Modèles de turbulence pour la simulation des écoulements

8. Apprentissage automatique de modèles et quantification des incertitudes

Paola CINNELLA

paola.cinnella@sorbonne-universite.fr

SCIENCES SORBONNE UNIVERSITÉ

∂'Alembert
Institut Jean le Rond d'Alembert

Overview

- Introductory thoughts
- Turbulence modelling : errors and uncertainties
- Uncertainty quantification for RANS turbulence modelling
  - Tackling parametric uncertainties
  - Tackling structural uncertainties
- Conclusions and future trends

Introductory thoughts

- 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
- Physical/mathematical models: error or uncertainty?
  - Modeling errors : conscious use of a possibly unsuitable/partially suitable model for a given problem
  - Modelling uncertainties :
    - does a model fit a given problem?
    - How close it is to reality?

  →lack of knowledge that could be improved → **epistemic uncertainty**

Turbulence modelling : errors and uncertainties

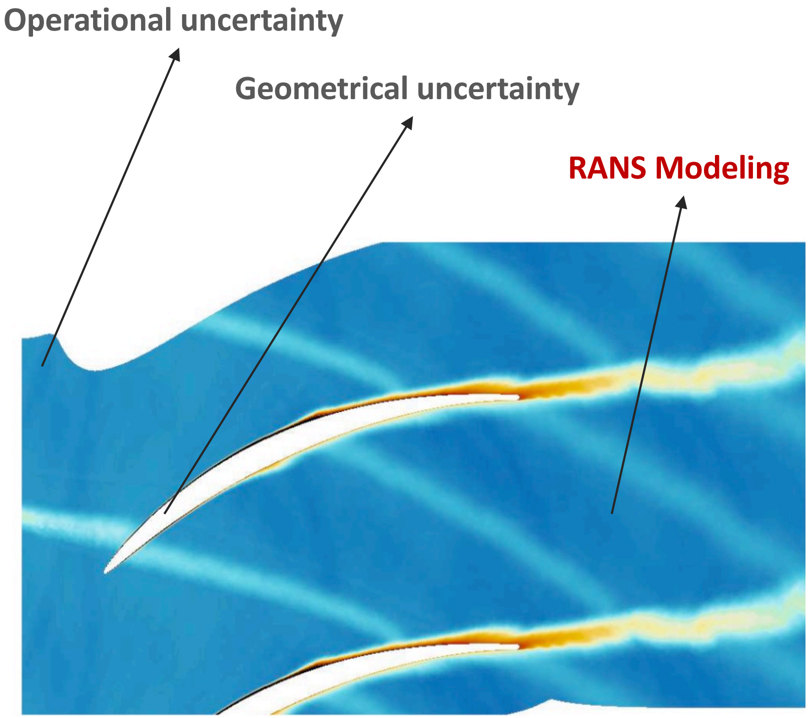
- Reminder:
  - DNS of turbulent flow fields computationally intractable for most practical cases
  - Several levels of approximation are possible, according to flow scales that are resolved/modelled (RANS, LES, hybrid RANS/LES)
  - Scale separation introduces unclosed terms that need to be modelled
  - Increasing the range of **modelled** scales lead to more flow-dependent, uncertain models
  - Increasing the range of **resolved** scales leads to a stronger influence of numerical errors/boundary condition treatment

Turbulence modelling : errors and uncertainties

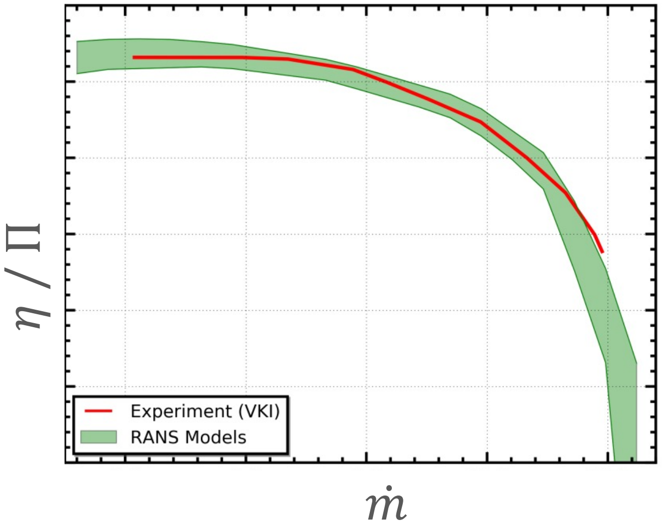
- Choice of the appropriate modelling level: essentially "expert judgement" :
  → **cost/accuracy considerations**
- For a given level
  - Several possible models, which differ by
    - Their mathematical structure
    - The associated closure parameters
- Common practice in turbulence modelling
  - Model structure chosen by <u>expert judgement</u>
  - Model constants <u>imperfectly known</u>/calibrated on simple flows

  → **Both are sources of uncertainty**

Uncertainties in CFD may lead to design failures



V103 Cascade – *(Leggett 2016)*



RANS modelling uncertainties

- Reynolds averaged Navier-Stokes (RANS) equations:
  - Define a suitable averaging operator (modeling choice)
  - Decompose field quantities into average and fluctuating parts

$$\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'; \qquad p = \bar{p} + p'$$

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

$$\frac{\partial \bar{\mathbf{u}}}{\partial t} + \bar{\mathbf{u}} \cdot \nabla \bar{\mathbf{u}} = -\frac{1}{\rho} \nabla \bar{p} + \nabla \cdot \left(\nu \nabla \bar{\mathbf{u}} - \boxed{\overline{\mathbf{u}'\mathbf{u}'}} \right)$$

Reynolds stresses

$$\tau_{ij} = \overline{u_i' u_j'} = 2k \left(b_{ij} + \frac{1}{3} \delta_{ij} \right)$$

b_{ij} = anisotropy tensor → must be modelled
k = turbulent kinetic energy

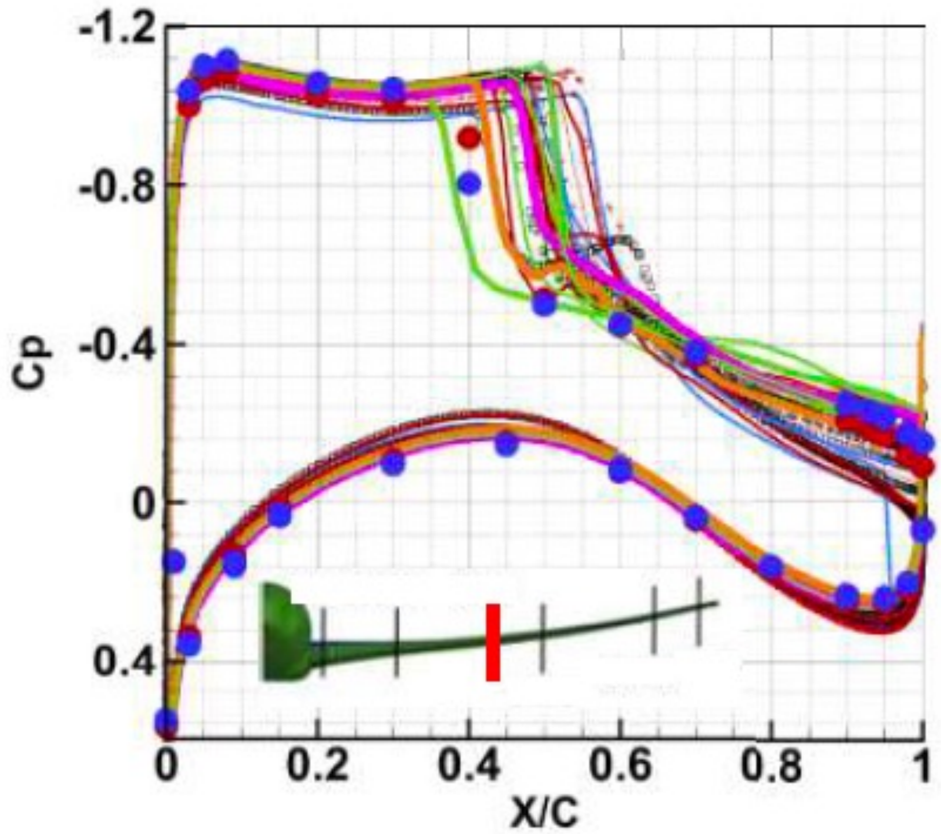
Reynolds stresses need a constitutive law: a **turbulence model**

1. Look for a mathematical formulation (model structure)
2. Look for closure coefficients (model parameters)

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

RANS modelling uncertainties

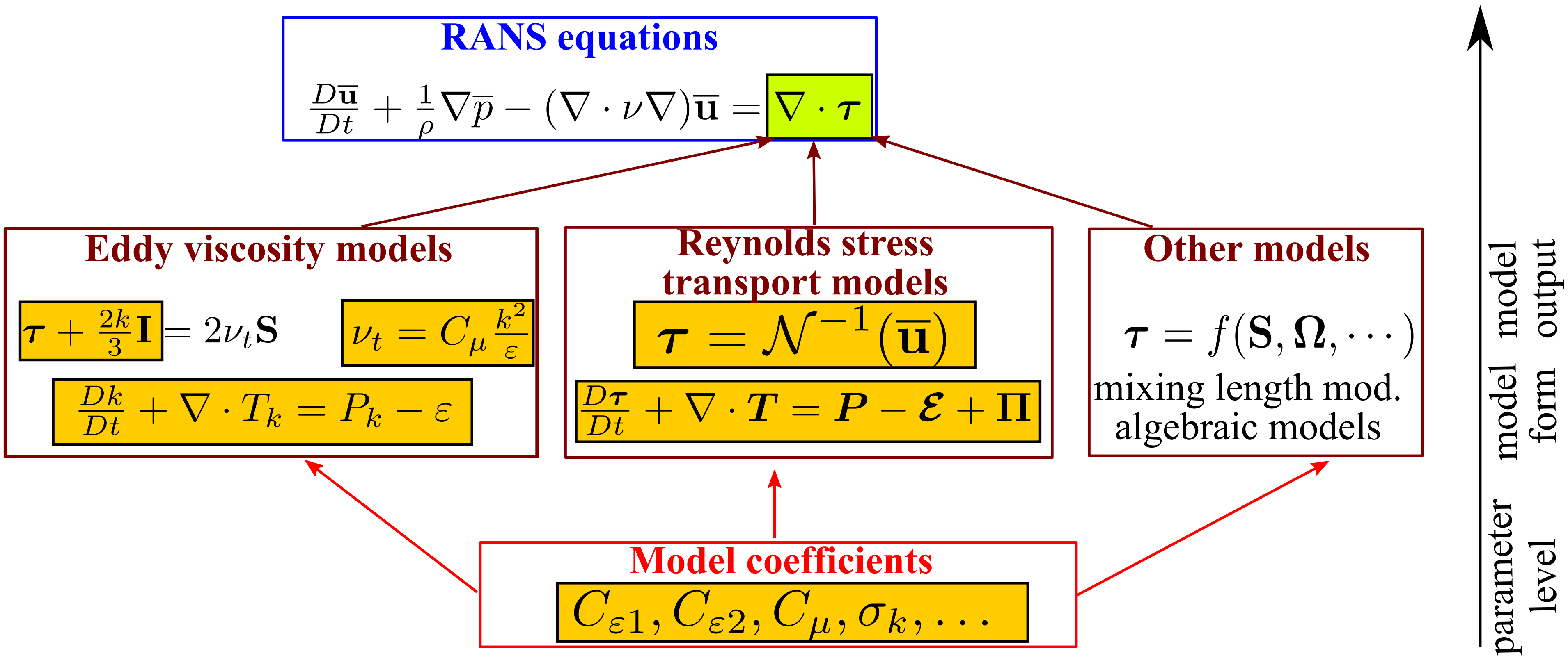
- **Model structure** classically derived from physical arguments
  - Integrates physical principles such as objectivity, symmetries, realizability
  - Relies on more or less crude modeling assumptions
- **Model parameters** calibrated for simple flows and from uncertain data
- Rich zoology of models of different complexities
- **No universally accepted model, no universal parameters**



Pressure distribution along a wing section from various RANS models (lines) and experiments (symbols).
6th AIAA Drag predition workshop.

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

RANS modelling uncertainties



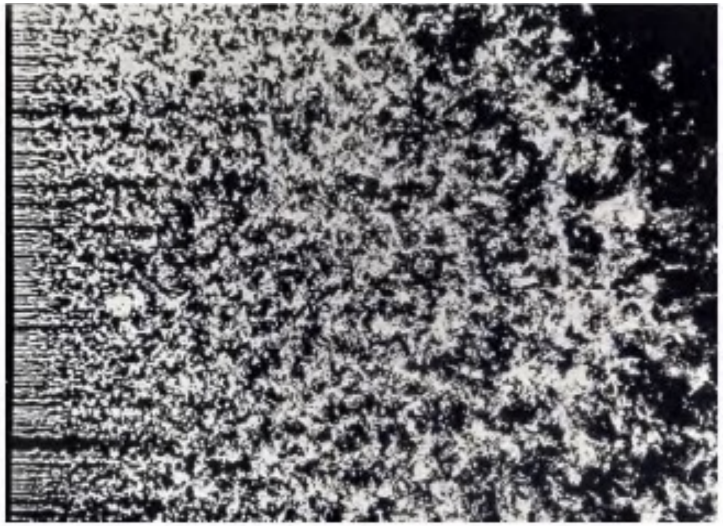
Comments on model coefficients

- Need for specifying closure coefficients
  - Calibrated in a deterministic (empirical) way

    …from uncertain experimental data

    …on simple flow configurations

    (e.g. isotropic decaying turbulence)
  - Example: k-ε model
    - Several versions (structure uncertainty)
    - Coefficients may vary a lot, e.g. $C_{\varepsilon 2}$
      - ✓ Commonly used 1.92
      - ✓ RNG k-ε 1.68
      - ✓ k-τ 1.83
      - ✓ Best-fit to data (n=1.3) 1.77

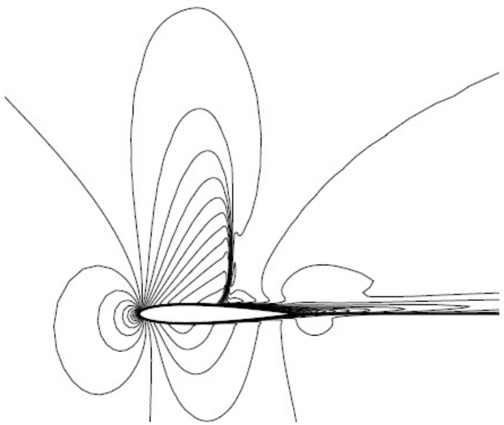


$$n = 1/(C_{\varepsilon 2} - 1)$$

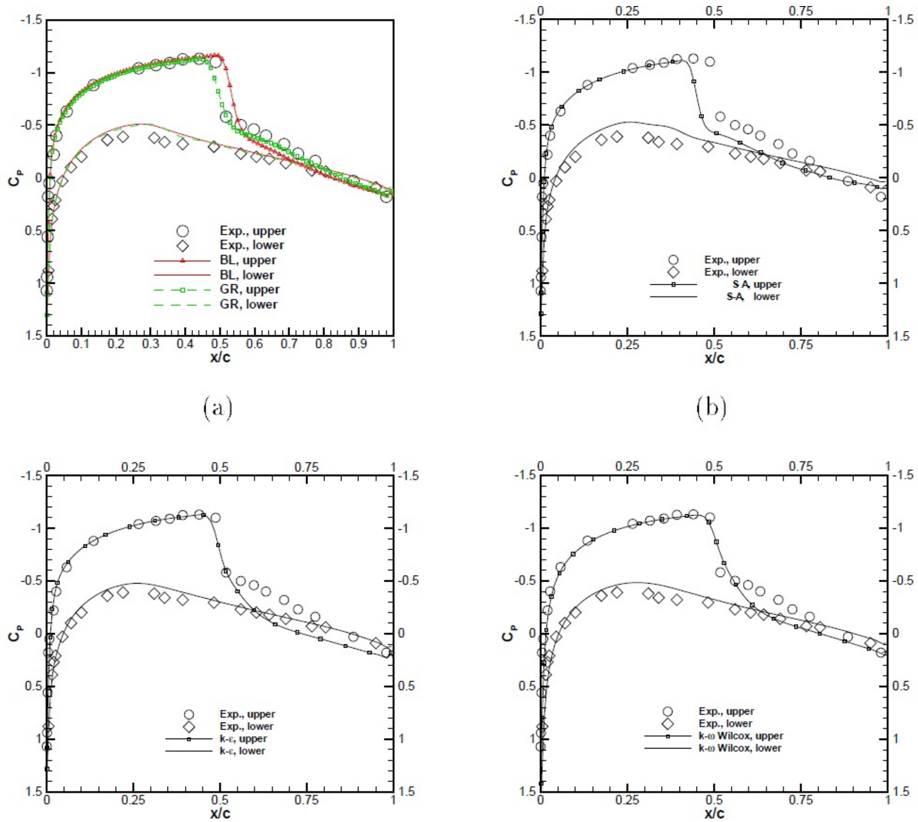
→ Model coefficients ARE NOT SACRED!!!

RANS modelling uncertainties

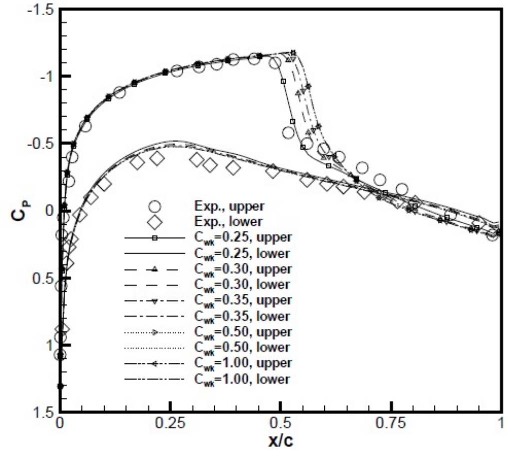
- Turbulent flow past an NACA0012 airfoil, M=0.8, AoA=2°, Re=9e6
- Perfect gas, Newtonian fluid, eddy viscosity turbulence model



IsoMach lines



Wall pressure coefficient, different models



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

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

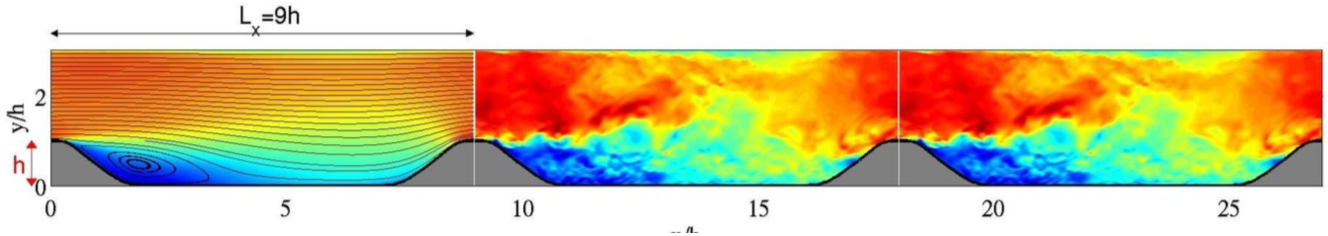
LES: Closure models for the subgrid stresses

- Not so many models as RANS, yet…
  - Smagorinsky model (1963)
  - Dynamic Smagorinsky model
  - Multiscale models
  - Transport-equation models
  - Implicit models
  - …

→ Model less critical then in RANS

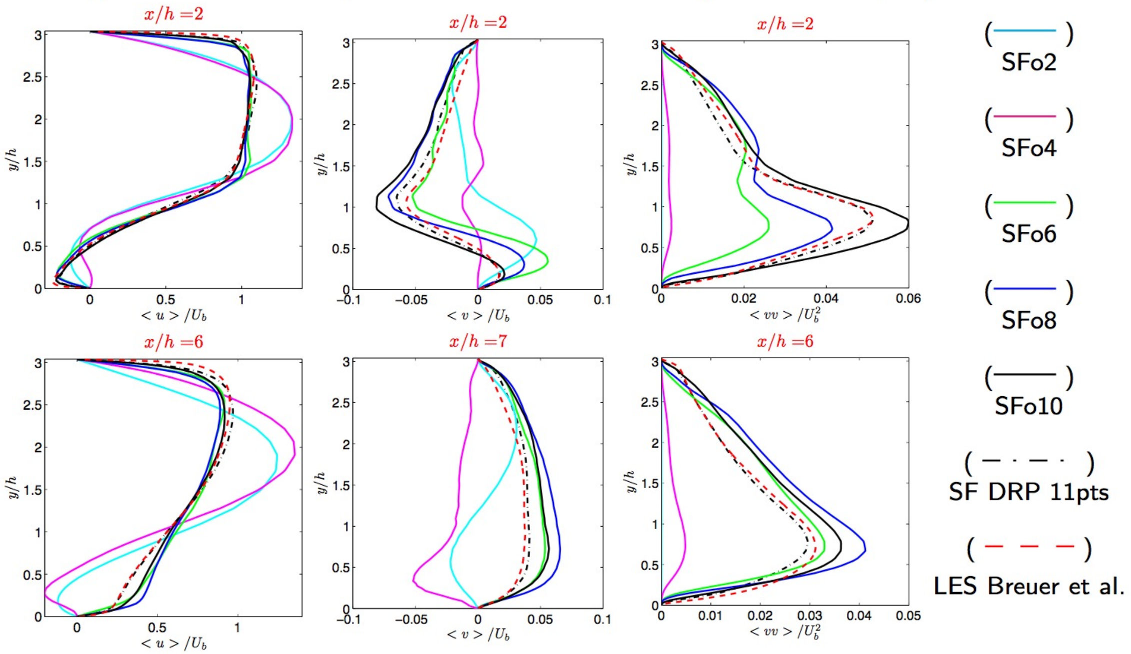
Numerics tends to be the primary source of uncertainty

LES errors/uncertainties



Periodic hill at Re=10595: numerical schemes

grid 64 × 33 × 32; influence of filter alone (FD DRP 11pts)



Average streamwise velocity (Gloerfelt & Cinnella, 2015)

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

RANS/LES uncertainties

- The most uncertain of all
  - RANS modeling uncertainties, e.g. in the boundary layers
  - LES errors/uncertainties: sensitivity to the numerical scheme, grid resolution, boundary conditions
  - Uncertainties associated with the treatment of grey zones
- Requires A LOT of expert judgement

In the following we focus on RANS models

Objectives

- "Improve" RANS models
  - More accurate prediction (not always possible)
  - **Reliability information** (how much wrong am I?)
    - Identify flows/flow regions with high RANS uncertainty
    - Quantify the uncertainty

- Two main approaches
  1. Model structure is not modified → parametric approach
     - Update model parameters
     - Bayesian mixture of models
  2. Model structure is modified → non parametric approach
     - « Data-driven » modelling, machine learning approaches

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

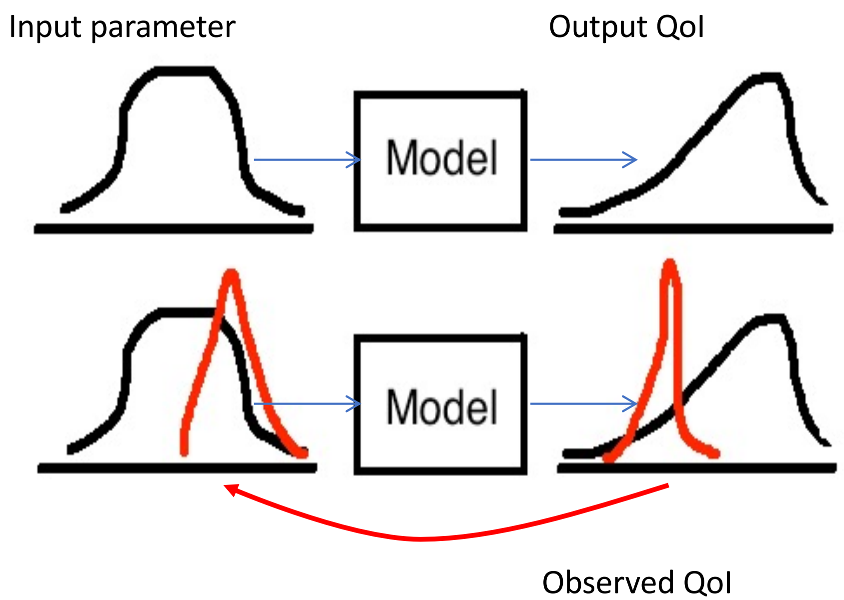
Interlude: a (very) short introduction to Uncertainty Quantification (UQ)

Quantification of parametric uncertainty

- Goal: quantifying and reducing uncertainties related with the choice of closure coefficients

Direct vs Backward UQ

- Direct
  - Sensitivity analysis of a QoI to model parameters
    [Platteuw et al. 2008, Roy & Oberkampf 2010, Edeling et al., 2013, Margheri et al. 2014]
- Backward
  - Infer posterior pdf of model coefficients from available observations
    [Cheung et al., 2011; Kato & Obayashi 2013; Edeling et al. 2013; Margheri et al. 2014…]



→Solve an inverse probabilistic problem

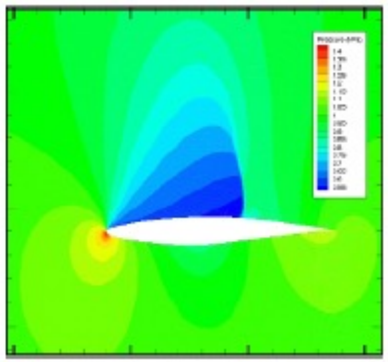
Forward step: direct UQ

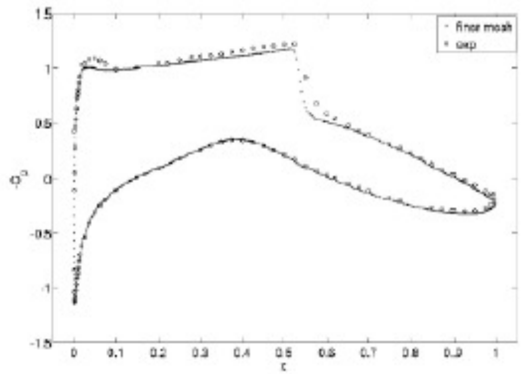
- Monte Carlo sampling converges slowly (as 1/√N) and is not applicable to computationally intensive problems
  - Multi-level Monte Carlo methods may reduce the cost
- Other methods
  - Non probabilistic methods:
    - Interval analysis
    - Method of moments (based on the use of sensitivity derivatives)
  - Probabilistic methods
    - Polynomial chaos
    - Probabilistic collocation
    - Surface response methods
    - …

Uncertainty quantification of complex models

- Consider a **computational model** for the transonic turbulent flow around an airfoil

$$M_\infty = 0.734$$
$$\alpha = 2.79^o$$
$$Re = 6.5 \times 10^6$$





- Common practice in Computational Fluid Dynamics (CFD):
  1. choose a model
  2. set the configuration
  3. run the corresponding computation
  4. Validate results against experiments (if available)

Uncertainty quantification of complex models

- Assume that free-stream conditions and geometry are not precisely known

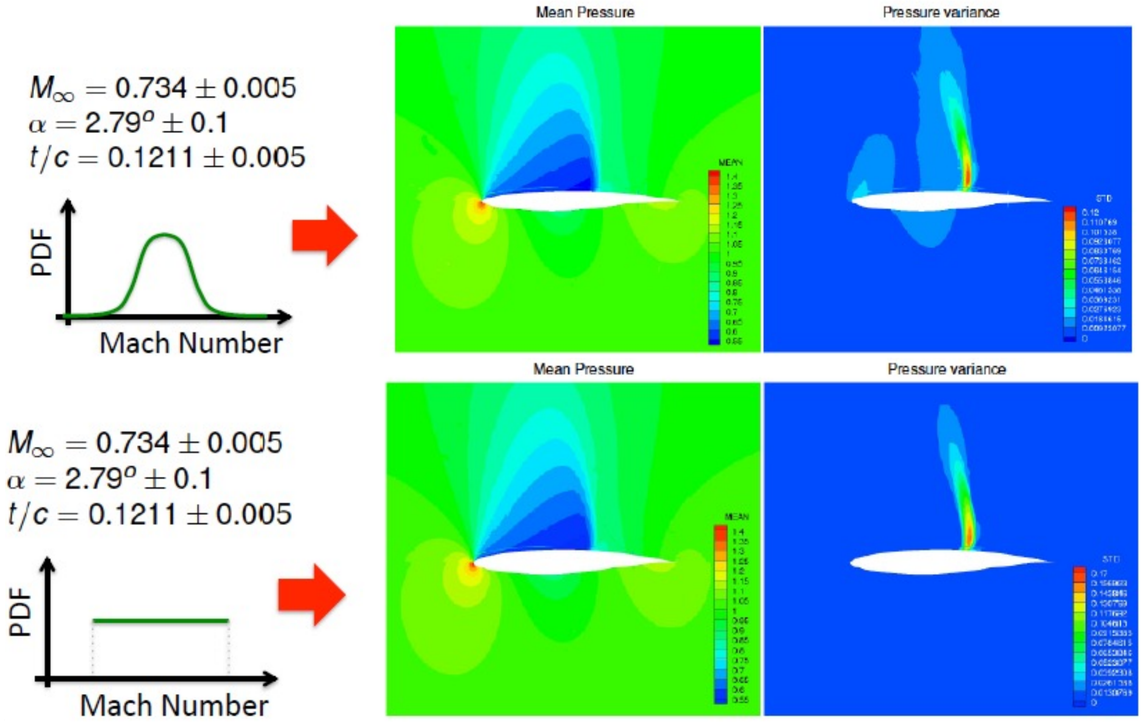
$$M_\infty = 0.734 \boxed{\pm 0.005}$$
$$\alpha = 2.79^o \boxed{\pm 0.1}$$
$$t/c = 0.1211 \boxed{\pm 0.005}$$

- How can I quantify the impact of this uncertainty on the outputs?

→ **<u>This is the role of UQ methods</u>**

Exemple: transonic airfoil flow

- The definition of the input pdf is critical
  - If the range is interpreted as a 95% confidence interval we can, e.g. model the input distribution as a multivariate normal
  - If we do not have any information → uniform pdf

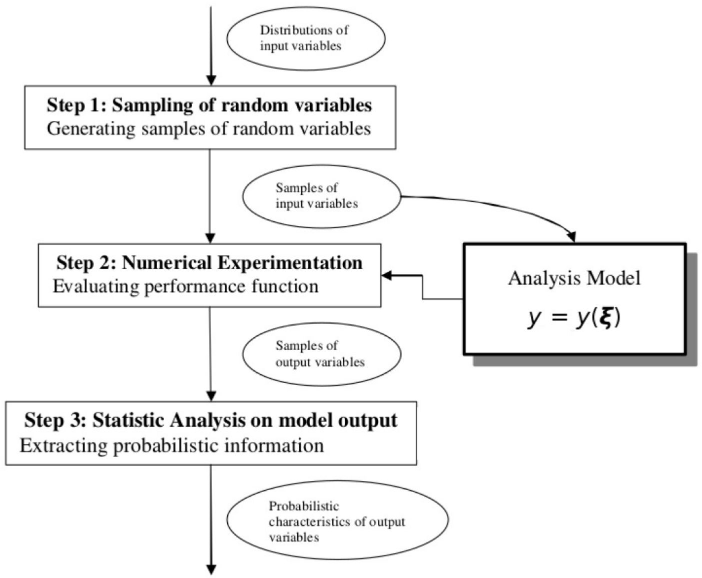


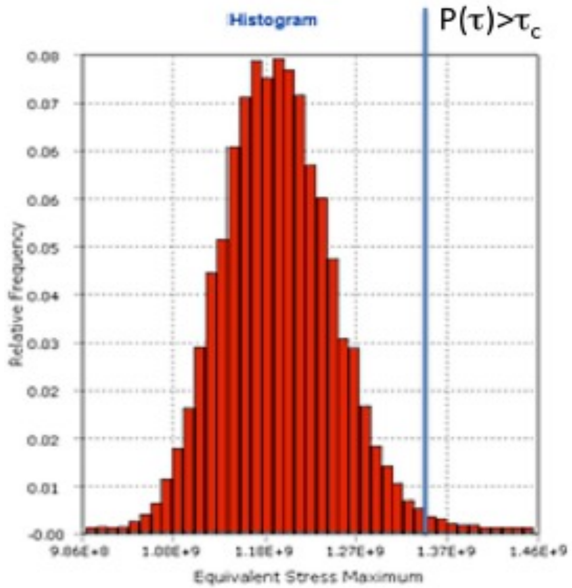
Monte Carlo Method

- Sample input random variables according to their pdf and generate N samples

$$\boldsymbol{y}^k \quad k = 1, \ldots, N$$

- Solve a deterministic model for each sample
- Compute an histogram of the output
- Compute solution statistics: mean, variance, probability of failure…
  - Central limit theorem → for a large enough sampling set, the approximated statistics converge to the true value





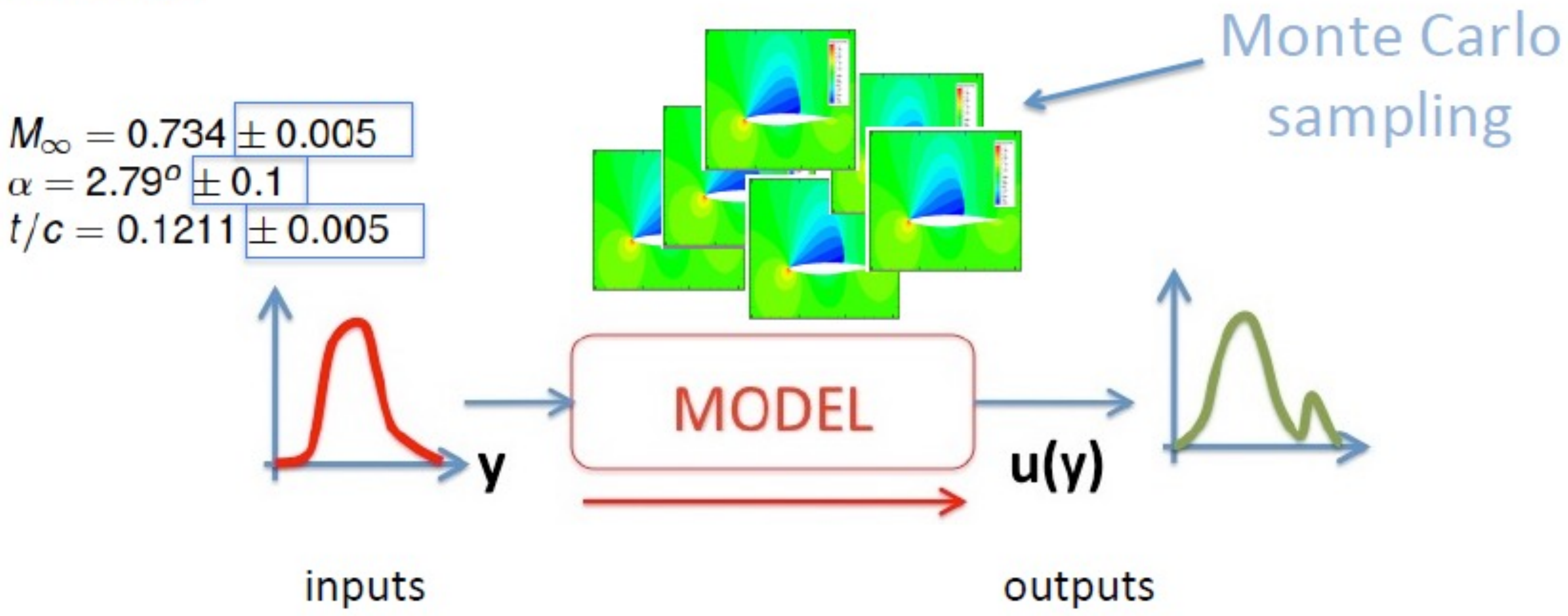
Monte Carlo Method

- Advantages:
  - Simple, parallel, non intrusive
  - The accuracy of MC increases an √N independently on the dimensionality d of the parameter space
  - May treat correlated parameters if their joint pdf is known

- Drawbacks:
  - Convergence in √N is too slow!
  - Inacceptably expensive for costly computer models even using multiple cores (maximum core number is limited)

Again the transonic airfoil…

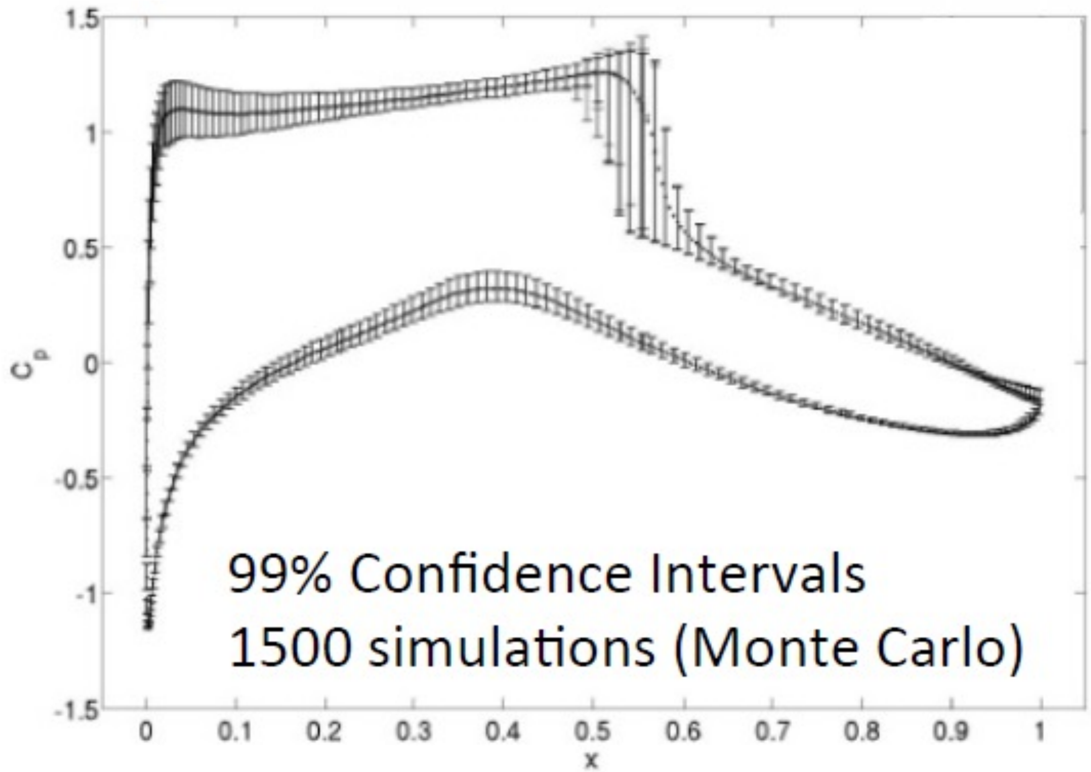
- Three input parameters are described in probabilistic terms, i.e. via a *pdf*



Again the transonic airfoil…

- Compute a set of simulations and determine statistics for the outputs (expectancy, confidence intervals, …)

$$M_\infty = 0.734 \pm 0.005$$
$$\alpha = 2.79^o \pm 0.1$$
$$t/c = 0.1211 \pm 0.005$$



Calibration

- The process of fitting the model to the observed data by adjusting the parameters.

- Calibration is typically effectuated by ad hoc fitting;
  after calibration the model is used, with the fitted input values, to predict the future behavior of the system

- Hereafter we look for statistical calibration techniques

Approaches to statistics

- Frequentist : assumes infinite sampling
- Likelihood : single-sample inference based on maximisation of the likelihood
  → disguised Bayesian…
- Bayesian (Bayes, Laplace) : unknown quantities are treated probabilistically and all knowledge can always be updated

Let us look in some more detail to the different philosophies

Frequentist/Likelihood vs Bayesian

- Non Bayesian: objective view of probability.
  - The relative frequency of an outcome of an experiment over repeated runs of the experiment
  - The observed proportion in a population

- Bayesian: subjective view of probability
  - Individual's degree of belief in a statement
  - Defined personally
  - Can be influenced in many ways (personal beliefs, prior evidence)

Bayesian statistics does not require repeated sampling or large *n* assumptions

Model calibration: problem statement

- Problem data
  - A model $\boxed{y=M(x, \vartheta)}$

    with $\boldsymbol{\vartheta}$ the unknown model random inputs and x the explanatory (known) variables
  - An *a priori* probability distribution for $\boldsymbol{\vartheta}$, $p(\boldsymbol{\vartheta})$ (for Bayesian only)
  - A sample of observations for y

- Problem outcome
  - An estimate for $\boldsymbol{\vartheta}$. This can be:
    - **STANDARD CALIBRATION (e.g. least mean squares)** : A « best fit » value for ϑ, ϑ^*, *no error estimate* or complicated
    - **BAYESIAN CALIBRATION**: The *a posteriori* probability distribution for $\boldsymbol{\vartheta}$

      → results from our *a priori* knowledge on $\boldsymbol{\vartheta}$, ***plus***

      the observation *likelihood*
    - An estimate of the model/measurement error variance

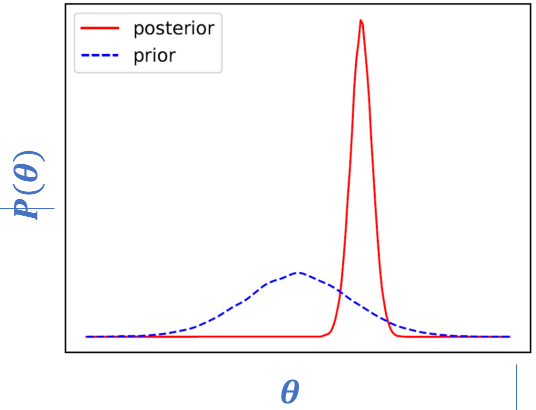
SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

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

Model calibration results from Bayes theorem on conditional probability:



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

where $p(\theta)$ is the prior belief about parameters θ;
$p(\mathbf{z}|\theta)$ is the likelihood (probability) of observing the data **z** given the parameters;
$p(\mathbf{z})$ is the evidence of the data can generally be treated as a normalization constant, such that the posterior integrates to 1
$p(\theta|\mathbf{z})$ is **the joint posterior distribution** of the parameters

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

It also updates the prior belief when new information becomes available

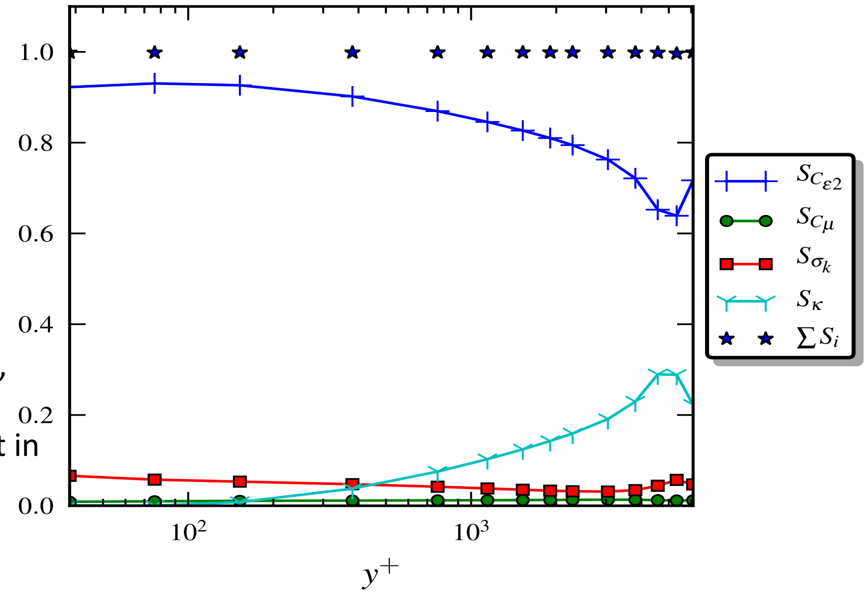
For most engineering problems, z results from running a computer code!!

→ **The posterior has to be computed numerically**

Incertitudes paramétriques des modèles RANS

Forward UQ of RANS models

- Sensitivity analysis of the QoI to model parameters (Platteuw, Loeven, Bijl, AIAA Paper 2008-2150, Edeling et al., JCP 2013, Margheri et al. 2014)
  1. Sample model coefficients
  2. Propagate through a CFD solver
  3. Summarize the output distributions for a
     parameters contributing the most to the

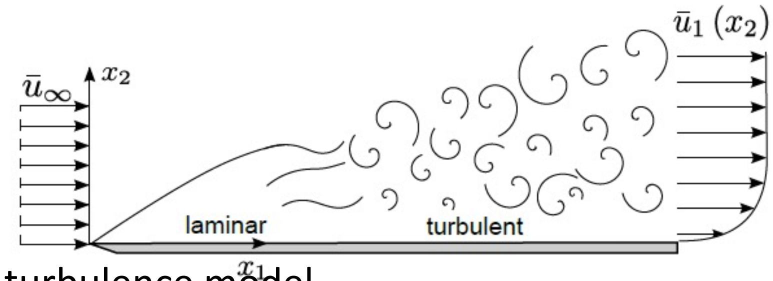


Sobol indexes of u+ vs y+ for a ZPG boundary layer (Edeling, 2013)

- $C_{\epsilon 2}$ is by far the most influential parameter
- The von Karman constant κ is also highly influential, espetially in the logarithmic and defect layers
- The diffusion coefficient σ_κ is particularly important in the viscous and inner logarithmic layers

Backward UQ: turbulent flow over a flat plate

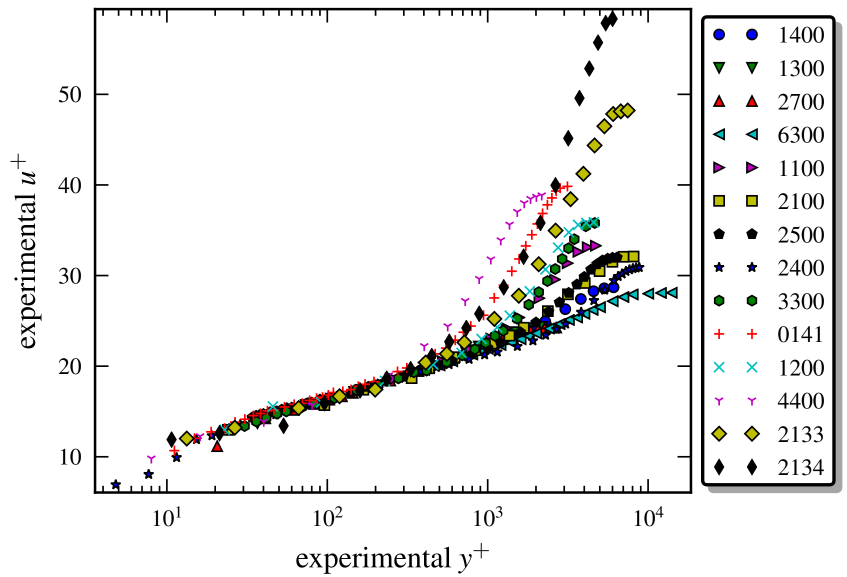
Objective: predict velocity profiles developing in the turbulent boundary layer close to the wall



- **Governing equations***:* RANS + turbulence model
  - Wilcox' boundary layer code (fast function evaluations)
  - Launder-Jones's (1972) k-ε model

- **Data** : 13 measured velocity profiles

  [Coles & Hirst, 1968]



Calibration setup

- **Likelihood model**: data **z** related to model outcomes $y(\theta_p)$, a function of the parameters via the multiplicative model error η and the observational error *e*, at each measurement point *i*

$$z_i = \eta_i y_i(\theta) + e_i \quad \text{with, e.g., the assuptions}$$

$$\eta \sim N(\mathbf{1}, K_{mc})$$, squared exponential correlation structure

$$e_i \sim N(0, \lambda^2)$$

- Numerical solutions: quick boundary-layer code

- Use Markov-Chain Monte-Carlo method to draw samples from the posterior

$$p(\theta|\mathbf{z})$$

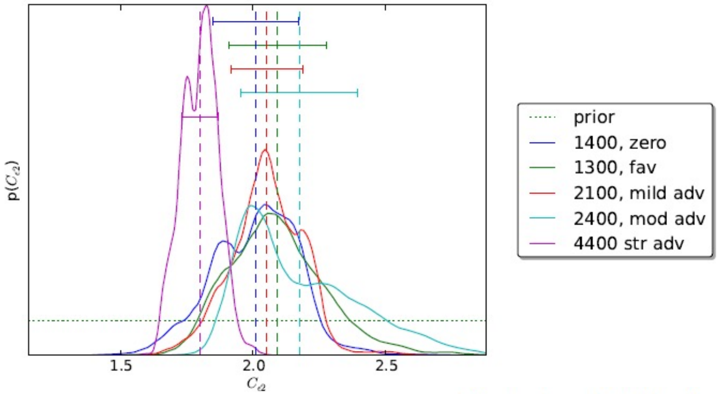
  - Python package pymc https://pymcmc.readthedocs.org/en/latest/#

- Other options: Kalman filters (MAP estimate), particle filters…

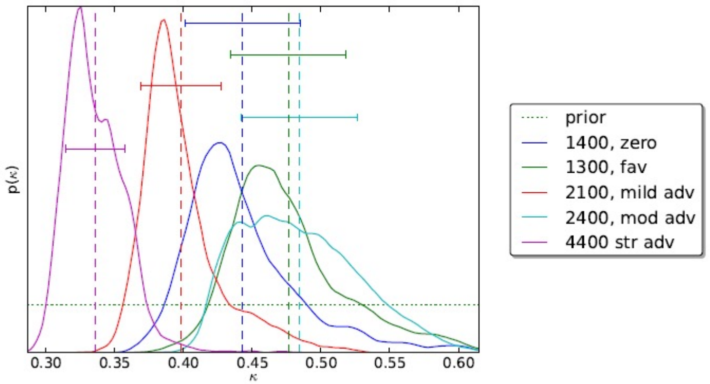
SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

Some results for the k-ε Jones-Launder model

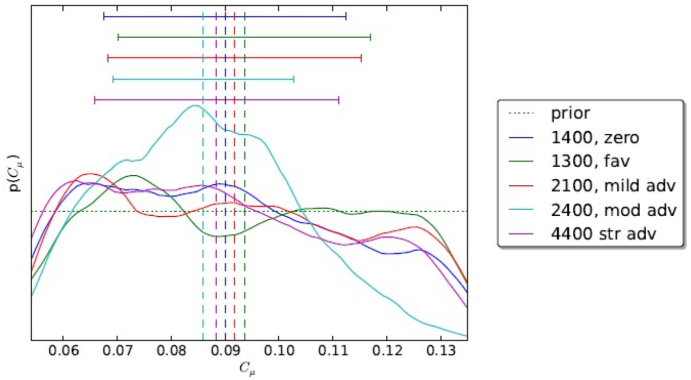
Posterior distributions for $C_{\varepsilon 2}$ for a favorable, zero, mild, moderate and strongly adverse $d\bar{p}/dx$.



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



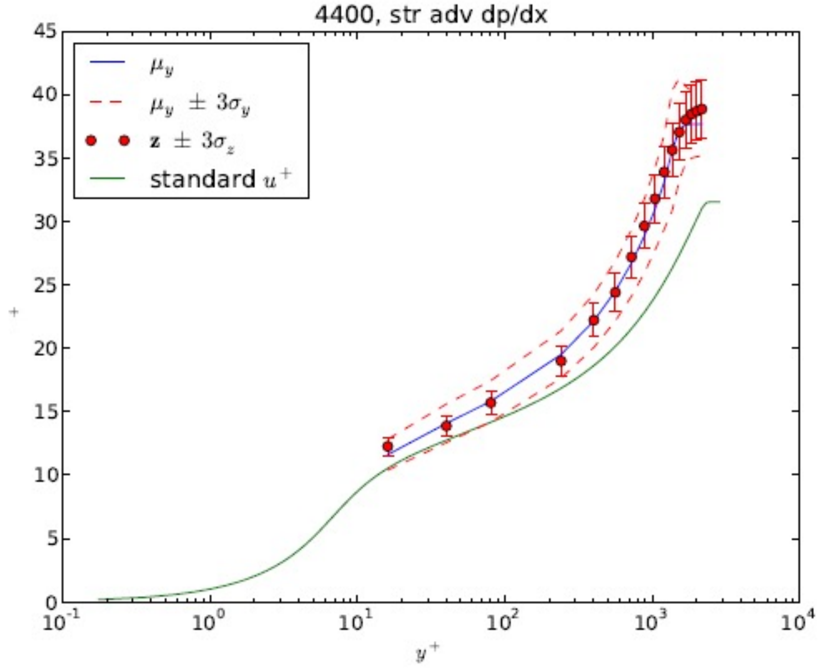
Posterior distributions for C_μ for a favorable, zero, mild, moderate and strongly adverse $d\bar{p}/dx$.



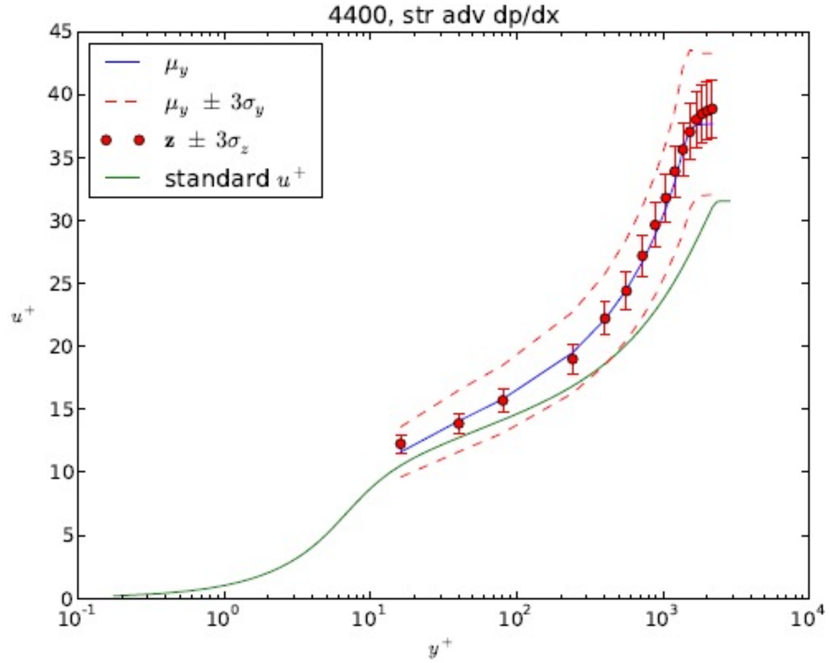
SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

Some results for the k-ε Jones-Launder model

- Posteriors are *propagated* through the RANS code to get the posterior estimate of the velocity profile
- Samples can also be drawn out of the model indequacy term



Posterior distribution of *y*



Posterior distribution of *ηy*

Lessons learned about parametric uncertainty

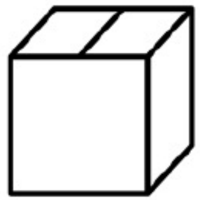
- Coefficients highly case-dependent
  - This reflects the structural inadequacy of the calibrated model
  - Including a model-inadequacy term (e.g. Kennedy & O'Hagan) partly alleviates overfitting, but is hardly extended to prediction cases
  - Parameter variability can be used to estimate structural uncertainty (p-box approach)

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

Incertitudes structurelles des modèles RANS

Dealing with structural uncertainties

- Much more challenging than parametric uncertainty
- Approaches applied to turbulence modelling
  1. Non-intrusive approach: Bayesian Model Averaging (BMA) [Edeling et al., 2014; 2018]
  2. Intrusive approaches: find a deterministic or stochastic correction of modeled terms (eddy viscosity, source terms in model transport equations, Reynolds stresses) by learning from data

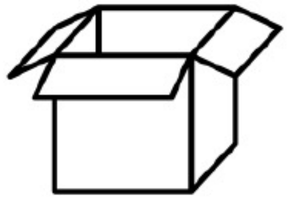


- **Black-box** machine learning → Difficult to get insight into models

[Singh & Duraisamy, 2016; Xiao et al., 2016; Ling et al., 2016]

- **Open-box** machine learning
  [Weatheritt & Sandberg, 2016; Edeling et al., 2018; Schmeltzer et al., 2018]

→ Tangible mathematical expressions



SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

Non-intrusive approach: Bayesian Model Averaging (BMA)

[e.g., Hoeting, 1999; Fragoso et al., 2018]

- Call $\mathcal{M}$ the space of all possible models, M a precise model, Δ a QoI which we want to predict for a new case, z the observed data (not necessarily for Δ):
- Approximate $\mathcal{M}$ using a finite discrete set of models $\mathcal{M}=(M_1, M_2,\ldots, M_N)$

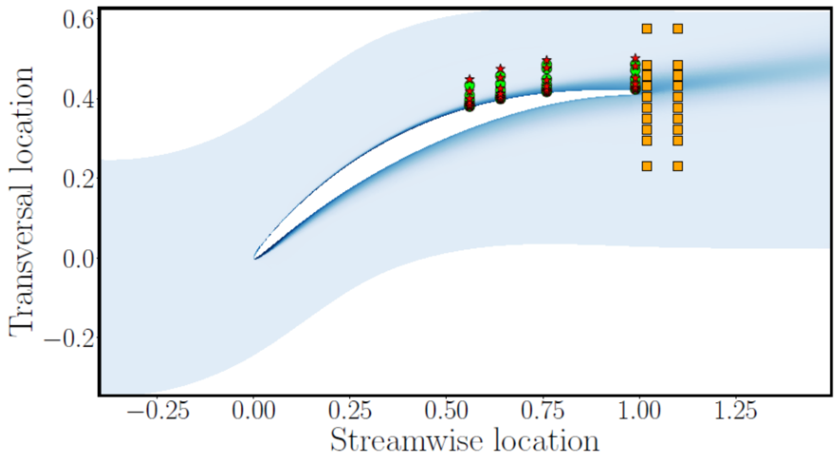
  → Use several concurrent models for predicting a new configuration
- Calibrate each model in $\mathcal{M}$ against z
- Make predictions from alternative models as a weighted average
  - Bayesian Model Average, BMA:

$$p\left(\Delta \mid \boldsymbol{M}, z\right) = \sum_{i=1}^{N} p\left(\Delta \mid M_i, z\right) P\left(M_i \mid z\right)$$

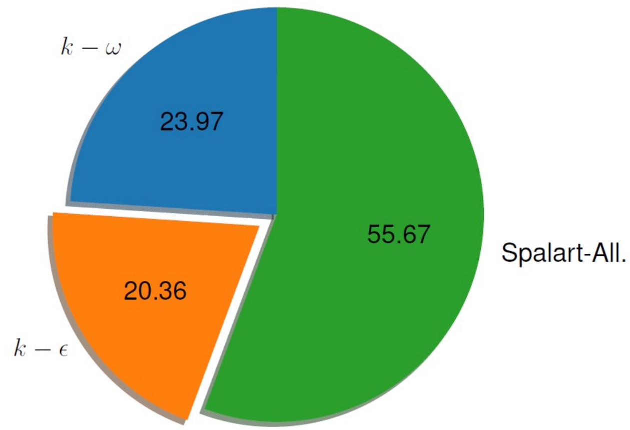
  - The weights are the posterior model probabilities (related to the denominators of Bayes' formula when calibrating each concurrent model M_i against z)

Non-intrusive approach: Bayesian Model Averaging (BMA)

- **Model Probabilities**
  - >We compare RANS models results with REFERENCE data (experiments, LES, DNS,…) for each probe.
  - >Statistical tools (Monte-Carlo) lead to a global probability for each model.
  - >We call it posterior model probability $P(M_i|D)$, and it will be the weight in the model averaging.



Probabilities of model for scenario S_3 $(AoA = 44°)$



Non-intrusive approach: Bayesian Model-Scenario Averaging (BMSA)

[Edeling, Cinnella, Dwight, JCP 2014, AIAA J 2018]

- Bayesian Model-Scenario Average, BMSA: extension of BMA
- Remark: $\mathcal{M}$ can be trained against multiple competing datasets!
- Which dataset is the most suitable to train the BMA for predicting a given new flow?
  - Call $\mathcal{S}=(S_1, S_2,\ldots, S_K)$ the set of all available calibration scenarios
  - Call z_k the data observed for scenario S_k (not necessarily for Δ)
  - Calibrate each model in $\mathcal{M}$ against each scenario in $\mathcal{S}$
  - Make predictions from alternative models as a weighted average over all models and scenarios

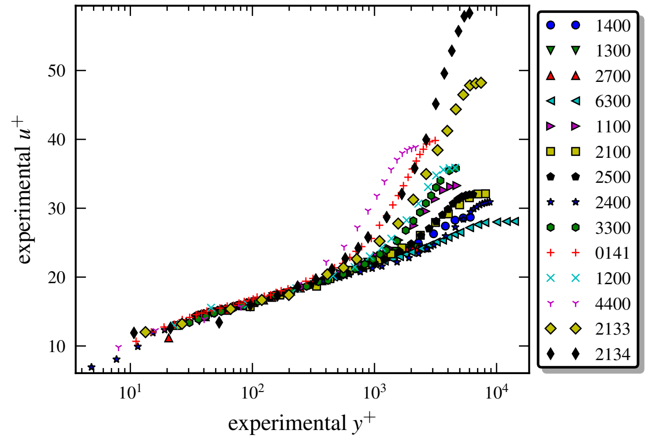
$$p\left(\Delta \mid \boldsymbol{M},\boldsymbol{S}\right)=\sum_{i=1}^{N}\sum_{k=1}^{K} p\left(\Delta \mid M_i,z_k\right)P\left(M_i \mid z_k\right)P\left(S_k\right)$$

  - The weights are the posterior model probabilities AND scenario probabilities (to be assigned a priori)

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

Scenarios

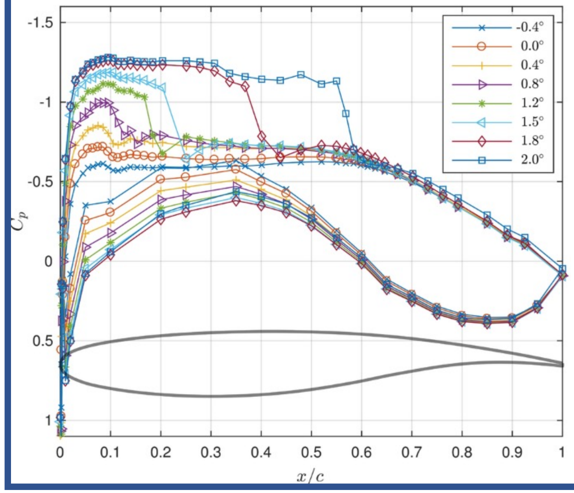
- **Simpler case**: Same class of flows, various operating conditions :
  - Example: boundary layers with various pressure gradients

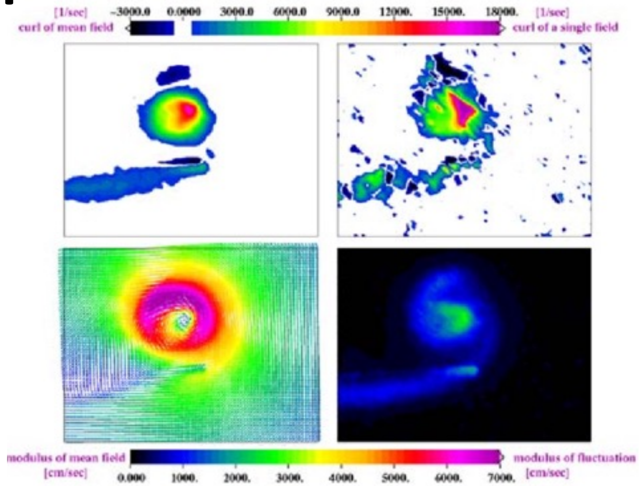


- **More complex case** : various flows (see exemple below)

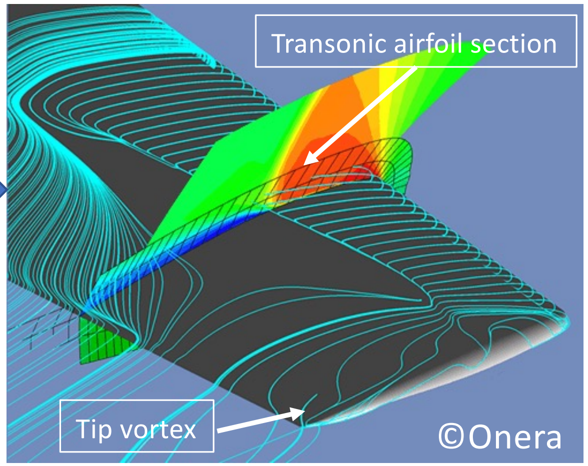
Calibration scenarios

« 2D airfoil » scenarios + « tip vortex » scenarios



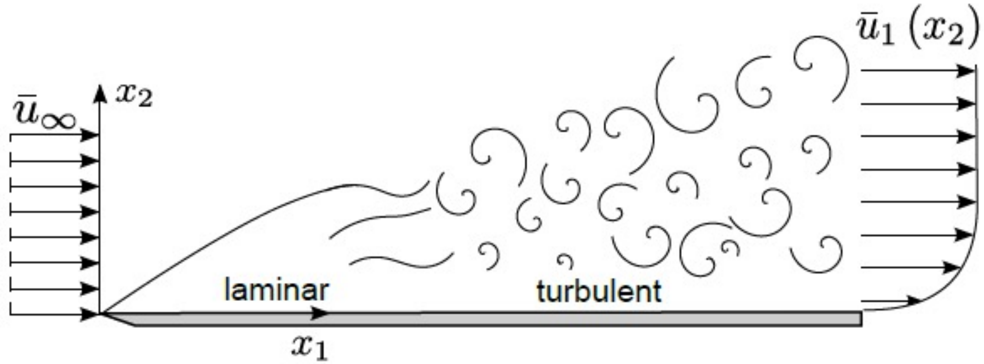


Prediction scenario



BMSA: Turbulent flow over a flat plate

Objective: predict velocity profiles for flat plate boundary layer subject to various gradients



Governing equations: RANS + turbulence model

- Algebraic Baldwin-Lomax' (1972) model
- Launder-Jones's (1972) k-ε model
- Menter's (1992) k-ω SST model
- Spalart-Allmaras (1992) model
- Wilcox' stress-ω model (2006)

Data: 13 velocity profile measurements from [Coles & Hirst, 1968]

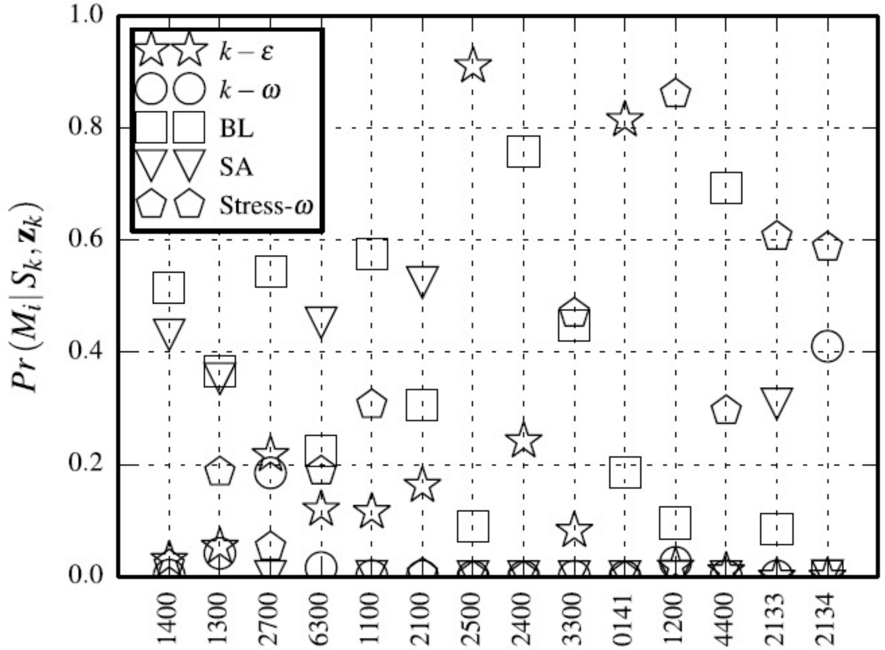
Bayesian model-scenario averaged prediction of the profiles : requires 5x13 UQ runs

BMSA : model probabilities

- Posterior model probabilities computed for all models in M for each z_k by sampling

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

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



Large spread in model probabilities when changing the pressure gradient scenario

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

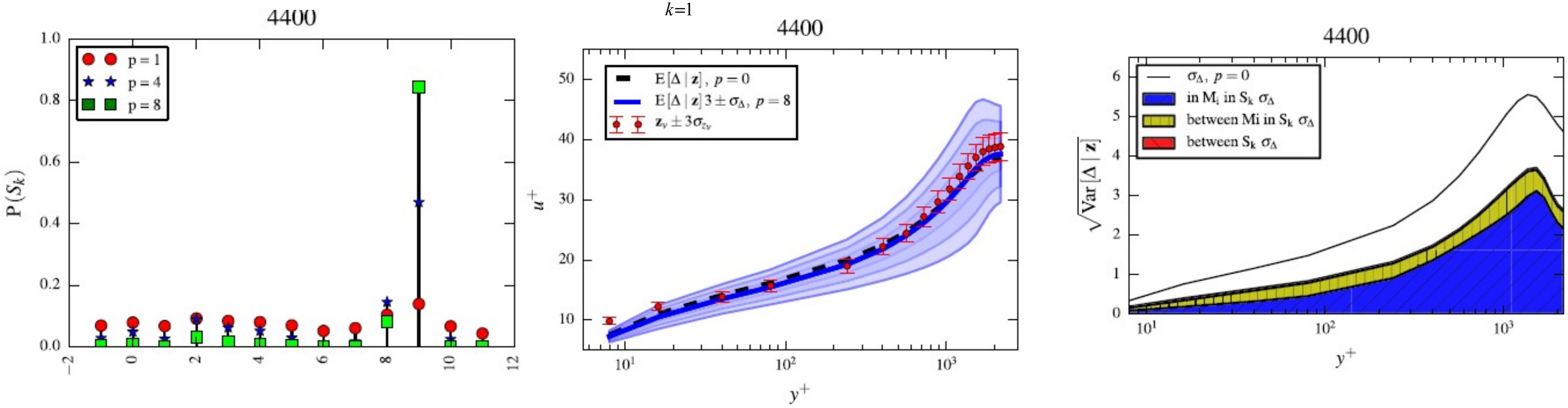
BMSA prediction

- Scenario pmf uniform (overconservative) or weighted according to an error measure

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

$$\varepsilon_j = \sum_{i=1}^{m} \left\| E\left[\Delta \,|\, z_j, M_i\right] - E\left[\Delta \,|\, z_j, \mathcal{M}\right] \right\|_{L_2} \qquad \forall z_j \in \mathcal{Z}$$

$$P(z_j) = \varepsilon_j^{-p} \,/\, \sum_{k=1}^{s} \varepsilon_k^{-p}$$



Prediction of an adverse pressure gradient BL using 5 models calibrated for 13 different pressure gradients

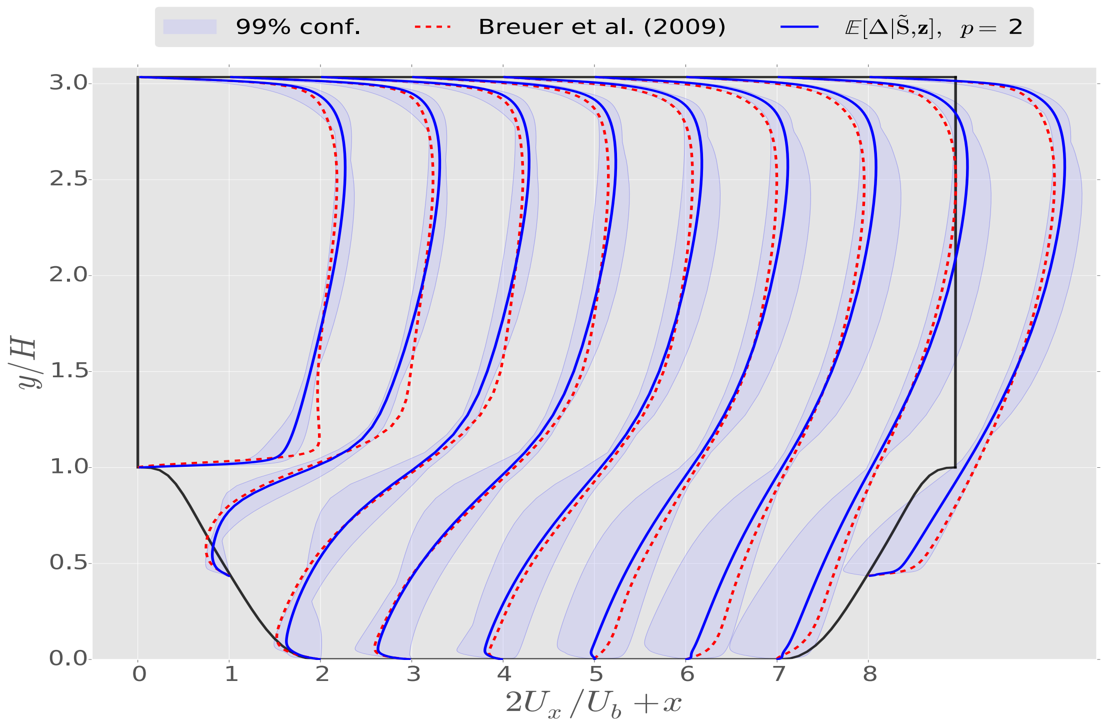
$\boxed{\text{Good prediction, variance consistent with experimental uncertainty}}$

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

BMSA: Flow over a periodic 2D hill at Re=5600

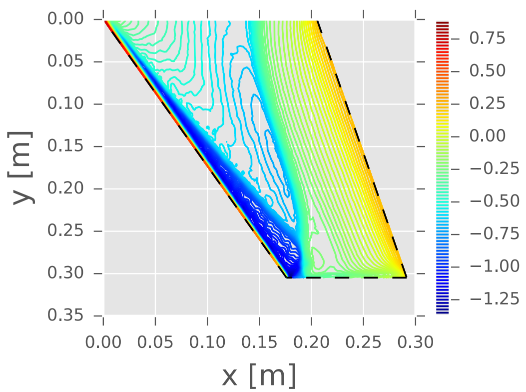
- Prediction of a flow configuration far from RANS safety margins
- Models: Spalart-Allmaras, Jones-Launder, Wilcox
- Propagation of the 13 boundary layer MAP estimates of the parameters through SIMPLEFOAM
- Comparison with DNS data of Breuer et al.

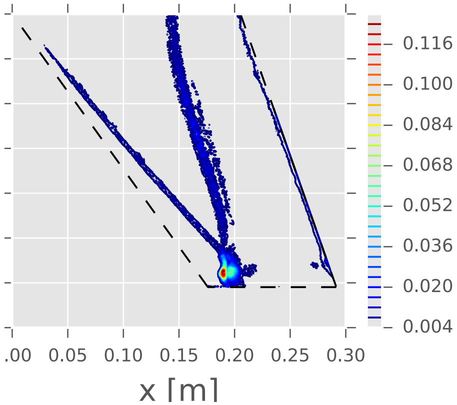
Velocity profiles at various downstream positions.



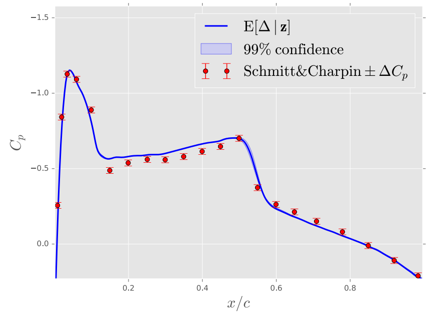
BMSA: Transonic flow past a wing

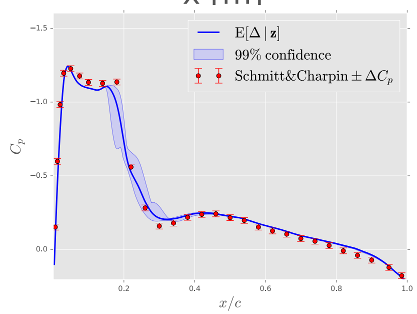
- Prediction of the pressure coefficient for transonic flow past the ONERA M6 wing:
  - M=0.8395, AoA=3.06°
- Results based on two models (Jones-Launder & Spalart-Allmaras)
- Propagation of the 13 boundary layer MAP estimates of the parameters through FLUENT
- Scenario weights computed locally in each section.





Predicted pressure coefficient C_p



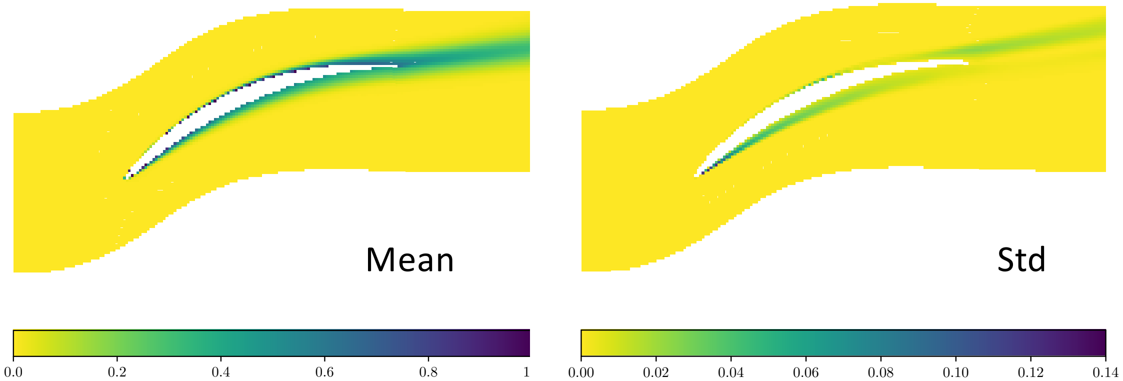


- Low model uncertainty at the wing root
- Uncertainty in shock locations
- High uncertainty in the tip region (shock/tip vortex interaction)

BMSA: Flow through a compressor cascade

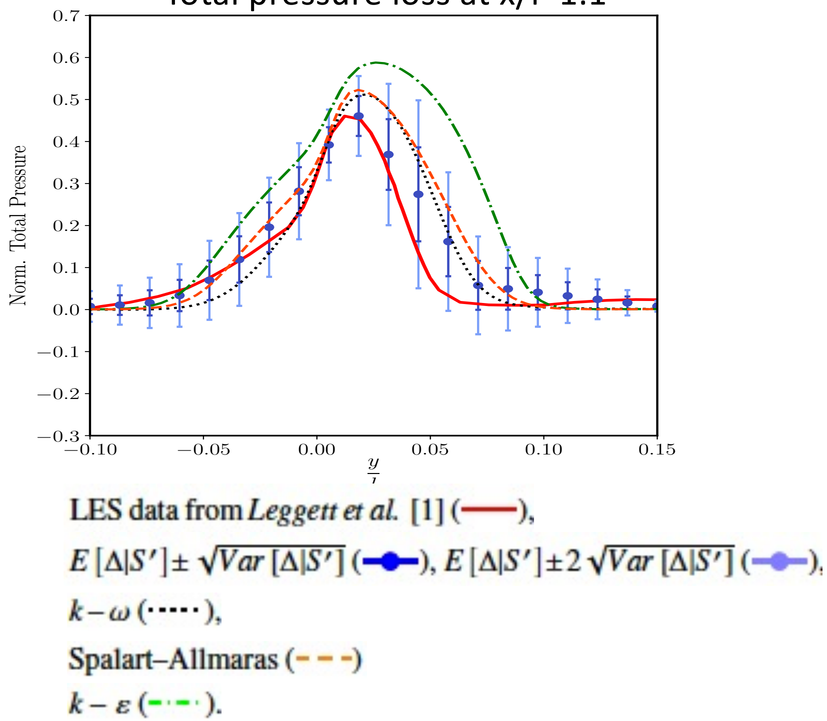
- Prediction of compressible flow through a compressor cascade (NACA65 V103) at off design conditions
- Results based on three models ($k - \omega$ Wilcox, $k - \varepsilon$ Launder-Sharma & Spalart-Allmaras)
- Propagation of the 13 boundary layer MAP estimates AND of 3 MAP estimates calibrated against LES data for the NAVA65 V103 cascade at operating conditions different from prediction ones

From De Zordo-Banliat et al., C&F, 2020



Total pressure loss field: mean and standard deviation

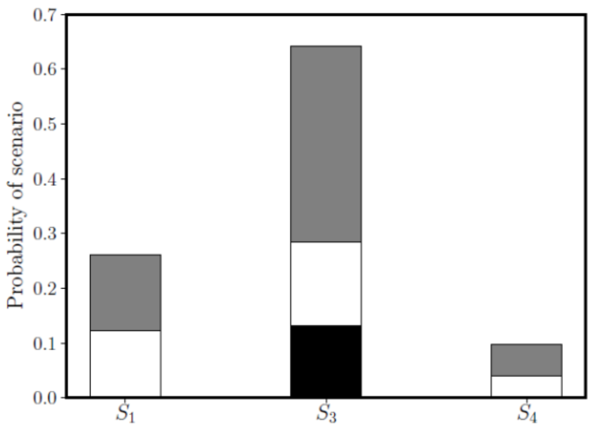
Total pressure loss at x/l=1.1



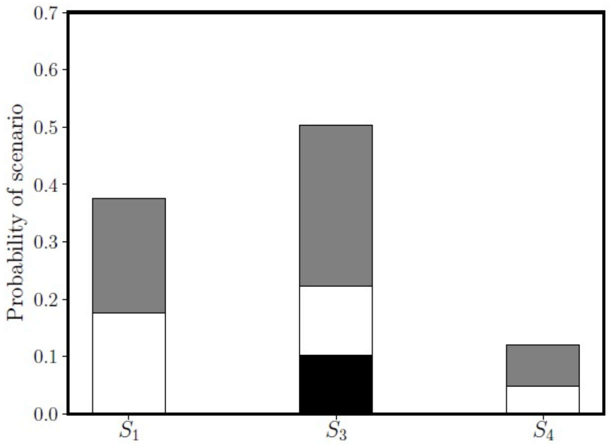
BMSA: Flow through a compressor cascade

- **Three formulations** for $P(S_k)$ were tested

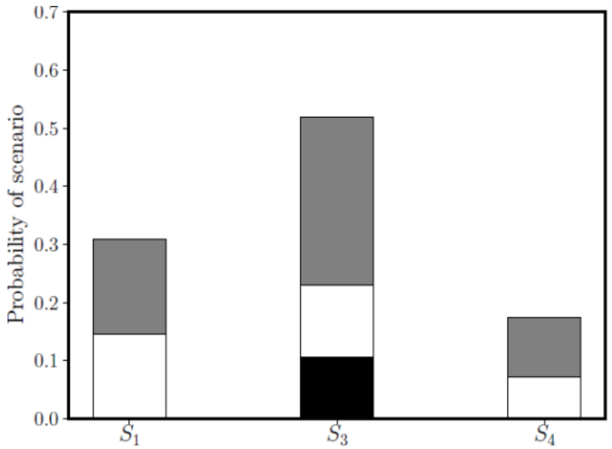
>Consensus based criterion



>Calibration-driven criterion



>Operating condition-based criterion

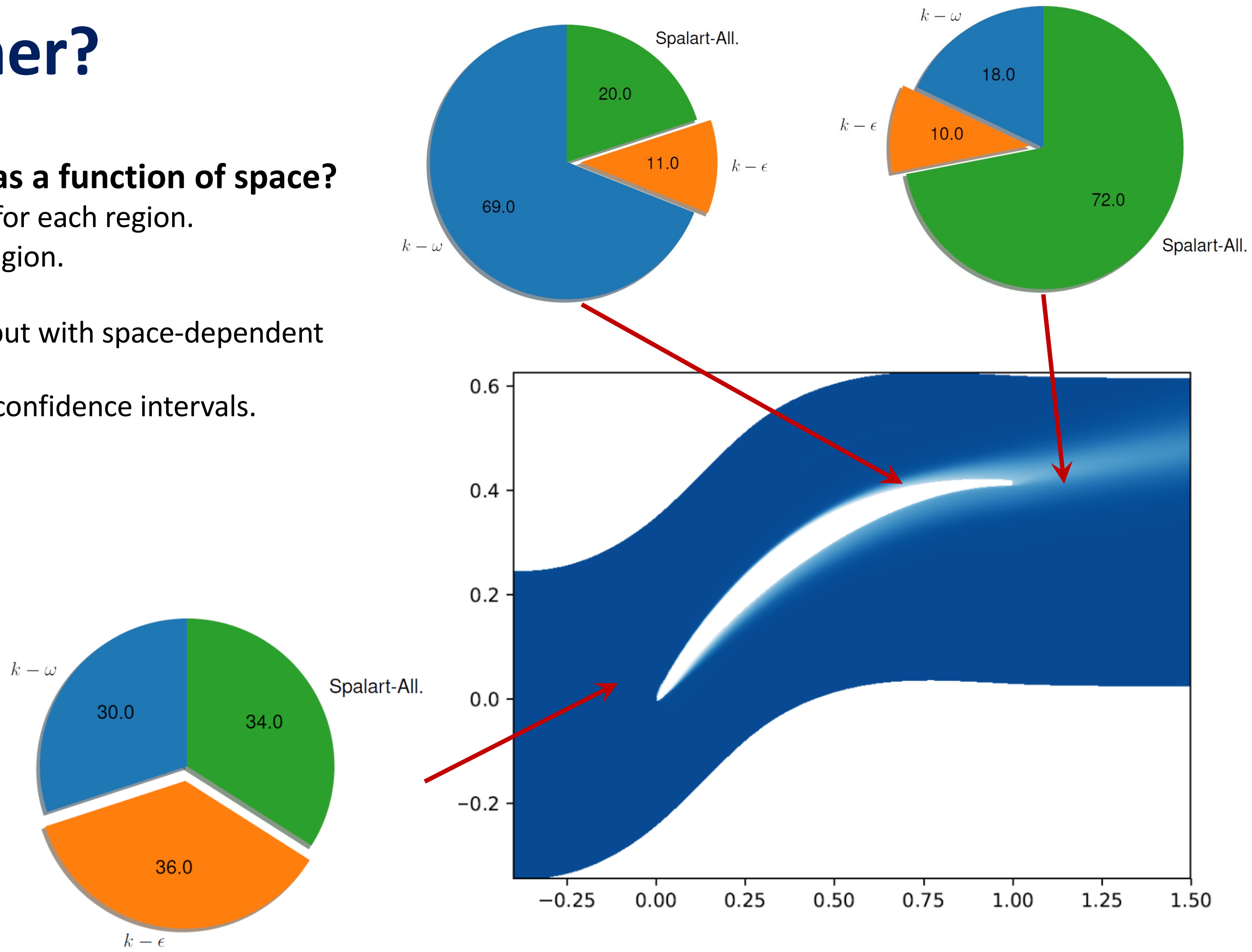


Scenario and model probabilities used for BMSA prediction:
$k-\omega$ Wilcox (white), $k-\varepsilon$ Launder-Sharma (black), Spalart-Allmaras (grey)

Can we go further?

- **Could we compute $P(M_i|D)$ as a function of space?**
  - Quantify the model probabilities for each region.
  - Identify the best model in each region.

  - Use similar model combinations but with space-dependent weights.
  - Further improve predictions and confidence intervals.



Space-dependent Bayesian Model Averaging

- BMSA uses the same weights throughout the flow field → contrary to expert judgment
- Further progress: compute $\boldsymbol{P}(\boldsymbol{M_i}|\boldsymbol{D})$ as a function of space
  - Infer model probabilities for each flow region
  - Identify the “best” model (if any) in each region
- Clustered Bayesian Model Averaging (CMA) [Yu, 2011]:
  regression of weights using decision trees
  - For a new point x_t, the average prediction on the ensemble of trees gives the weights of the models.
  - The final prediction is a space-dependent model average with weights w_j

$$y_{final} = \sum_j w_j \, y_j$$

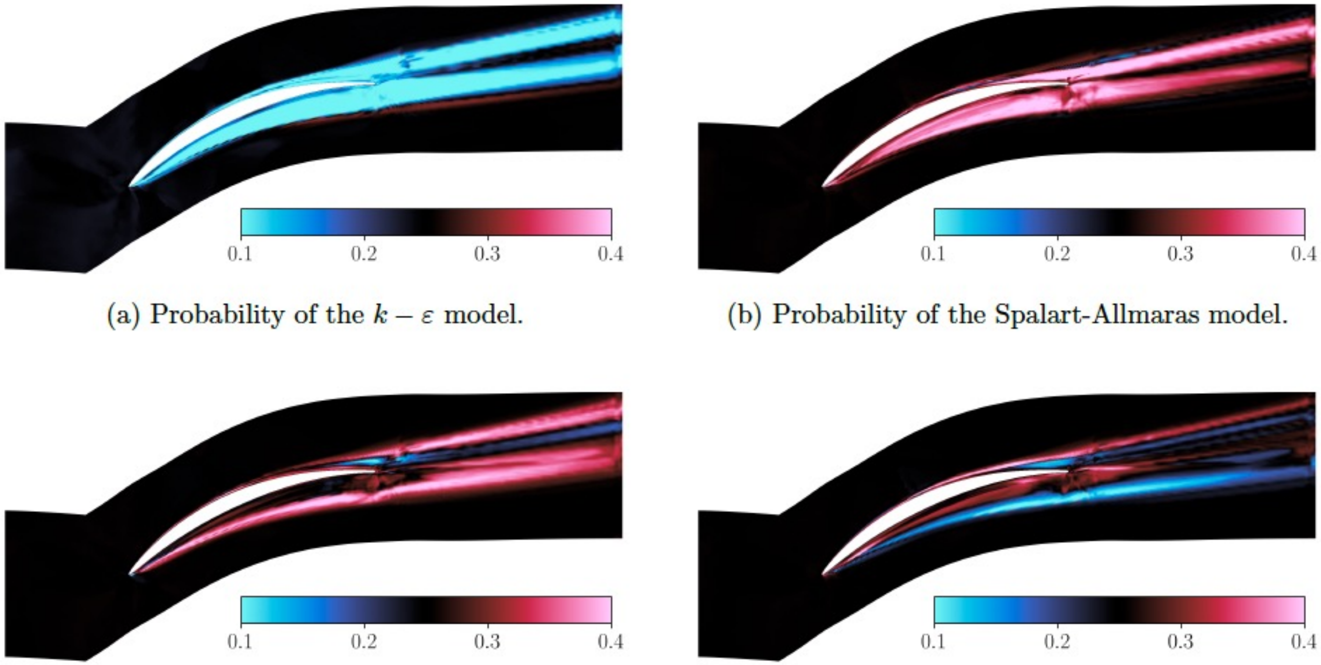
- Adaptation of CMA to CFD problems. Use of random forests instead of decision trees → XBMA (thèse M. de Zordo-Banliat)

XBMA : spatially-varying BMA

- XBMA of $k-l, k-\omega$, $k-\epsilon$ & Spalart-Allmaras as models. EARSM as reference data

  Data: 40080 points per scenario **Random forest**: 1000 trees

2D distribution of model weights



(a) Probability of the $k-\varepsilon$ model.

(b) Probability of the Spalart-Allmaras model.

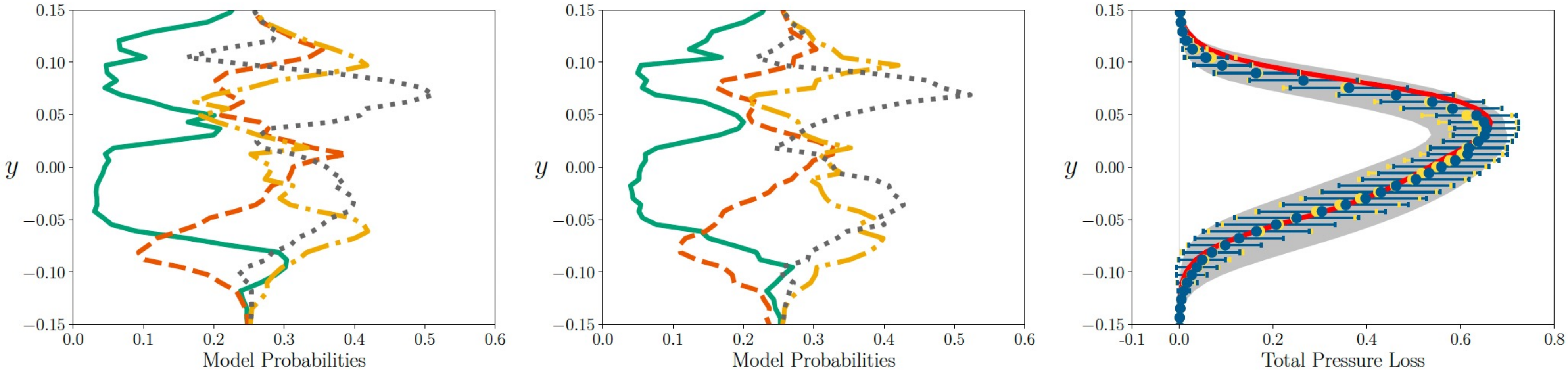
(c) Probability of the $k-l$ model.

(d) Probability of the $k-\omega$ model.

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

XBMA : spatially-varying BMA

- XBMA of $k-l, k-\omega\ , k-\epsilon$ & Spalart-Allmaras as models. EARSM as reference data
  Data: XBMA1, 40080 points per scenario, XBMA2, 820 points per scenario **Random forest**: 1000 trees



(a) Model probabilities obtained with $XBMA_1$.

(b) Model probabilities obtained with $XBMA_2$.

(c) $XBMA_1$ and $XBMA_2$ predictions superposed.

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

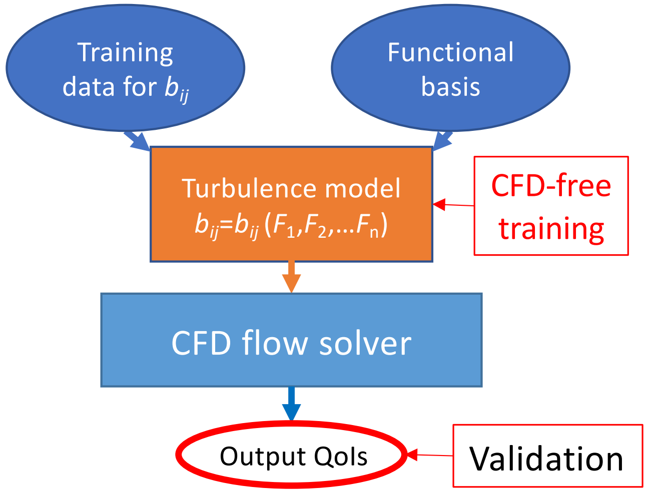
Apprentissage de modèles RANS

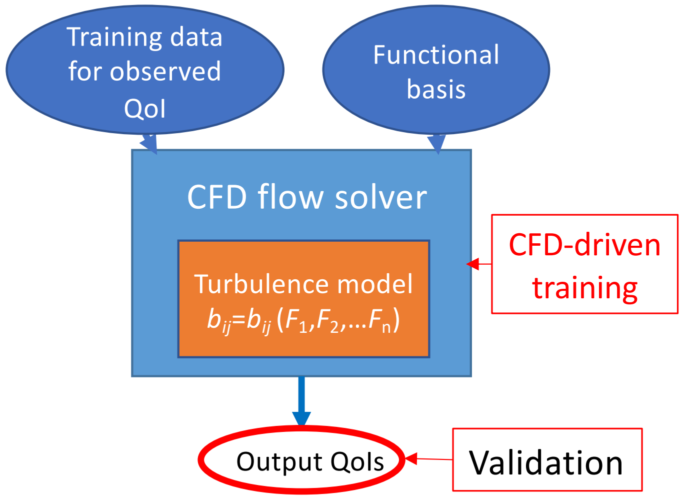
Using data for predicting turbulence?

- Rapidly increasing mass of **high-fidelity** flow field data
  - Turbulence-resolving simulations
    - Complete flow-field description, low residual uncertainty
    - Limited to simple configurations, low to moderate Reynolds numbers
  - Flow measurements (highly resolved PIV, stress-sensitive films, MEMS):
    - More complex configurations, high Reynolds numbers
    - Incomplete and possibly noisy data
- Use data to inform lower-fidelity RANS model
  - Inform parameters without changing model structure (model calibration)
  - Inform model structure (model identification)
- **Challenges:**
  - Much smaller (but well resolved) amount of training data than in typical IA applications
  - Use of possibly incomplete and noisy data
  - Estimate predictive uncertainties

Using data for predicting turbulence?

- General framework
  - No longer a « universal » model, but a model that generalises as well as possible to a **class** of flows
    - Choose a functional basis
    - Enforce physical constraints (whenever possible)
    - Train against data
- Two training strategies
  - CFD-free training
    - ☺ Inexpensive (manipulate analytical expressions)
    - ☹ Requires high-fidelity, low noise data for turbulent quantities
    - ☹ Does not warrant exact energy conservation
    - ☹ May lead to non robust models
  - CFD-driven training
    - ☺ May use virtually any data (mean flow and turbulent quantities)
    - ☺ May ensure energy conservation
    - ☺ Produces robust models
    - ☹ Requires the solution of a VERY costly multidimensional optimization problem





SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

« Data-driven » frameworks

Characterized by:

- (Spatially) **local** information on model quality
- Use of **lots** of data (DNS/LES)
- High-dimensional parameter-spaces.
- Usually machine-learning to identify mean-flow => error relationship.

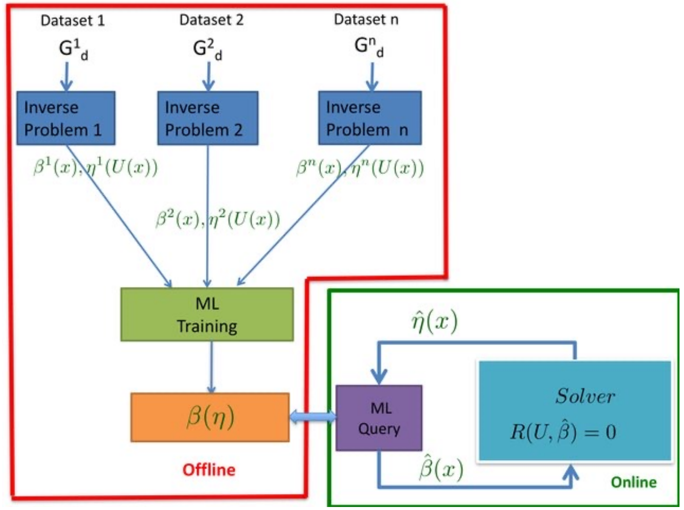
Possibilities:

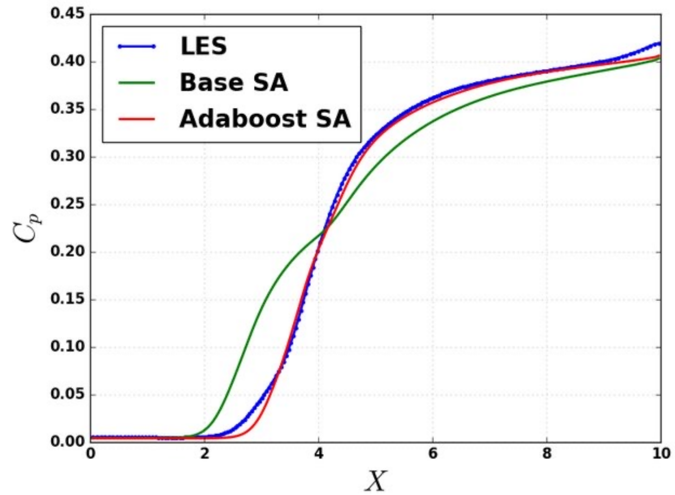
- Calibrate correction terms for a baseline turbulence model (Parish & Durasaimy, JCP 2013; Ling & Templeton, POF 2015, Xiao et al., FTC 2015)
- Calibrate Reynolds stresses by enforcing realizability constraints
- Develop correction terms by means of Genetic Programming (Weatheritt&Sandberg, 2016) or deterministic identification (Schmeltzer et al.,2019)

Field inversion and machine learning

Two-step approach (Duraisamy and coworkers, AIAA 2015-1287, JCP 2016)

- Step I: calibrate a spatially-varying field.
  - Stochastic error field β (defined at every grid-point), modelled as a Gaussian Process
    - E.g. ntroduced as **multiple of production term in k-ω**
    - Validation of Gaussian assumption performed
  - Calibration method: Gradient-based optimization for MAP estimate of β using adjoint
    - And Hessian for covariance
- Step II: Use Machine Learning to identify relationships between resolved/calibrated quantities.
  - Prediction: Use Supervised Machine Learning to convert spatially varying βinto a function of physical input features
  - Features: *Sk/ε*, *d√k/ν*, *P/ε*, y^+ (« flow-specific »)
  - Bias due to the choice of the baseline model?
- Requires a lot of data and efficient high-dimensional optimization methods
- Does not extrapolate very far from the training set
- As hoc choice of learning features
- If the baseline model is Boussinesq, corrected model is Boussinesq





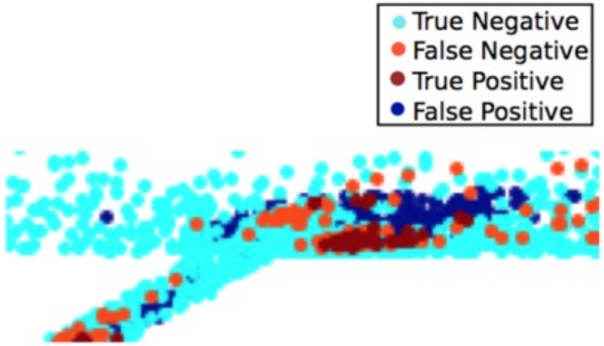
Machine learning of RANS error metrics

One-step approach (Ling and Templeton, POF 2015)

- Goal*:* Automatically identify regions of high-uncertainty in RANS models
- How: Train ML to classify typical error metrics based on mean-flow criteria
- Method: Random forests, Adaboost decision trees.
  - Inputs: 12 non-dimensional criteria based on mean-flow (turbulence Intensity, pressure gradient, streamline curvature …)
  - Output: 3 binary error metrics (eddy-viscosity, anisotropy, non-linearity)
- Data: LES/DNS dataset over 7 cases.
  - Error metrics based on combination of LES/DNS solutions and RANS solution.

Prediction: Ling et al., JCP 2016; Ling & Kutz, JFM 2016

- Pope's **generalized eddy viscosity** approach + deep neural networks or random forests to predict coefficients of the **tensor expansion**



Example result, from Ling et al.: Inclined jet in Crossflow, non-linearity error metric

Calibration of the Reynolds stress tensor

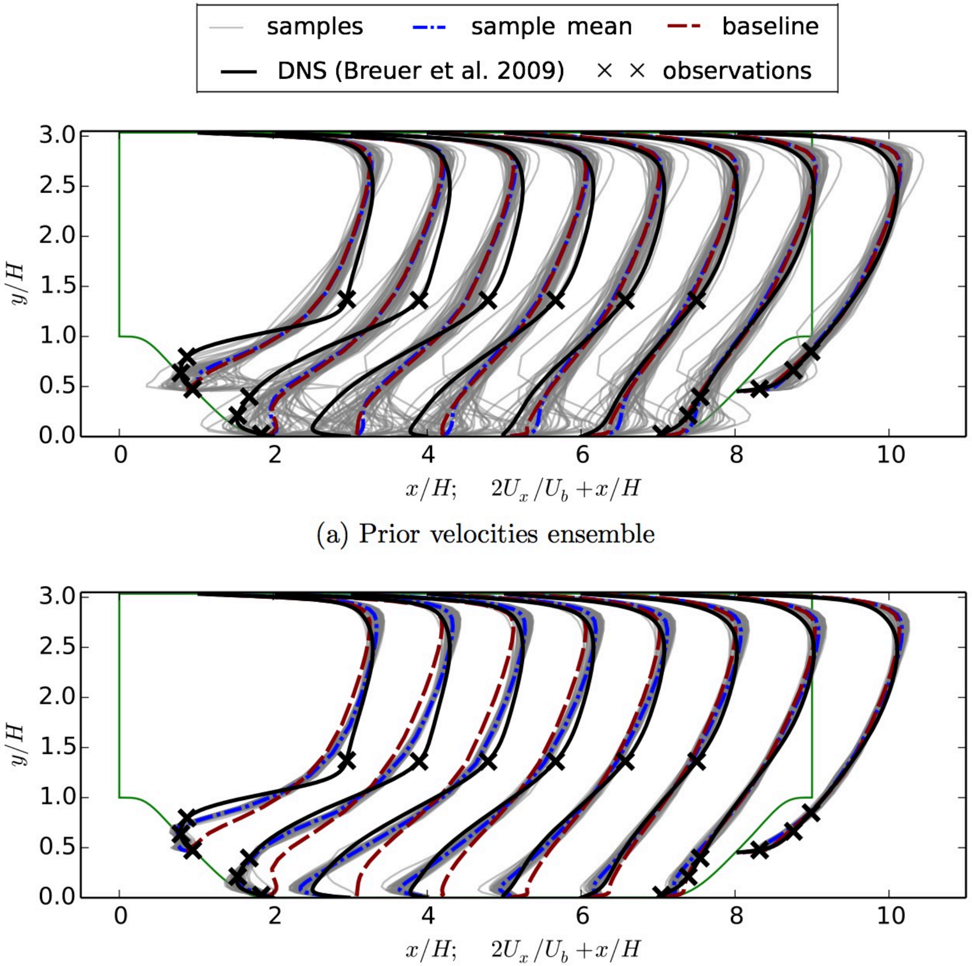
Inverse formulation of the UQ framework of Iaccarino et al. (Xiao et al., series of papers, Flow Turb Comb 2016)

- Emphasis on physical priors: realizability, problem-specific prior-knowledge.
- Parameters*:* Eigenvalues of RST, as function of space (GP via KL-expansion)
  - Positive kinetic-energy is assured
  - Reynolds tensor perturbed around a preliminary RANS solution (bias?)
  - KL used the reduce dimension of stochastic space
- Calibration method: Iterative Ensemble Kalman Method
  - Large number of evaluations of RANS (with different Reynolds stresses)
- Data: Velocity at a small number of locations
  - Periodic-hills and square duct

Calibration of the Reynolds stress tensor

- Prediction: Extrapolation to geometrically identical configuration with different Re or geometrically similar configuration with similar operating conditions

- Demonstrated ability to improve RANS predictions with sparse data



(a) Prior velocities ensemble

(b) Posterior velocities ensemble

Example result, from Xiao&al.: Wall-shear-stress on periodic Hill case – DNS reference

Sparse identification of the Reynolds stress tensor: the SparTA algorithm

[Schmeltzer, Dwight, Cinnella, FTaC 2020]

- SpaRTA = SPArse Regression of Turbulent-stress Anisotropy
  - Open-box machine learning algorithm
  - CFD-free traning
  - Uses a pre-defined library of explicit functions for learning
- Start with linear eddy viscosity model (here, Menter's $k-\omega$ SST)

$$\tau_{ij} = 2k\left(b_{ij} + \frac{1}{3}\delta_{ij}\right); \quad b_{ij} = -\frac{\nu_t}{k}S_{ij} \qquad \nu_t = f(k,\omega)$$

+ transport equations for k and ω

→ Not suitable for flows separation, streamline curvature, strong gradients, etc.

- Internal additive corrections of Reynolds stress anisotropy (b_{ij}^{Δ}) and turbulent transport equations (R):

$$b_{ij} = -\frac{\nu_t}{k}S_{ij} + b_{ij}^{\Delta}$$

- Learn b_{ij}^{Δ} and R from high-fidelity data

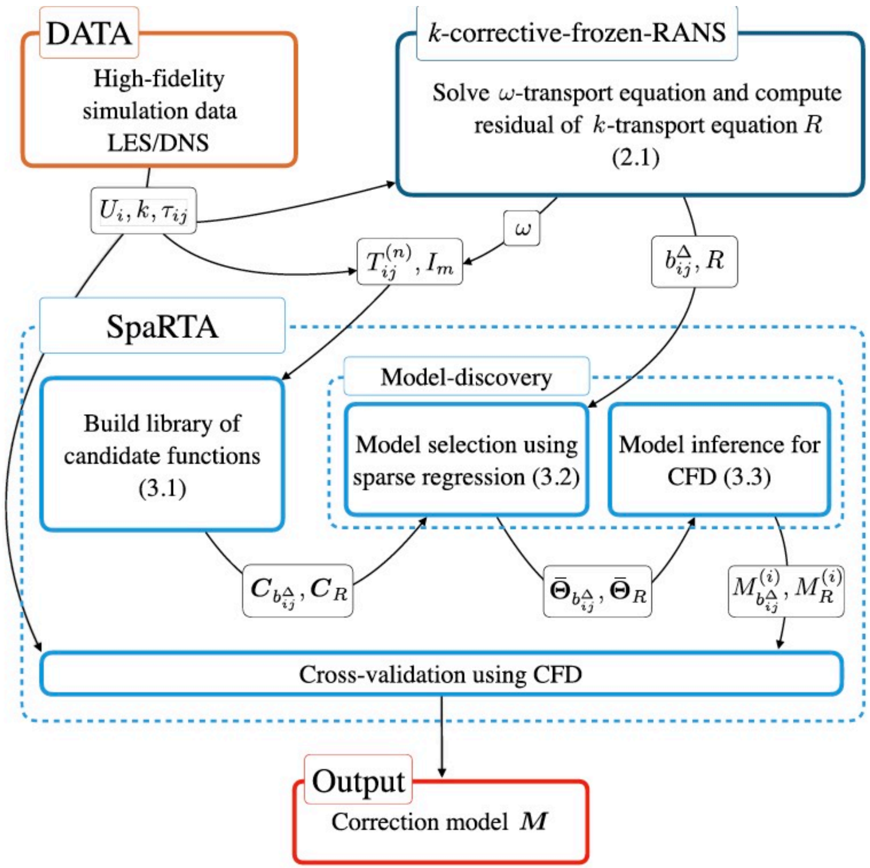
$\boxed{\text{Data-augmented SpaRTA model}}$

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

Sparse identification of the Reynolds stress tensor: the SparTA algorithm

[Schmeltzer, Dwight, Cinnella, FTaC 2020]

- Create a database of ”exact” DNS/LES data for b_{ij}^{Δ} and R
  - Frozen approach: passively solve turbulent equations using high-fidelity mean-flow and Reynolds-stress data
- Discovery step: use sparse elastic-net regression to identify suitable model structures
- Inference step: use ridge regularized least mean square regression to identify coefficients
- Run competing models through the CFD code and select best model



SpaRTA workflow (from Schmeltzer et al.)

Sparse identification of the Reynolds stress tensor: the SparTA algorithm

[Schmeltzer, Dwight, Cinnella, FTaC 2020]

- In practice: we construct b_{ij}^{Δ} by using the effective eddy viscosity approach of Pope (JFM, 1975)
- Assume $\tau_{ij} = \tau_{ij}\left(\frac{\partial \overline{u}_i}{\partial x_j}\right)$ and project τ_{ij} onto a minimal integrity basis:

$$b_{ij}^{\Delta} = \sum_{l=1,\ldots,10} \alpha_l(I_1, I_2, I_3, I_4, I_5) T_{ij}^{l}$$

- For 2D flow, b_{ij} depends on three tensor polynomials of the mean strain rate S_{ij} and rotation Ω_{ij} + 2 invariants $I_1=|S_{ij}|^2$, $I_2=|\Omega_{ij}|^2$:

$$b_{ij}\left(S_{ij},\Omega_{ij}\right) = \alpha_1\left(I_1,I_2\right)S_{ij} + \alpha_2\left(I_1,I_2\right)\left(S_{ik}\Omega_{kj} - \Omega_{ik}S_{kj}\right) + \alpha_3\left(I_1,I_2\right)\left(S_{ik}S_{kj} - \frac{1}{3}\delta_{ij}S_{mn}S_{nm}\right)$$

- Model R as : $R = 2kb_{ij}^{R}\frac{\partial \overline{u}_i}{\partial x_j}$, where is modelled similarly to b_{ij}^{Δ}
- Build libraries of polynomial functions of the invariants

$$\mathcal{B}_l = [C, I_1, I_2, I_1^2, I_2^2, I_1 I_2, \ldots] \quad \text{so that} \quad b_{ij}^{\Delta} = \sum_{l=1,\ldots,10} \Theta \cdot \mathcal{B}_l\, T_{ij}^{l}$$

with Θ a vector of coefficients

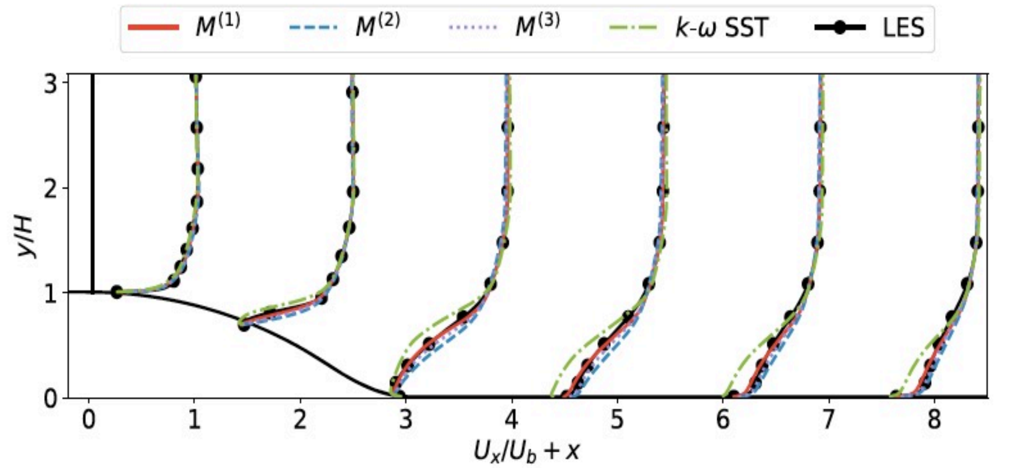
- Find Θ by solving a regularized (elastic net) regression problem

OUTCOME: sparse data-driven Explicit Algebraic Reynolds-Stress Model (EARSM)

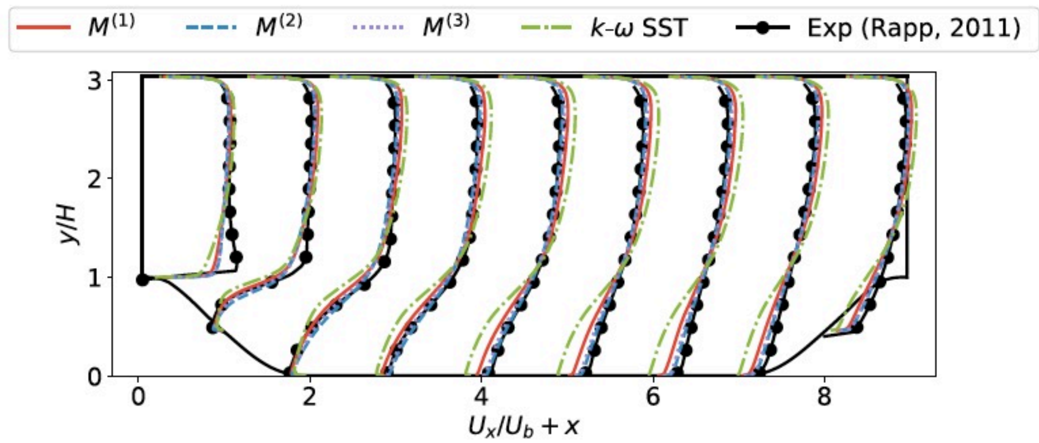
Results

- Model corrections to k-ω SST derived using LES/DNS data for:
  - Periodic 2D-hill flow (PH) at Re=10595 → $M^{(1)}$
  - Converging-diverging channel (CD) at Re=12600 → $M^{(2)}$
  - Curved backward-facing step (CBFS) at Re=13700 → $M^{(3)}$
- Corrections propagated through the OPENFOAM open source CFD solver

Data-driven models (including those trained for PH and CBFS) outperform the baseline for all cases



CBFS results for various models



PH results for various models at Re=37000

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

CFD-driven SparTA algorithm [Ben Hassan-Saidi, Cinnella, Grasso 2020]

- Plug generic model into the CFD solver
- Collect high-fidelity data for any QoI (e.g., velocity)
- Find coefficients by solving the optimization problem :

$$\Theta^* = \operatorname{argmin}_\Theta \left\| U_{Sparta}(\Theta) - U_{HF} \right\| + \lambda \|\Theta\|_1 + 0.5\lambda(1-\rho)\|\Theta\|_2$$

- Preliminary local sensitivity analysis for reducing problem dimensionality –> 12 parameters
- Enforcement of realizability constraints
- Optimization based on blackbox python library : CORS algorithm (constrained optimization using response surfaces)
  - Cubic radial basis function surrogate + resampling
  - Candidate samples preventing the CFD solver to converge are discarded and resampled

OUTCOME: data-driven Explicit Algebraic Reynolds-Stress Model (EARSM)

- Only one step needed (simultaneous discovery and inference)

SCIENCES SORBONNE UNIVERSITÉ ∂'Alembert Institut Jean le Rond d'Alembert

Results

- PH flow at Re=37000 (out of training set)

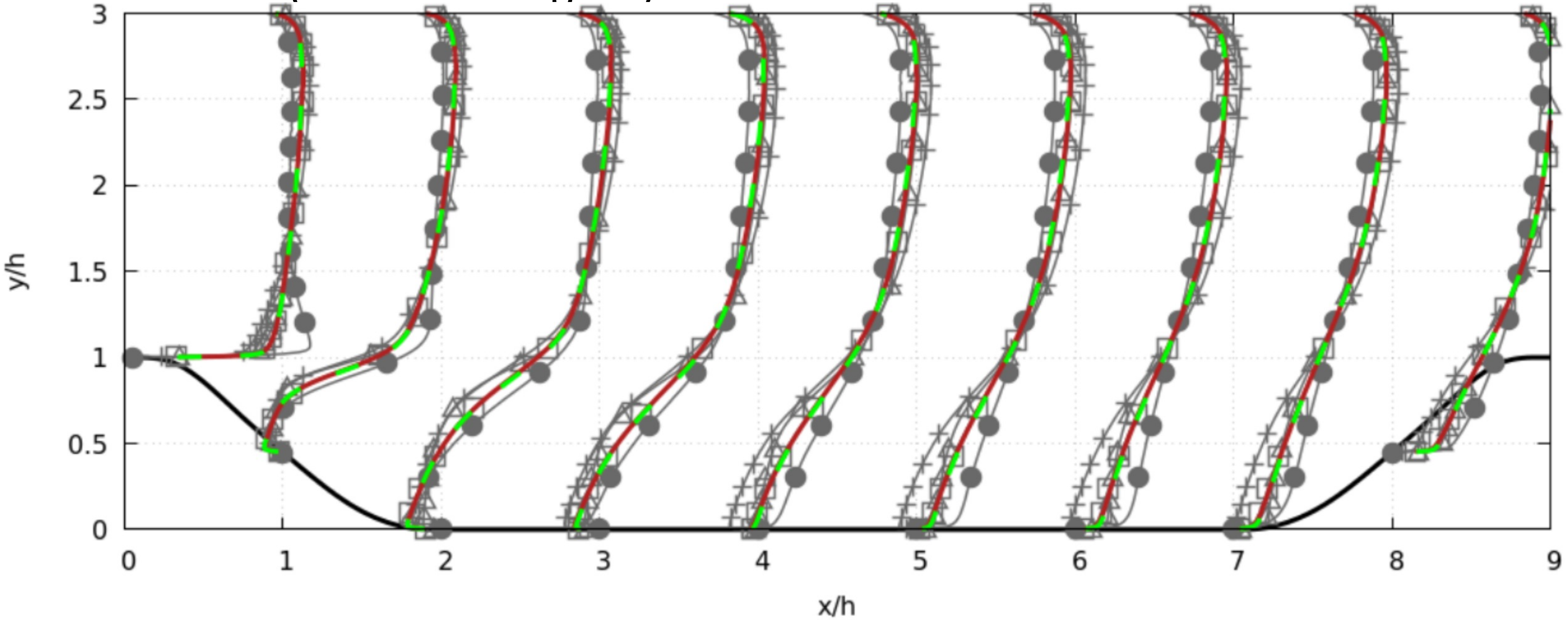


Figure 13: Longitudinal velocity profiles normalized with bulk velocity : $\overline{u_1}/U_b + x$ for PH_{37000}. k-ω-SST: (—+—), High fidelity data : (—●—), Model 1 : (——), Model 2 : (- - -), frozen-training: (—⊟—), BSL-EARSM : (—△—).

- CFD-driven SpaRTA outperforms the baseline and delivers results comparable to CFD-free SpaRTA

Récapitulatif

- Model discovery by learning from data represents an attractive opportunity for developing improved RANS models, customized for reproducing classes of flows
  - Encouraging results obtained for a variety of 2D flows, including massively separated flows and turbomachinery flows
  - Work needed for better improving the algorithms and reducing computational cost
- Bayesian inference provides a systematic framework for
  - updating coefficients associated to turbulence models,
  - selecting or averaging (BMSA) concurrent models
  - Providing estimates of confidence intervals