C_{\mu} = 0.09, \quad \sigma_k = 1.0, \quad \sigma_{\varepsilon} = 1.3$$

Not all the coefficients are independent

- $C_{\varepsilon 2}$ is related to the rate of decay of homogeneous isotropic turbulence, *n*

$$C_{2,\epsilon} = \frac{n+1}{n}$$

- $C_{\varepsilon 1}$ is related to $C_{\varepsilon 2}$ via the production-to-dissipation rate (about 1.9)

$$C_{1,\epsilon} = \frac{n+1}{n(\frac{P}{\epsilon})^{k-\epsilon}} + \frac{(\frac{P}{\epsilon})^{k-\epsilon} - 1}{(\frac{P}{\epsilon})^{k-\epsilon}}$$

- The other parameters are usually assumed to satisfy:

$$\sigma_{\epsilon} = \frac{\kappa^2}{C_{\mu}^{1/2}(C_{2,\epsilon} - C_{1,\epsilon})}$$

with κ the von Karman "constant"

Model parameters: priors (2)

According to the available experimental measurements and to the different values proposed in the litterature, we assume the following uniform priors

coefficient	left boundary	right boundary
$C_{\epsilon 2}$	1.152 (-40%)	2.88 (+50%)
C_μ	0.054 (- 40 %)	0.135 (+50 %)
σ_k	0.450 (-45 %)	1.15 (+50 %)
κ	0.287 (-30 %)	0.615 (+50 %)
σ	0.0	0.1
$\log \alpha$	0.0	4.0

Calibration

- Calibration based on experimental data (velocity measures) for 13 boundary layers subject to both positive and negative pressure gradients

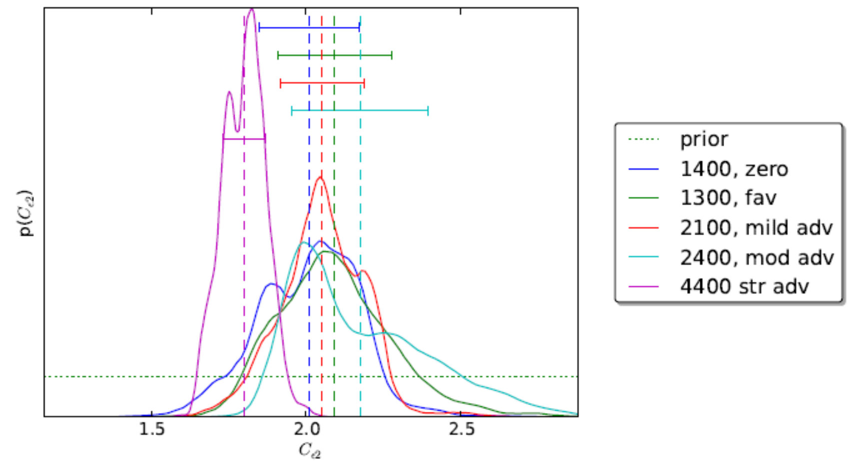
- Numerical solutions obtained through a fast boundary-layer code, more complex flow topologies will require the use of a surrogate model.
- Use Markov-Chain Monte-Carlo method to draw samples from

$$p(\theta_k \mid \mathbf{z}_k)$$

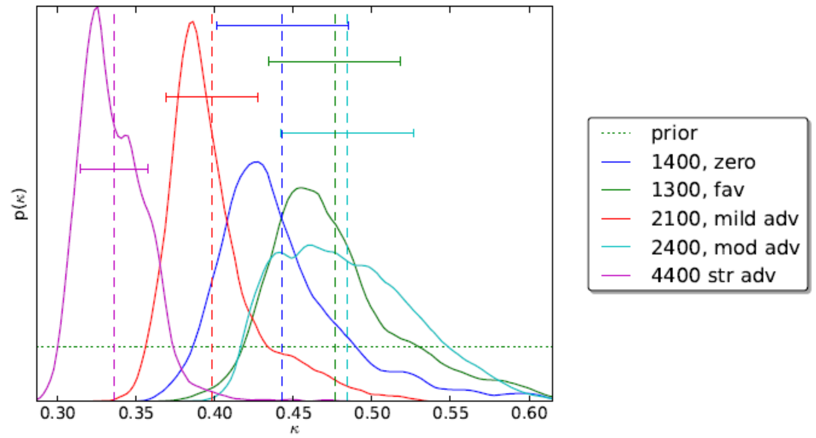
- Use these samples of θ to construct approximate pdfs through a kernel-density estimation.

Some results

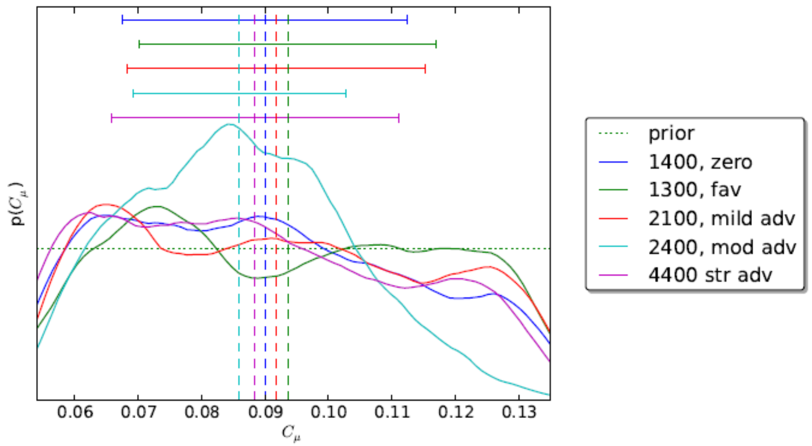
Posterior distributions for $C_{\varepsilon 2}$ for a favorable, zero, mild, moderate and strongly adverse $d\bar{p}/dx$.



Posterior distributions for κ for a favorable, zero, mild, moderate and strongly adverse $d\bar{p}/dx$.

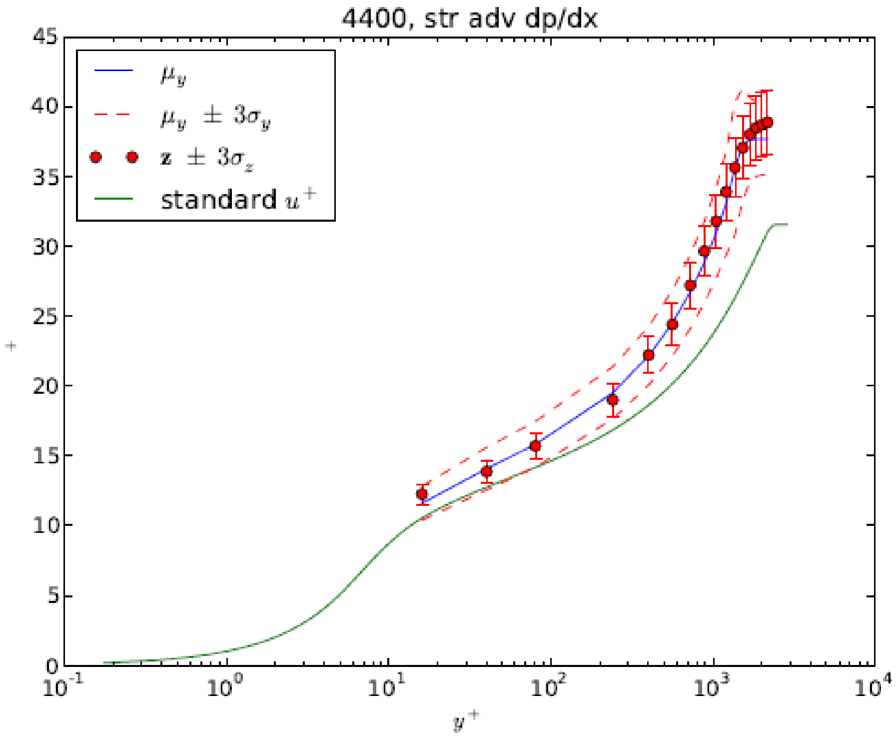


Posterior distributions for C_μ for a favorable, zero, mild, moderate and strongly adverse $d\bar{p}/dx$.

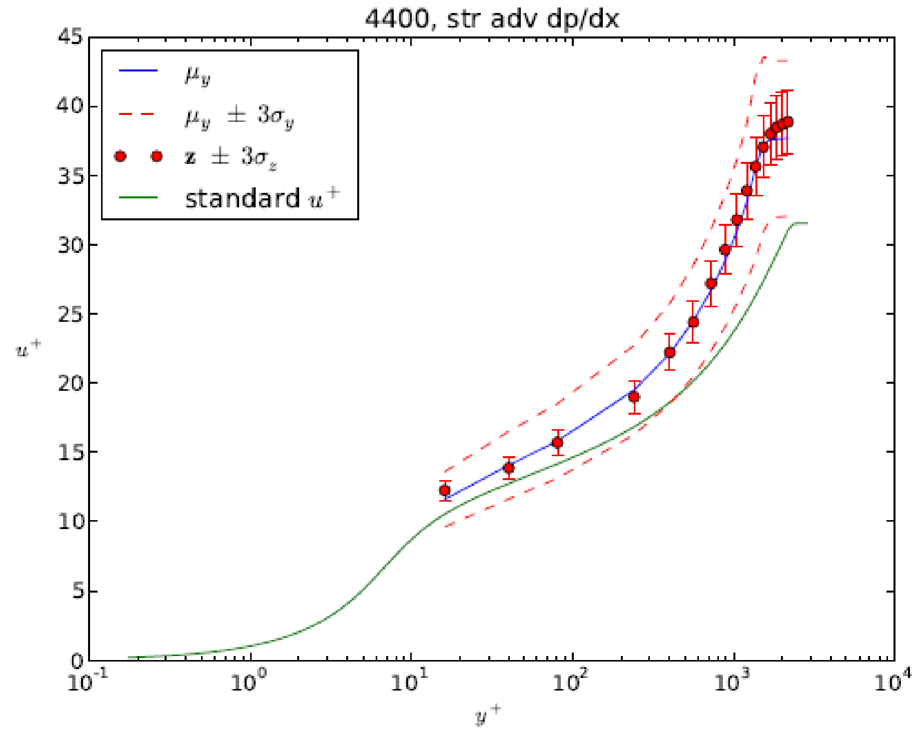


Some results

- **Posteriors are *propagated* through the RANS code to get the posterior estimate of the velocity profile**
- **They can also be used to sample from the true process**



Posterior distribution of y



Posterior distribution of ηy

Remarks

• The spread in most-likely closure coefficients due to different pressure gradients is significant, thus there is no such thing as a true value for the closure coefficients.

• Performing a calibration with a structural uncertainty term tells you something about the structural uncertainty of that case alone.

• How to summarize the effect of both parametric and model-form uncertainty to make predictions of new cases?

Talk overview

- Introductory thoughts
- Parametric and model-form uncertainties in Reynolds-Averaged Navier-Stokes (RANS) solutions of turbulent flows
- Bayesian calibration of turbulence models
- **Quantifying model-form uncertainty**
  - **Bayesian model-scenario averaging (BMSA)**
- Conclusions and perspectives

Model-form uncertainty

Draper (1997)

The model M is not univocally determined because of both parameter and structural uncertainty

Call $\mathcal{M}$ the space of all possible models, M a precise model : $M = (S, \theta)$

→ M composed by two parts: the structure S and the model parameters θ

$$p(y|x,\mathcal{M}) = \int_{\mathcal{M}} p(y|x,M)\, p(M|y)dM = \int\int p(y|x,\theta,S)\, p(\theta,S|y)d\theta dS$$

Weighted average of the posterior distributions using posterior model probabilities (model plausibilities)

Computation of model probabilities? → Again Bayes theorem

$$p(M|y) = p(S|y)\, p(\theta|y,S) = C\; \underline{p(S)}\, \underline{p(\theta|S)}\, \underline{p(y|\theta,S)}$$

Prior probability of S — Prior probability of parameters — Evidence

Remark : $\mathcal{M}$ infinite !! → use of a discrete set of models

Model plausibility

Posterior plausibility quantifies the relative probability that a particular model in a given set produces the observed data

- For a model M in $\mathcal{M}$ the posterior plausibility is:

$$p(M \mid y, \mathcal{M}) = C \;\; p(M \mid \mathcal{M}) E(M; y, \mathcal{M})$$

$$p(M \mid \mathcal{M}) = \text{prior model plausibility}$$

$$E(M; y, \mathcal{M}) = \text{model Evidence:=} p(y \mid M, \mathcal{M}) = \int p(\theta) p(y \mid \theta, M) d\theta$$

- If the prior is a flat non informative function, E is a measure of the "peakedness" of the posterior

Bayesian model averaging (BMA)

- Obtained by applying Draper's decomposition to a discrete set of models
  - Let M_i be a turbulence model in set $\mathcal{M}$, S_k a pressure gradient scenario in set $\mathcal{S}$, and $\mathcal{Z}$ the set of all calibration data
  - The BMA prediction of the expectancy a quantity of interest Δ is then:

$$\mathrm{E}\left(\Delta \mid \mathcal{Z}\right) = \sum_{i=1}^{I} \sum_{k=1}^{K} \mathrm{E}\left(\Delta \mid M_i, S_k, \mathbf{z}_k\right) \mathrm{pr}\left(M_i \mid S_k, \mathbf{z}_k\right) \mathrm{pr}\left(S_k\right)$$

→ Introduction of an additional decomposition over different calibration scenarios (BMSA)

→ The scenario of Δ is not in the calibration set S

- Each individual expectation is weighted by:
  - The posterior model plausibility $\mathrm{pr}\left(M_i \mid S_k, \mathbf{z}_k\right)$
  - The prior scenario probability $\mathrm{pr}\left(S_k\right)$

Bayesian model averaging (BMA)

- Similarly, the variance of Δ writes:

$$\mathrm{var}[\Delta \mid \mathcal{Z}] = \underbrace{\sum_{i=1}^{I}\sum_{k=1}^{K} \mathrm{var}[\Delta \mid M_i, S_k, \mathbf{z}_k]\,\mathrm{pr}(M_i \mid S_k, \mathbf{z}_k)\,\mathrm{pr}(S_k)}_{\text{In-model, in-scenario variance}} +$$

In-model, in-scenario variance

$$\underbrace{\sum_{i=1}^{I}\sum_{k=1}^{K} \left(\mathrm{E}[\Delta \mid M_i, S_k, \mathbf{z}_k] - \mathrm{E}[\Delta \mid S_k, \mathbf{z}_k]\right)^2 \mathrm{pr}(M_i \mid S_k, \mathbf{z}_k)\,\mathrm{pr}(S_k)}_{\text{Between-model, in-scenario variance (model error)}} +$$

Between-model, in-scenario variance (model error)

$$\underbrace{\sum_{k=1}^{K} \left(\mathrm{E}[\Delta \mid S_k, \mathbf{z}_k] - \mathrm{E}[\Delta \mid \mathbf{z}_k]\right)^2 \mathrm{pr}(S_k)}_{\text{Between-scenario variance (spread)}}$$

Between-scenario variance (spread)

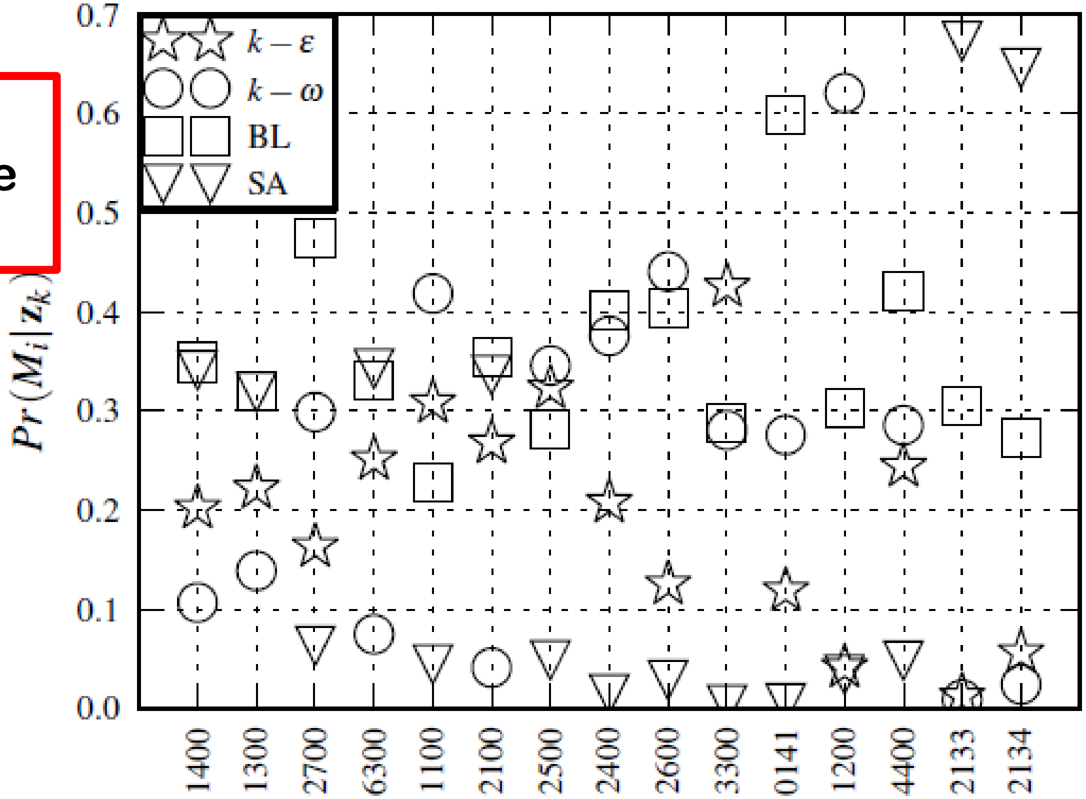
Application of BMSA to the incompressible turbulent flow over a flat plate

- Posterior model plausibilities computed for all models in $\mathcal{M}$ for each S_k using samples from

$$p\left(\theta_k \mid \mathbf{z}_k\right)$$

- Can be considered as a measure of consistency of calibrated model M_i with data $\mathbf{z}_k$

Large spread in model plausibilities, according to the pressure gradient scenario

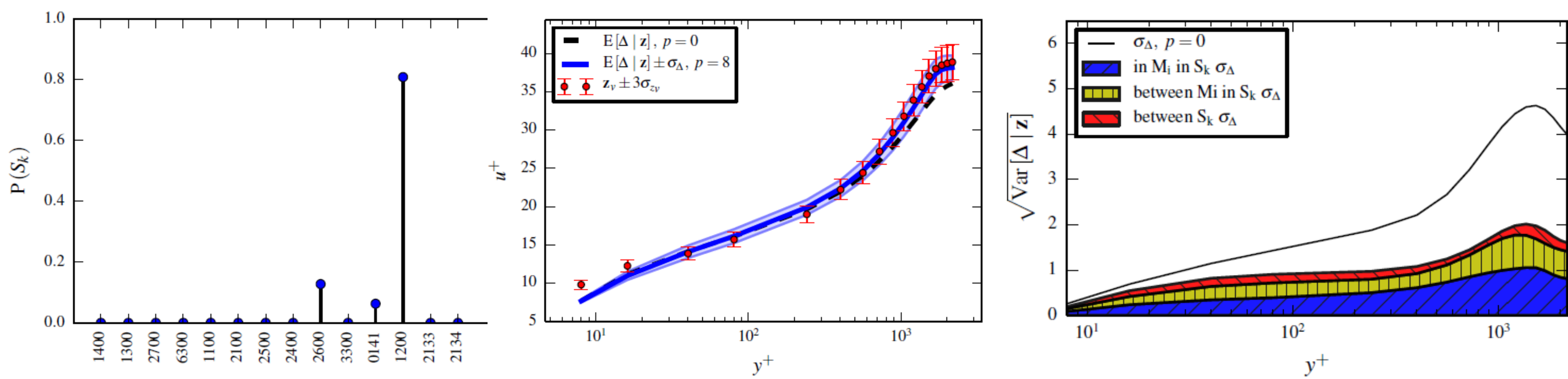


BMSA prediction

- BMA prediction for a validation case (not in the calibration set)
  - Strong adverse pressure gradient
- Scenario pmf weighted according to an error measure

  → penalizes scenarios with a large between-model, in-scenario variance

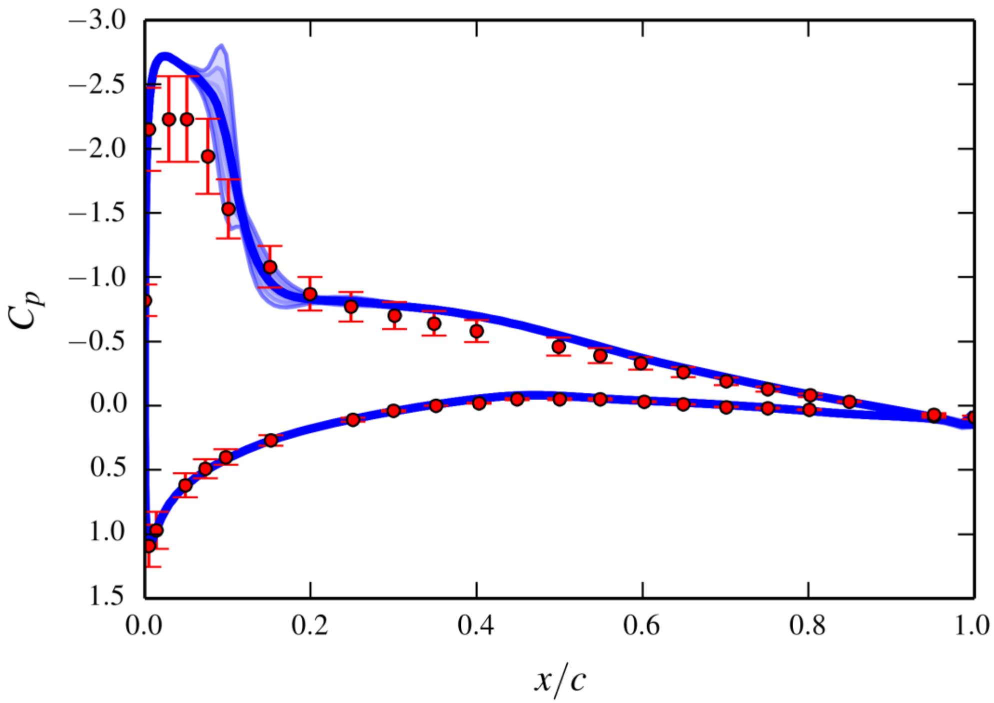
$$\mathcal{E}_k = \frac{\sum_{i=1}^{I} \| \mathrm{E}[\Delta \mid M_i, S_k, \mathbf{z}_k] - \mathrm{E}[\Delta \mid S_k, \mathbf{z}_k] \|_2 \, \mathrm{pr}(M_i \mid S_k, \mathbf{z}_k)}{\| \mathrm{E}[\Delta \mid S_k, \mathbf{z}_k] \|_2} \rightarrow \mathrm{pr}(S_k) = \frac{\mathcal{E}_k^{-p}}{\sum_{k=1}^{K} \mathcal{E}_k^{-p}}, \quad k = 1, 2, \cdots, K$$



- Good prediction, variance consistent with experimental uncertainty
- Significant contribution of the between-scenario variance

BMSA applied to the prediction of a more complex flow

- Coefficient and plausibility database applied to the prediction of the pressure coefficient for high-subsonic flow past an NACA64A010 airfoil (M=0.61, AoA=6.2°)
- Preliminary results with one model, several scenarios
- Propagation through an improved Simplex Stochastic Collocation method (Edeling, Dwight, Cinnella, 2014 – submitted)



- Parametric uncertainty mostly relevant in the high-gradient region

Talk overview

- Introductory thoughts
- Parametric and model-form uncertainties in Reynolds-Averaged Navier-Stokes (RANS) solutions of turbulent flows
- Bayesian calibration of turbulence models
- Quantifying model-form uncertainty
  - Bayesian model-scenario averaging (BMSA)
- **Conclusions and perspectives**

Conclusions

- Bayesian inference allows updating engineering models as soon as some observations become available
  - Particularly suitable for problems such that only a few data are available
  - Calibration allows to unfold parameters not informed by the data and correlated parameters
  - It not only provides optimal values for the parameters, but also error estimates (e.g. coefficients of variation)
- It offers criteria for model selection
- Bayesian model/scenario averaging can be used to summarize predictions made from alternative mathematical models/calibration cases

Thank you