Data-driven discovery and uncertainty quantification of turbulence models for Fluid Dynamics

Paola CINNELLA (paola.cinnella@sorbonne-universite.fr)

in collaboration with: I. Ben Hassan Saidi, IAT; G. Dergham, Safran Tech; M. De Zordo-Banliat, Safran Tech/ENSAM DynFluid; R. Dwight, TU Delft, NL; W.N. Edeling, CWI, NL; X. Merle, DynFluid; M. Schmeltzer, TU Delft, NL

2e Rencontres Chercheurs/Ingénieurs « Saisir le mouvement », 25 novembre 2020
Décrire et inférer la dynamique des Systèmes

SCIENCES SORBONNE UNIVERSITÉ

∂'Alembert
Institut Jean le Rond d'Alembert

Talk overview

- Turbulence modeling: why and how
- Using data for predicting turbulence?
- Quantifying modeling uncertainties
- Conclusions and future trends

Talk overview

- Turbulence modeling: why and how
- Using data for predicting turbulence?
- Quantifying modeling uncertainties
- Conclusions and future trends

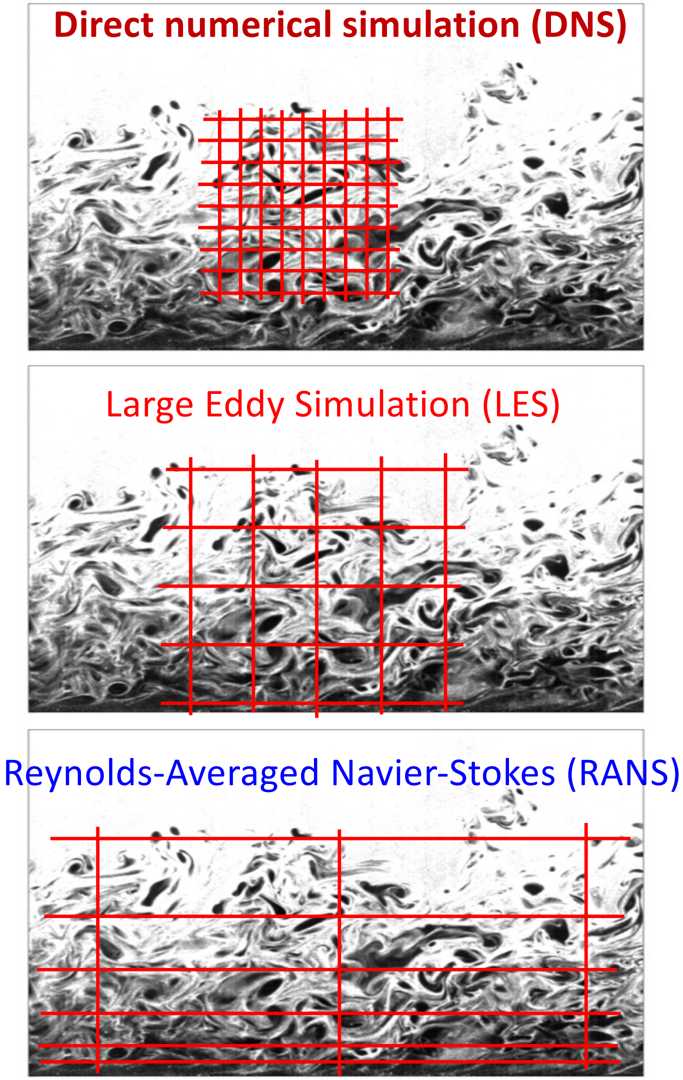
Turbulence modeling: why

- Turbulent flows omnipresent in Engineering sciences
  - Wide range of spatial and temporal scales
  - Key parameter: Reynolds number (ratio of inertia to viscous forces)
- **Navier-Stokes equations** contain all the necessary information:
  - **DNS** (computationally intractable for most practical cases)
  - Alternatives: hierarchy of approximations, depending the amount of resolved vs modelled scales

LES→ (RANS/LES)→**RANS**

- More resolved scales → higher cost (especially for wall-bounded flow) and sensitivity to numerical errors & boundary conditions
  - Not suitable for routine use in industry
- More modeled scales → lower cost, more flow-dependent, and uncertain **turbulence models**

RANS: workhorse for CFD simulations in engineering



Visualization of turbulent boundary layer (Mach 3.0) via nano-tracer-based planar laser scattering [Ding et al., 2018]

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

Turbulence modeling: how

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

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

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

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

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

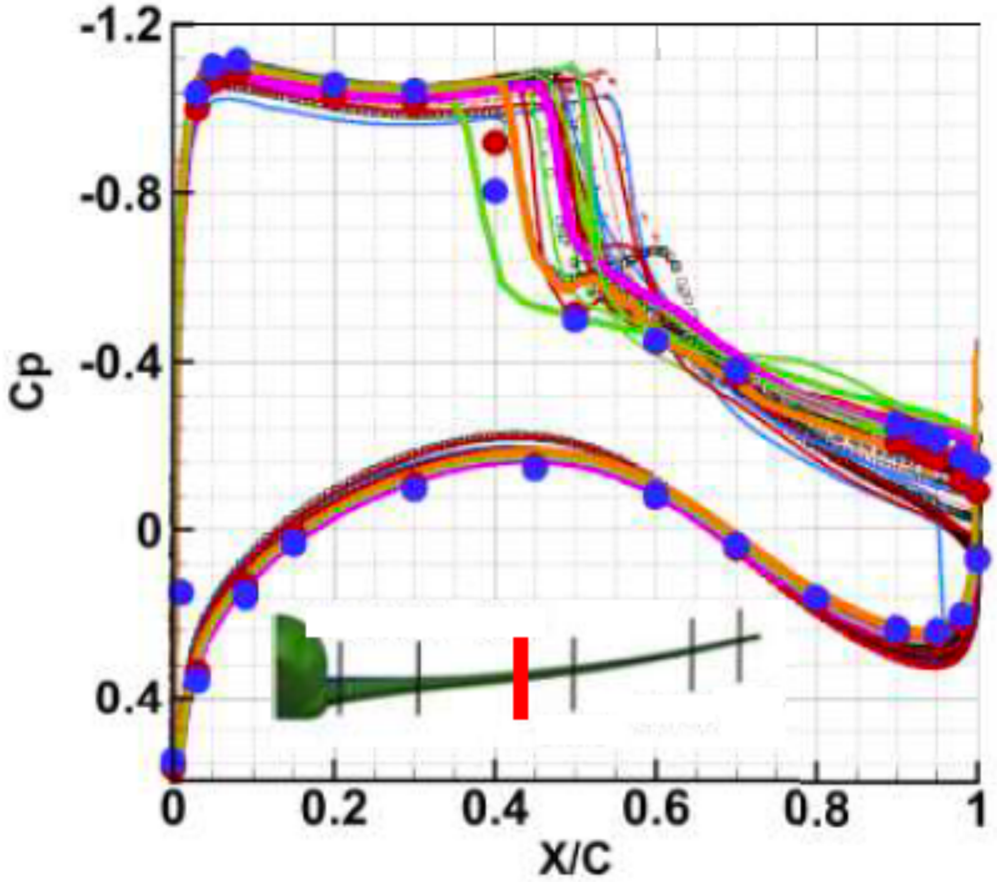
Reynolds stresses need a constitutive law: a **turbulence model**

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

Turbulence modeling: how

- **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 prediction workshop.

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

Talk overview

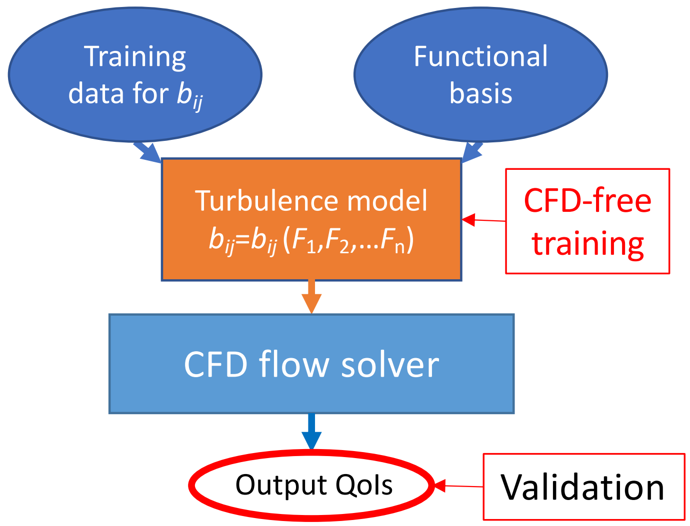
- Turbulence modeling: why and how
- Using data for predicting turbulence?
- Quantifying modeling uncertainties
- Conclusions and future trends

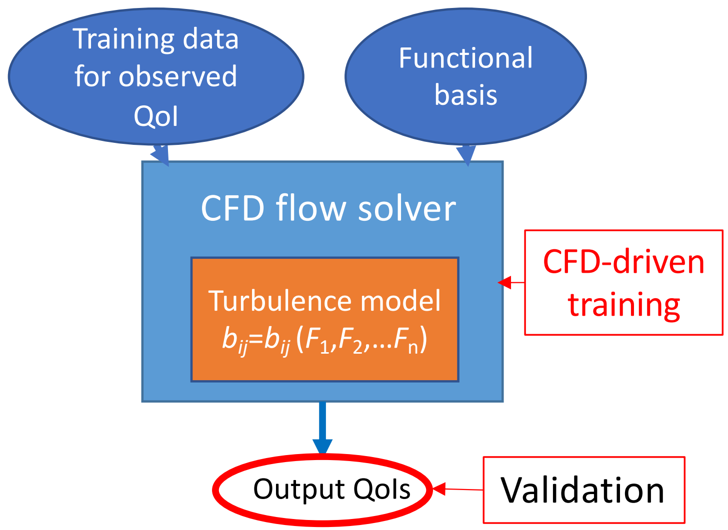
Using data for predicting turbulence?

- 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 model discovery: the SparTA algorithm [Schmeltzer, Dwight, Cinnella, FTaC 2020]

- SpaRTA = SPArse Regression of Turbulent-stress Anisotropy
  - Open-box machine learning algorithm
  - CFD-free training
  - 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):

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

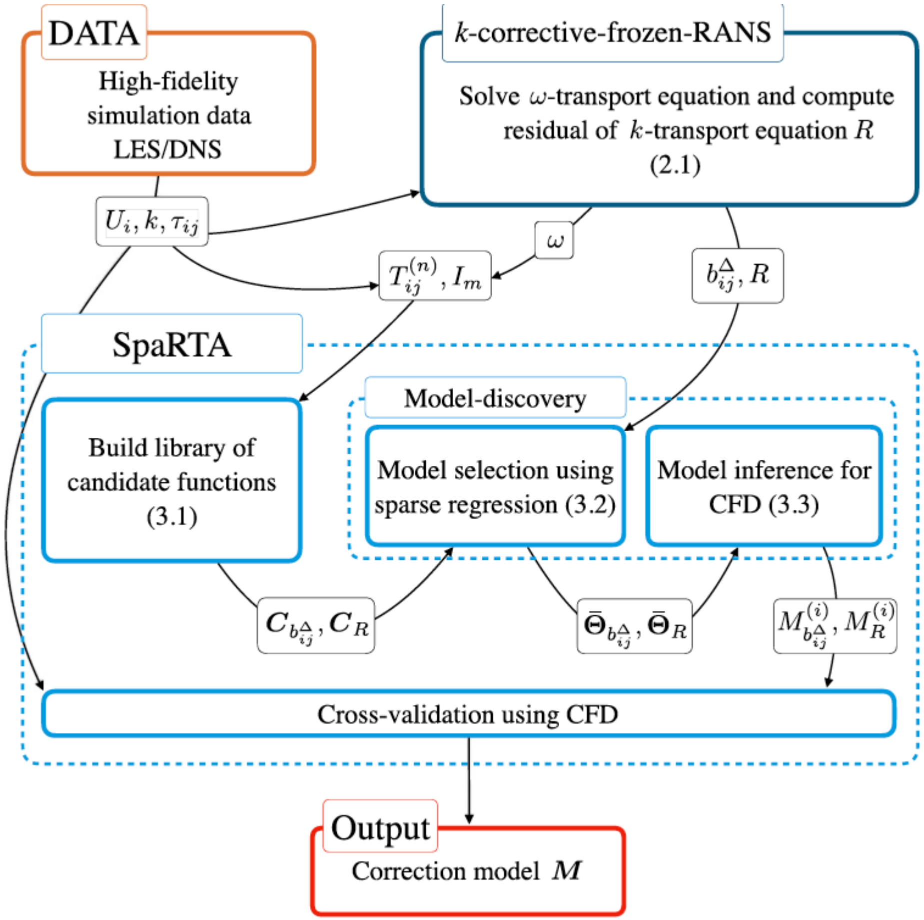
- Learn b_{ij}^{Δ} and R from high-fidelity data

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

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

Data-driven model discovery: 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.)

Data-driven model discovery: 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 $\tau_{ij} = \tau_{ij}\left(\frac{\partial \bar{u}_i}{\partial x_j}\right)$ and project τ_{ij} onto a minimal integrity basis:

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

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

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

  with Θ a vector of coefficients

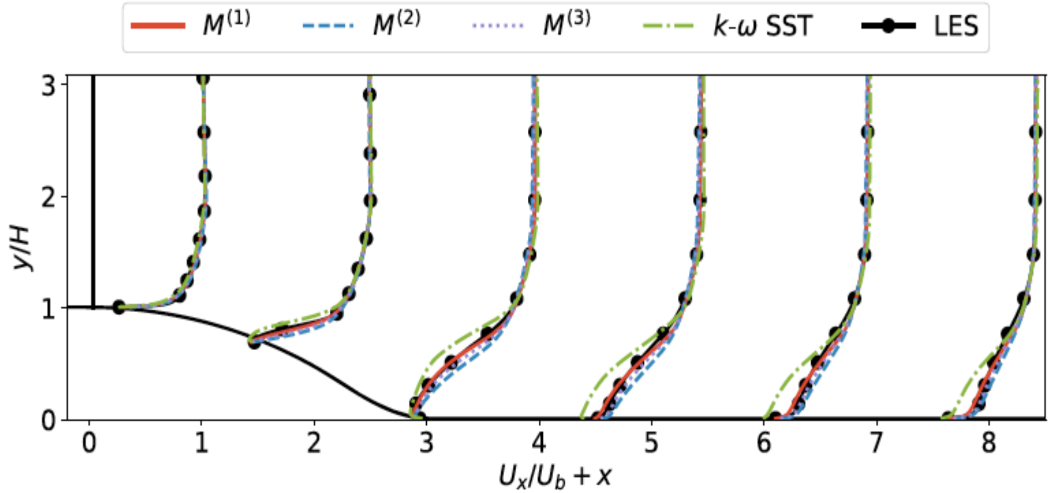
- Find Θ by solving a regularized (elastic net) regression problem

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

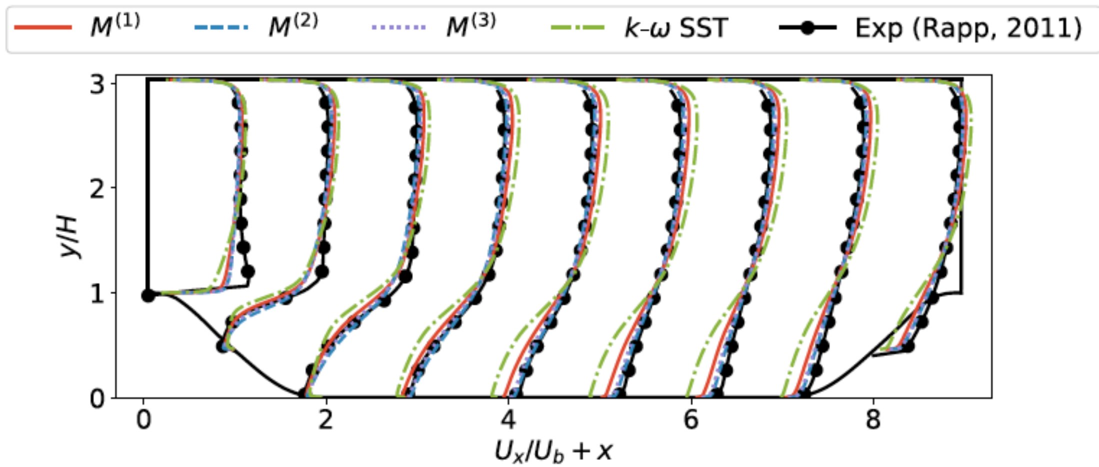
Results

- 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 :

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

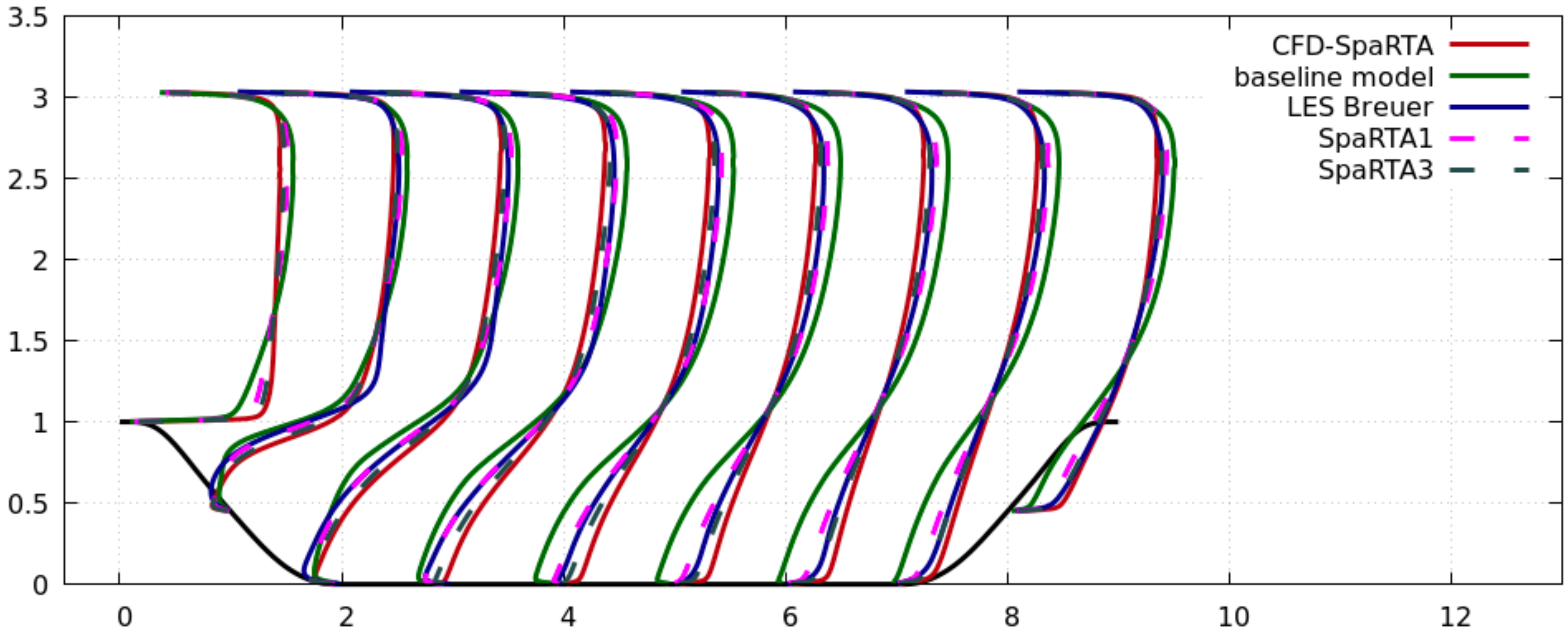
- Preliminary local sensitivity analysis for reducing problem dimensionality –> 12 parameters
- Enforcement of realizability constraints
- Optimization based on blackbox python library : CORS algorithm (constrained optimization using response surfaces)
  - Cubic radial basis function surrogate + resampling
  - 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)

Results

- Preliminary results for the PH flow at Re=10595
- Comparison with baseline model and CFD-free SpaRTA models. Optimization based on 75 CFD runs.



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

Talk overview

- Turbulence modeling: why and how
- Using data for predicting turbulence?
- Quantifying modeling uncertainties
- Conclusions and future trends

Quantifying modeling uncertainties : Bayesian model averaging

- Model identification procedures generally <u>deterministic</u>
  - No estimate of predictive uncertainties provided
- Use of Bayesian statistics to infer on model coefficient posterior distributions
  - **Bayesian inference** pictures the relation between data, *a priori* knowledge and updated knowledge of the coefficients

$$\boldsymbol{p}(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{\boldsymbol{p}(\boldsymbol{D}|\boldsymbol{\theta})}{\boldsymbol{p}(\boldsymbol{D})}\,\boldsymbol{p}(\boldsymbol{\theta})$$

- Well-informed posteriors are peaked → Maximum A Posteriori (MAP) approximation of the posteriors

- Bayesian model and scenario averaging (**BMSA**) to account for uncertainties in the choice of model structure and of training scenarii (geometry and flow conditions)

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

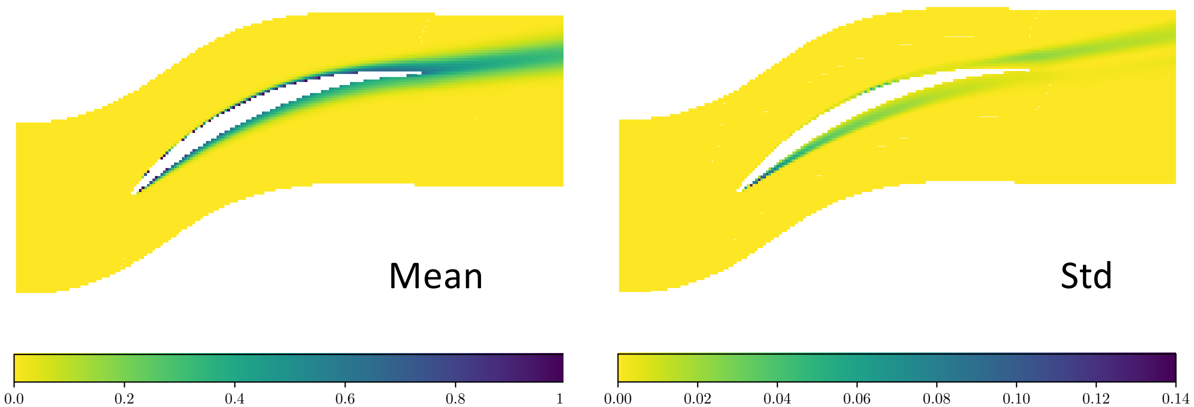
with $\mathcal{M} = (M_1, M_2, \ldots, M_N)$, $\mathcal{S} = (S_1, S_2, \ldots, S_K)$ a set of concurrent models and scenario, respectively.

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

BMSA: Flow through a compressor cascade

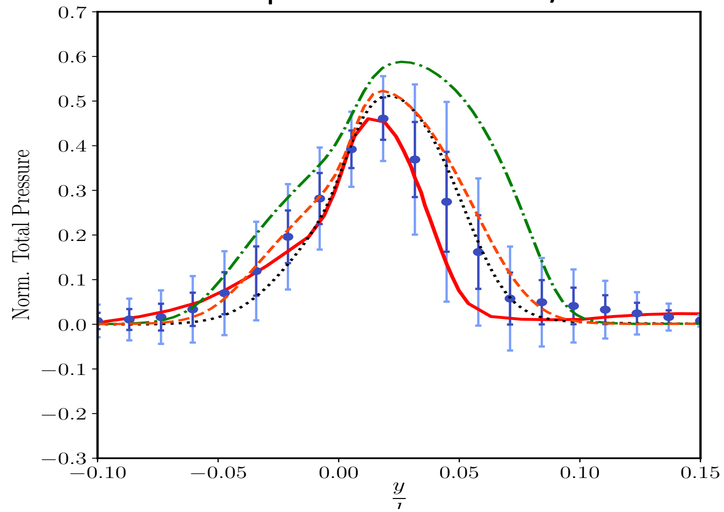
- Prediction of compressible flow through a compressor cascade (NACA65 V103) at off design conditions
- Results based on three models ($k-\omega$ Wilcox, $k-\varepsilon$ Launder-Sharma & Spalart-Allmaras)
- Propagation of the 13 boundary layer MAP estimates AND of 3 MAP estimates calibrated against LES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

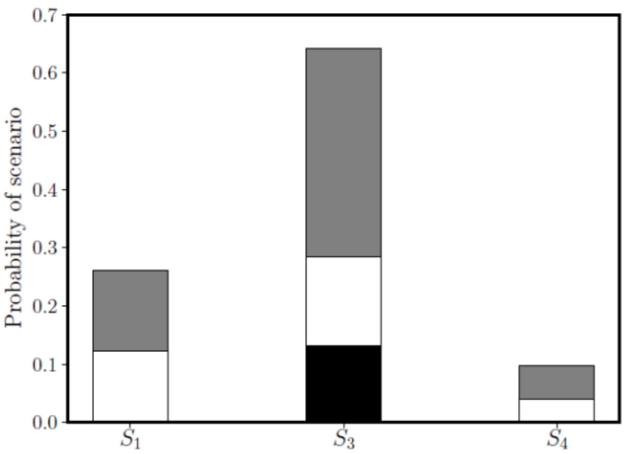


LES data from *Leggett et al.* [1] (—),
$E[\Delta|S']\pm\sqrt{Var[\Delta|S']}$ (—●—), $E[\Delta|S']\pm 2\sqrt{Var[\Delta|S']}$ (—●—),
$k-\omega$ (······),
Spalart–Allmaras (- - -)
$k-\varepsilon$ (-·-·-).

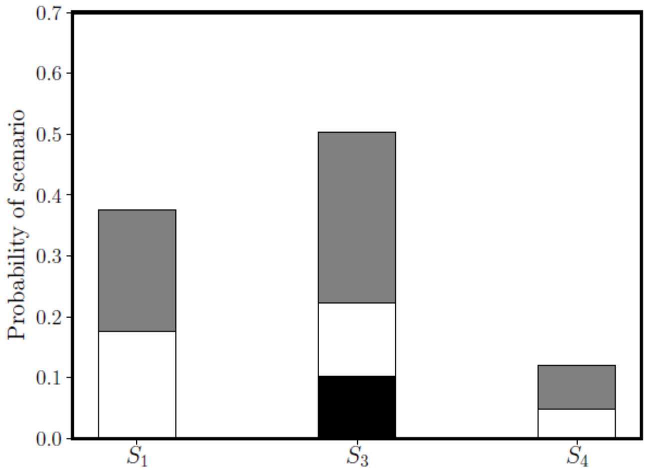
BMSA: Flow through a compressor cascade

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

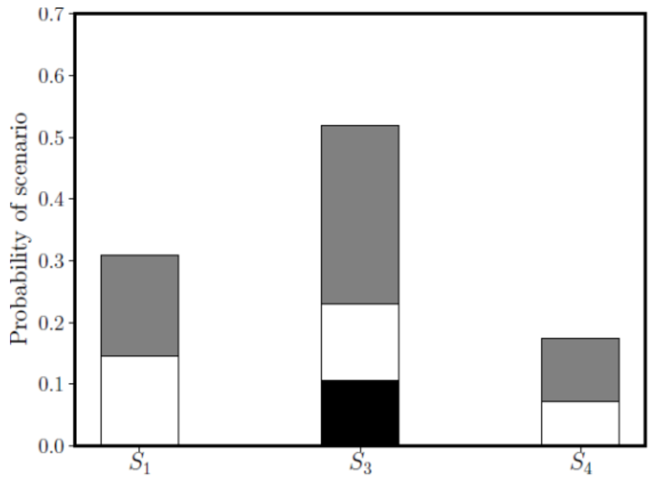
>Consensus based criterion



>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)

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

Space-dependent Bayesian Model Averaging

- BMSA uses the same weights throughout the flow field → contrary to expert judgment
- Further progress: compute $\boldsymbol{P(M_i|D)}$ as a function of space
  - Infer model probabilities for each flow region
  - Identify the “best” model (if any) in each region
- Clustered Bayesian Averaging [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

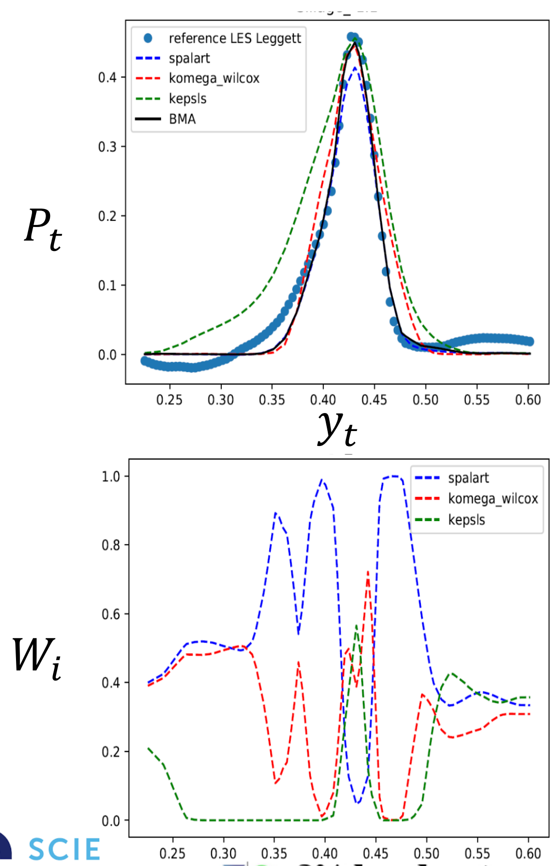
$$y_{final} = \sum_j w_j \, y_j$$

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

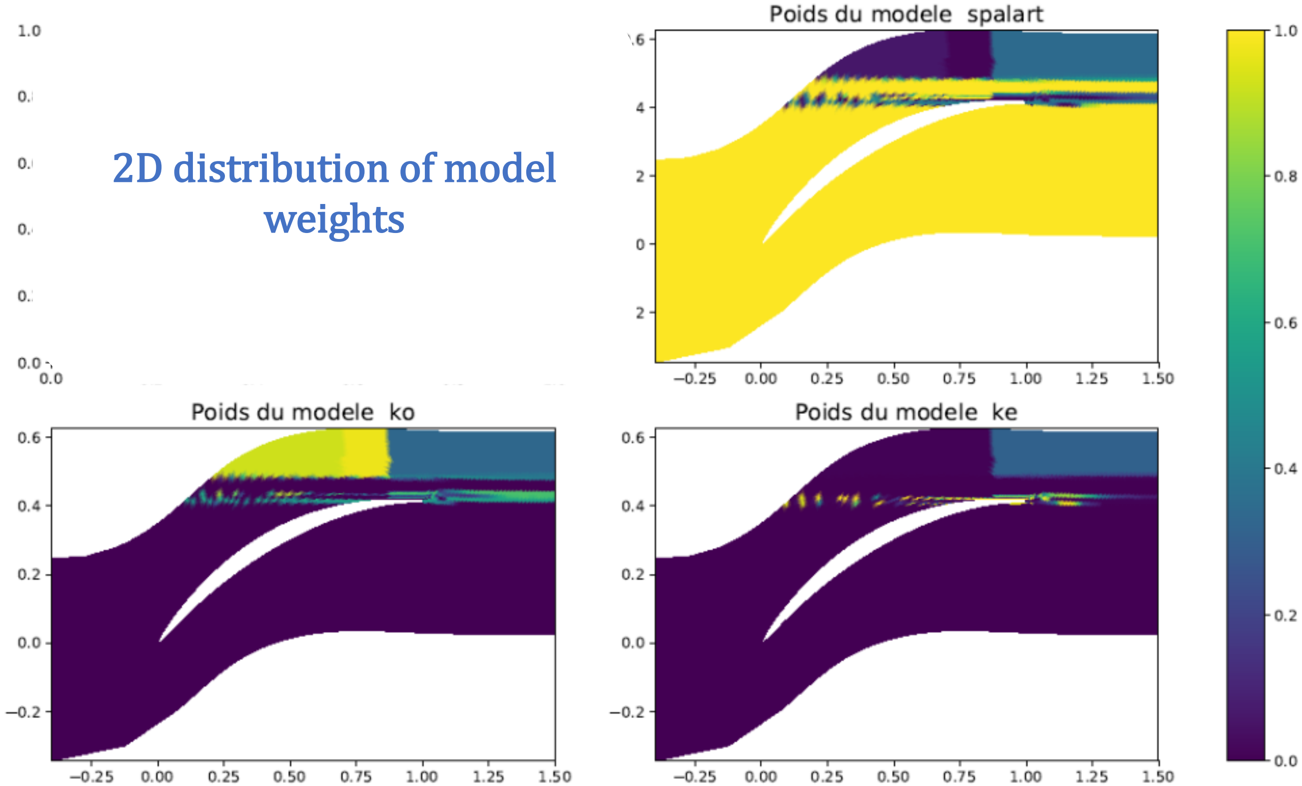
CBMA: preliminary results

- CBMA of $k-\omega$, $k-\epsilon$ & Spalart-Allmaras as models. LES as reference data

  Data: 300 points **CBMA**: 1000 trees



2D distribution of model weights



Conclusions

- 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
- Work in progress:
  - Bayesian formulation of SpaRTA
  - Combination of concurrent SpaRTA models via BMSA and/or CBMA

Questions?

[1] Edeling, W.N., Cinnella, P., Dwight, R., Bijl, H., 2014. Bayesian estimates of parameter variability in the k ε turbulence model. J Comp Phys 258 : 73-94.

[2] Edeling, W.N., Cinnella, P., Dwight, R., 2014. Predictive RANS simulations via Bayesian Model-Scenario Averaging. J Comp Phys 275 : 65-91.

[3] Edeling, W.N., Schmeltzer, M., Dwight, R., Cinnella, P., 2018. Bayesian predictions of RANS uncertainties using MAP estimates. AIAA J. 56(5):2018-2029

[4] Edeling, W.N., Iaccarino, G., Cinnella, P., 2017. Data-Free and Data-Driven RANS Predictions with Quantified Uncertainty. Flow Turb & Comb 100(3):593-616

[6] Xiao, H., Cinnella, P., 2019. Quantification of model uncertainty in RANS simulations: a review. Progress in Aerospace Sciences 108: 1–31.n

[5] Schmeltzer, M., Dwight, R., Cinnella, P., 2020. Discovery of Algebraic Reynolds-Stress models Using Sparse Symbolic Regression. Flow Turb & Comb 104:579-603.

[7] De Zordo-Banliat M., Merle X., Dergham X., Cinnella P., 2020. Bayesian model-scenario averaged predictions of compressor cascade flows under uncertain turbulence models. Comp Fluid 201:104473.

Bayesian model-scenario averaging (BMSA)

- Let M_i be a model structure in set M of m models
- Be z_j a calibration dataset taken in a training set Z of s calibration scenarios
- BMSA prediction of the expectancy of QoI Δ for a new scenario :

$$E\left[\Delta \mid Z\right]=\sum_{i=1}^{m}\sum_{j=1}^{s}E\left[\Delta \mid z_j, M_i\right]P\left(M_i \mid z_j\right)P\left(z_j\right)$$

The scenario of Δ is **<u>NOT</u>** in the calibration set Z

$$E\left[\Delta \mid z_j, M_i\right]$$

is the expectancy of Δ for the new scenario, under model M_i calibrated on dataset z_j

Bayesian model-scenario averaging (BMSA)

- Similarly, the variance of Δ may be written as:

$$\underbrace{\mathrm{var}\left[\Delta \,|\, Z\right] = \sum_{i=1}^{m}\sum_{j=1}^{s} \mathrm{var}\left[\Delta \,|\, z_j, M_i\right] P\left(M_i \,|\, z_j\right) P\left(z_j\right) +}_{\text{In-model, in-scenario variance}}$$

$$\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{s}\left(E\left[\Delta \,|\, z_j, M_i\right] - E\left[\Delta \,|\, z_j\right]\right)^2 P\left(M_i \,|\, z_j\right) P\left(z_j\right) +}_{\text{Between-model, in-scenario variance (model error)}}$$

$$\underbrace{\sum_{j=1}^{s}\left(E\left[\Delta \,|\, z_j\right] - E\left[\Delta \,|\, Z\right]\right)^2 P\left(z_j\right)}_{\text{Between-scenario variance (spread)}}$$