LIFD

Workshop: Data Driven Methods in Fluid Dynamics

Leeds Institute for Fluid Dynamics, March 30th-31st, 2023

Turbulence modelling: artificial vs human intelligence

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

SCIENCES SORBONNE UNIVERSITÉ

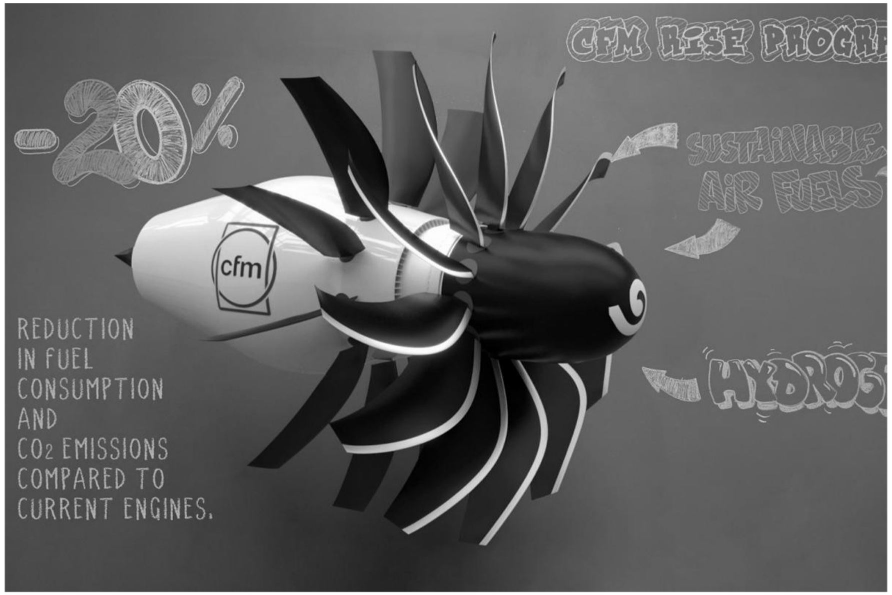
∂'Alembert Institut Jean le Rond d'Alembert

Overview

- Turbulence modeling:
  known facts, uncertainties, and the potential for machine learning
- Sparse Bayesian learning for data-driven turbulence modeling
- Generalization through model mixtures
- Conclusions and outlook

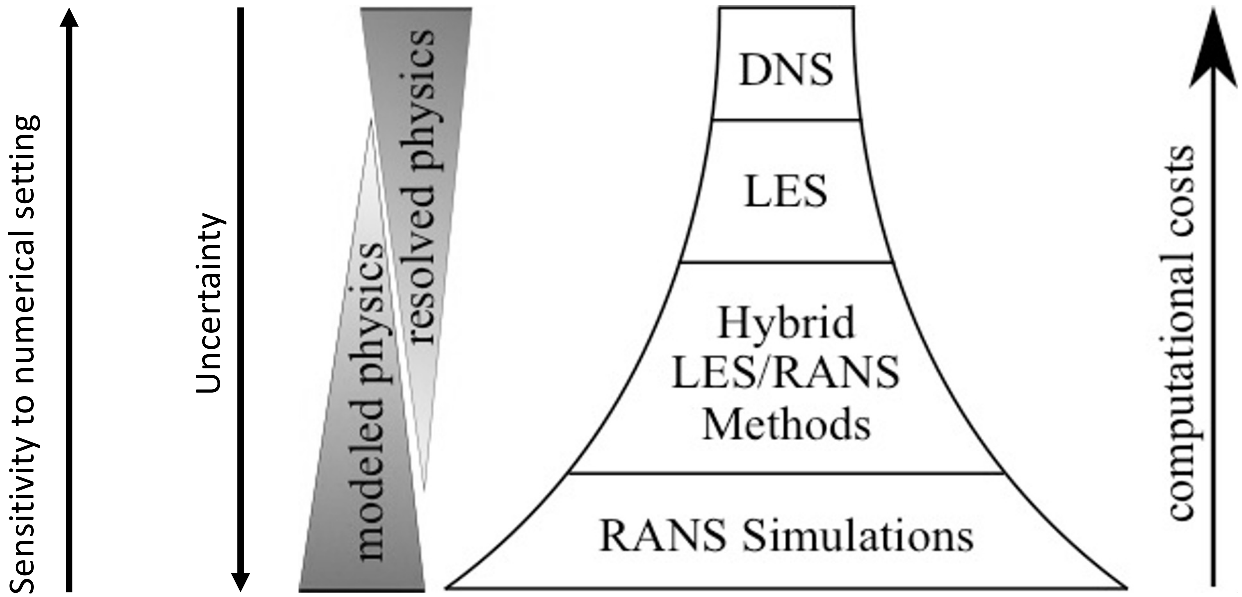
Motivation

- Increasing pressure for breakthrough designs: new fuels, extreme operating conditions
- Environmental constraints require drastically reduced energy consumption and carbon emissions
- Intensive use of **high-fidelity Computational Fluid Dynamics** (CFD) to explore uncharted designs or extreme operating conditions → ”**virtual rig**”



Motivation

- Flows of engineering interest: high **Reynolds** number
  - Large range of scales
  - Accurate numerical simulations require **billions of degrees of freedom, huge computational cost**
  - Low-fidelity RANS (or lower) models **not reliable** for configurations **off the beaten paths**



The closure problem

- 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

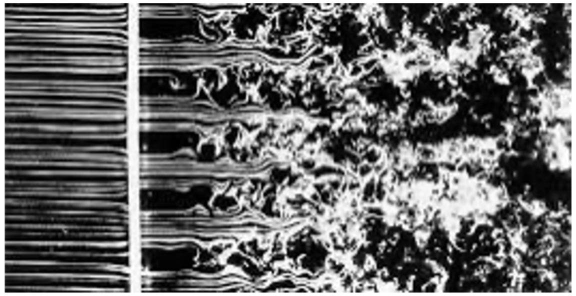
Reynolds stresses need a constitutive law: a **turbulence model**

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

$$\boxed{\tau = \tau(\bar{u}, \bar{p}; \theta)}$$

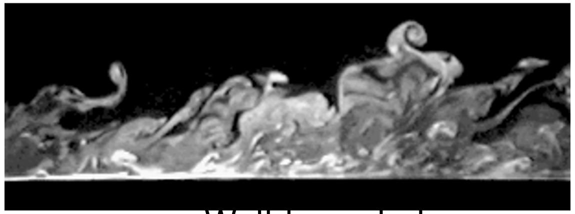
Known facts [P. Spalart lectures, July 2022, NASA Turbulence modeling Symposium]

- Classical turbulence models are already “data-driven”, to some extent
  - **Transport-equation model ingredients**:
  1. Constitutive Relation: combination of turbulence “scales” and velocity gradient
  2. Transport equations (imitating the exact equation for the Reynolds stresses)
     - Production + Dissipation + Diffusion
  3. Damping functions active in the viscous sublayer
  4. Tuning coefficients (not all independent)
- Fine-tuned to capture a few “turbulence facts”
  - Mostly the flat-plate boundary layer



Decaying grid turbulence

Free shear



Wall-bounded

Turbulence modelling uncertainties

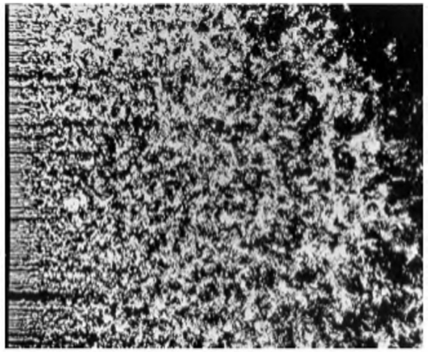
Quest of the Holy Grail: the "universal" RANS model

- **Model structure** based on first principle
  - Constraints: objectivity, symmetries, realizability
  - "Exact" equations from high-order moments of NS

    → 800-pound gorilla: the **closure problem**!
  - End up with **crude modeling assumptions**
  - **Closure coefficients are not carved in stone**!

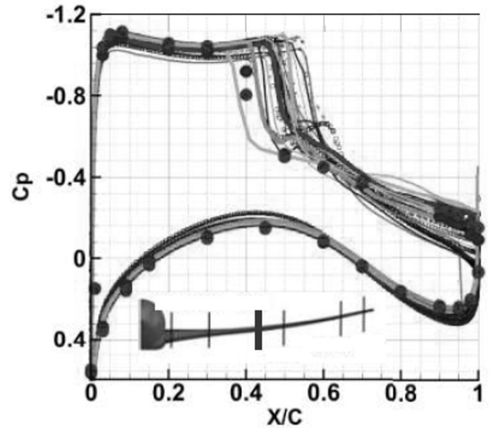
  $C_{\varepsilon 2} = 1.92; 1.68; 1.83\ldots$; best-fit to data: 1.77!

No universally accepted model, no universal parameters, variable performance



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

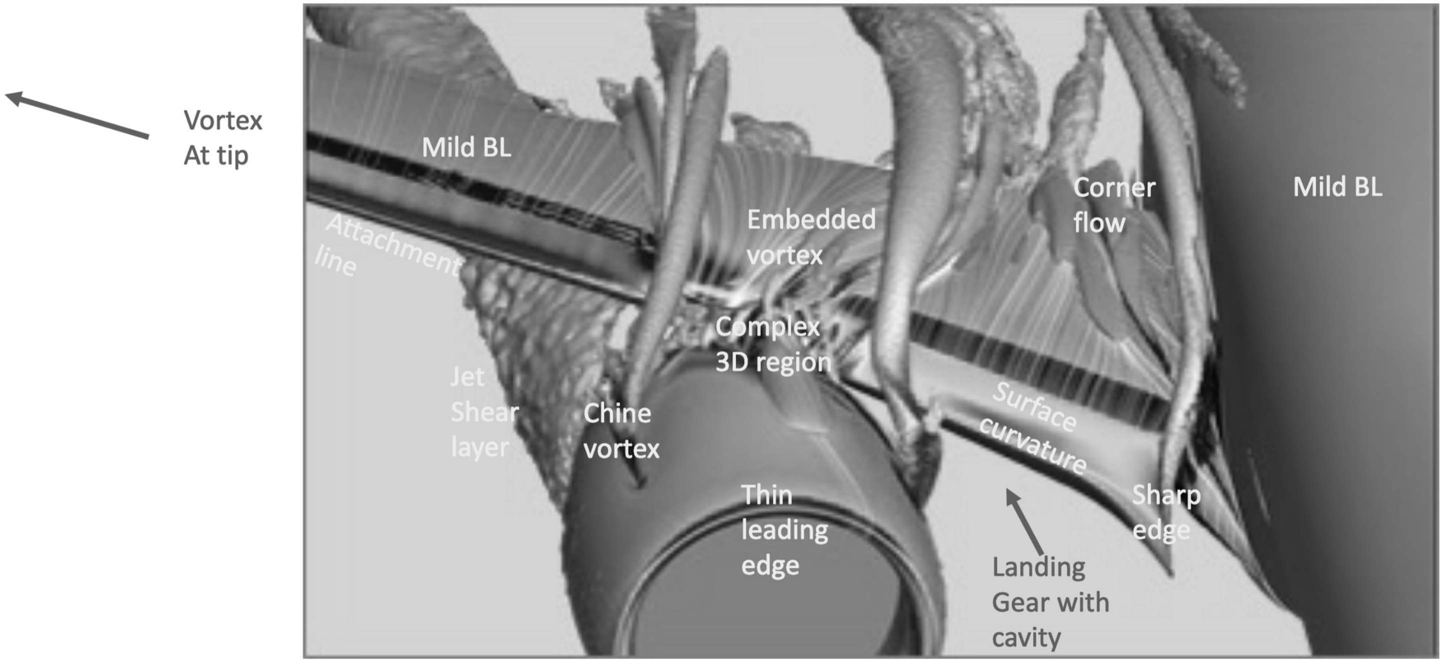
Decay rate of homogeneous isotropic turbulence



<u>Blind</u> pressure coefficient predictions from various RANS models (lines) and experiments (symbols).
6th AIAA Drag prediction workshop.

Quest for the “universal” model

- Some degree of generality needed
- Hand set “zonal” models not acceptable for industry

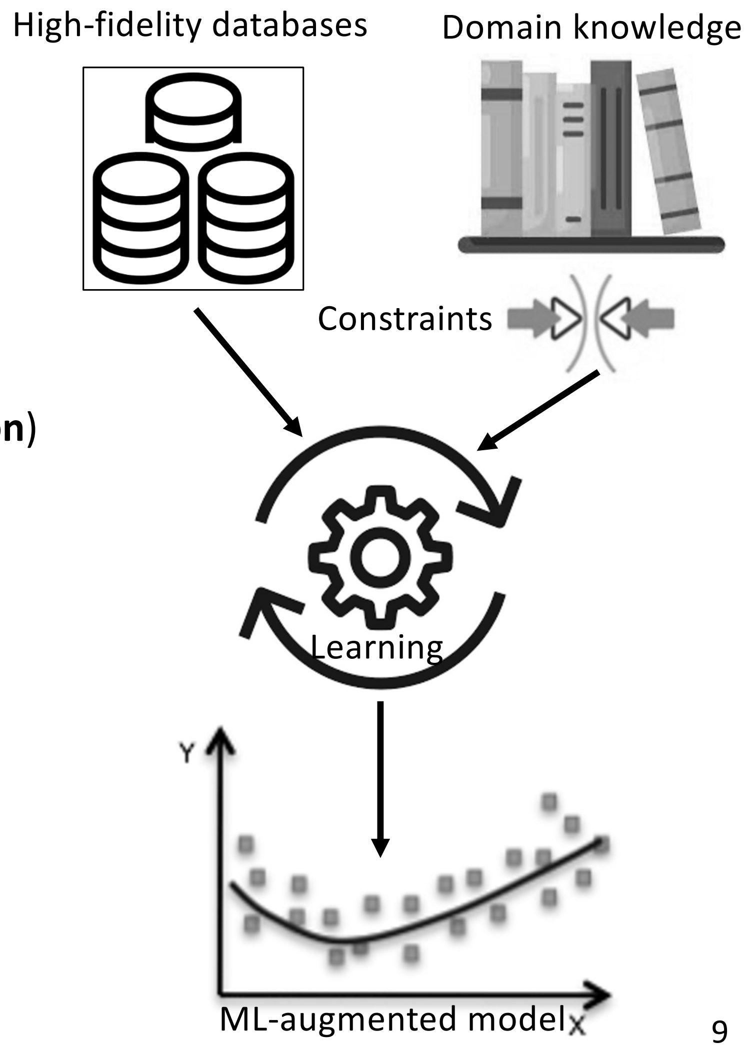


[from P. Spalart lectures, July 2022, NASA Turbulence modeling Symposium]

The potential for machine learning

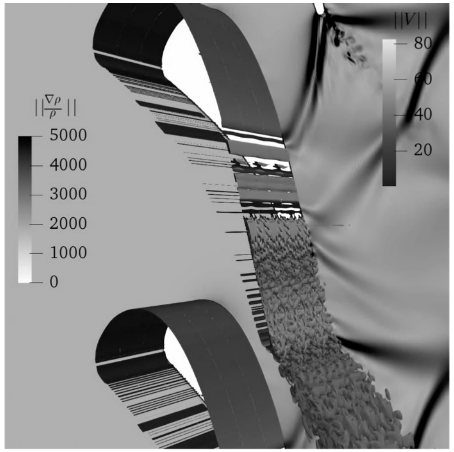
Machine learning of data-driven turbulence models

- General philosophy
  - No « universal » model
  - **Customized models ("experts") for flow classes**
  - Mostly formulated as a "correction" (data-driven **augmentation**)
    1. Collect data
    2. Choose a functional basis
    3. **Enforce physical constraints** (whenever possible)
    4. Train against **data**
    5. Validate for a test set

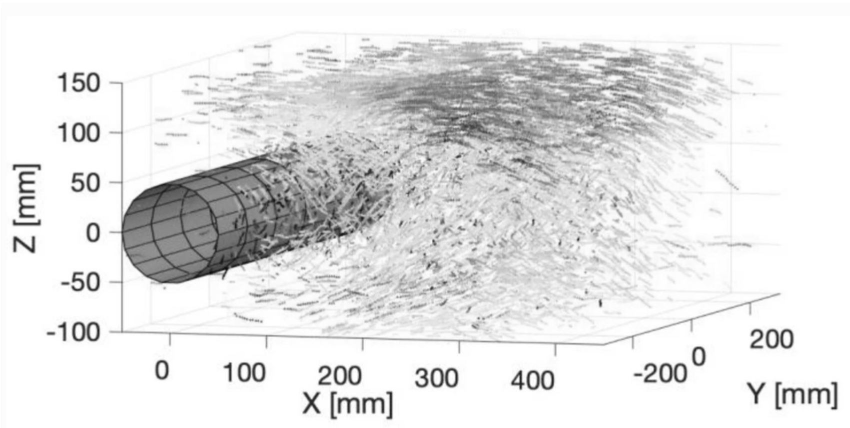


Challenges

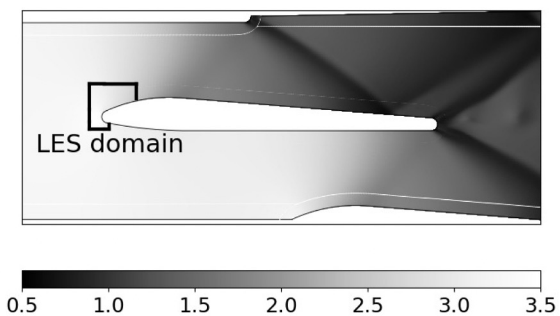
- Large, but **scarce** and extremely **costly** data bases
- HiFi-CFD limited to low/moderate-Reynolds numbers
- Experimental databases: higher-Re, but **noisy**/incomplete
- **ML out-of-sample performance**?
- **ML uncertainty quantification**?

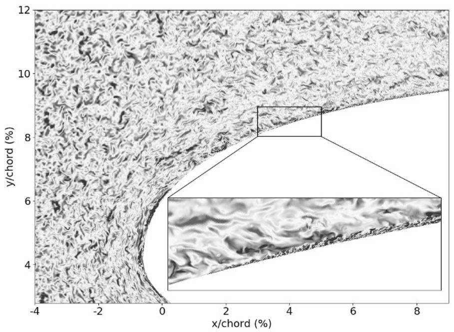


Transonic flow of a non-ideal gas (fluorocarbon PP11) through a linear cascade, coarse LES, Re= 8×10^5, 3×10^7 *points,* ~10^5 CPUh *(Hoarau et al., 2021)*



Time-resolved tomographic PIV of incompressible flow past a cylinder at Re_D*=27000 (Scarano et al., 2022)*





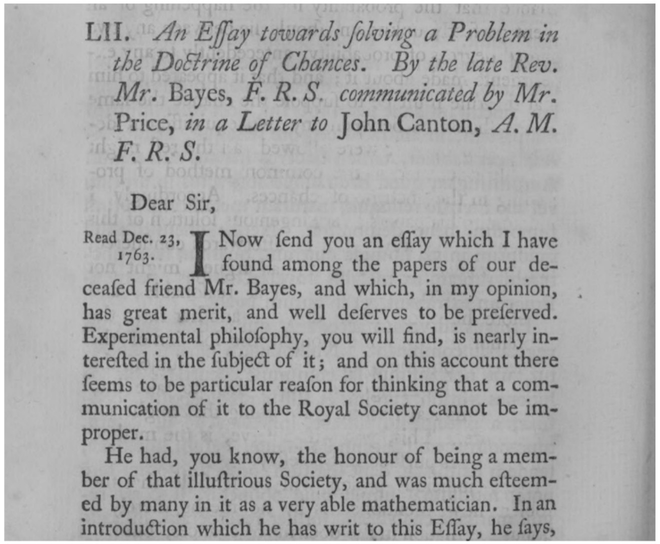
Free-stream transition of a non-ideal gas boundary layer (Novec649 advanced, ozone-friendly working fluid) around a turbine leading edge at Mach 0.9, Re_{L11}*~11000,* 2.15×10^9 *points, CFL~5, ~*10^6 *CPUh (Bienner et al., 2022)*

Strategy

- Large, but scarce and extremely costly data bases
  - Parsimonious approaches
- HiFi-CFD limited to low/moderate-Reynolds numbers
- Experimental databases: higher-Re, but noisy/incomplete
  - CFD-in-the-loop training, data assimilation
- Out-of-sample performance?
  - Model mixtures
- Uncertainty quantification?
  - Stochastic Machine Learning

Bayesian Machine Learning

Rev'd Thomas Bayes (1702-1761)



LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

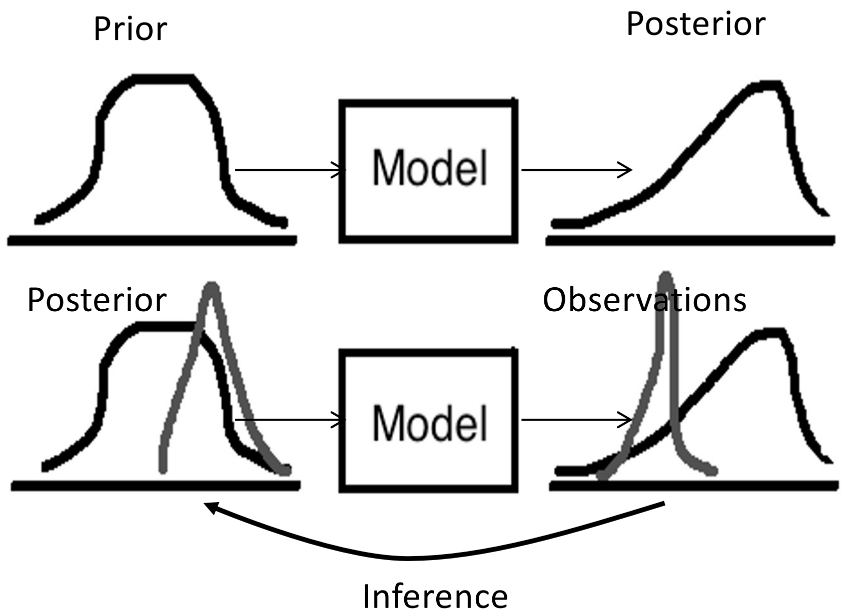
Dear Sir,

Read Dec. 23, 1763. I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illuſtrious Society, and was much eſteemed by many in it as a very able mathematician. In an introduction which he has writ to this Eſſay, he ſays,

Probabilistic modeling

- Express all forms of uncertainty and noise associated with the model via probability theory
- Law of inverse probability (**Bayes' rule**) allows to infer unknown quantities, adapt models to new observations, make predictions and learn from data



Modeling framework

- Consider a model

$$\mathbf{y} = M(\mathbf{x}; \boldsymbol{\theta})$$

$\mathbf{x}$=vector of model inputs; $\mathbf{y}$=vector of model outputs;
$\boldsymbol{\theta}$=vector of model parameters

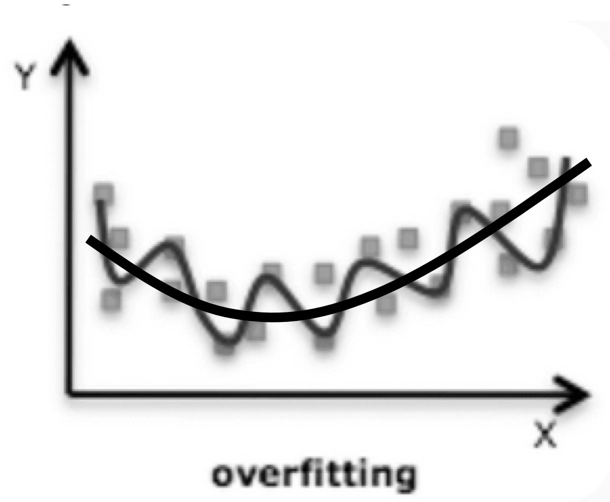
- Choose a functional space for M
  - To fix ideas, we focus on generalized linear models

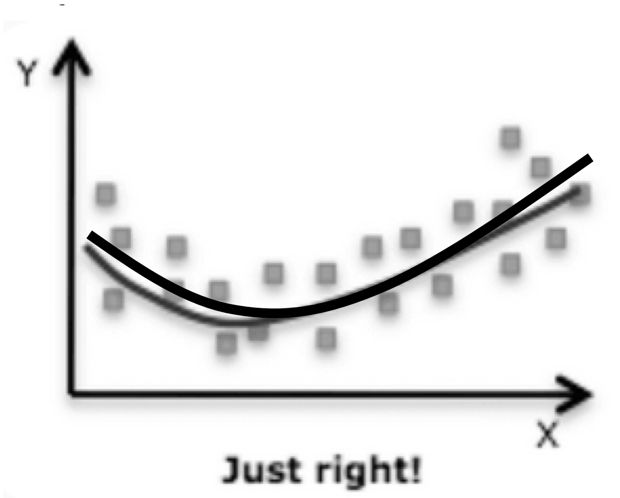
$$M(\mathbf{x}; \boldsymbol{\theta}) = \sum_{m=1}^{M} \theta_m \Phi_m(\mathbf{x})$$

  - $\Phi_m(\mathbf{x})$=basis functions (e.g. radial-basis functions)

- Consider a set of N data, we may write the design matrix:

$$\boldsymbol{\Phi} = \begin{bmatrix} \Phi_1(\mathbf{x}_1) & \dots & \Phi_M(\mathbf{x}_1) \\ \dots & \ddots & \dots \\ \Phi_1(\mathbf{x}_N) & \dots & \Phi_M(\mathbf{x}_N) \end{bmatrix}_{N\times M}$$



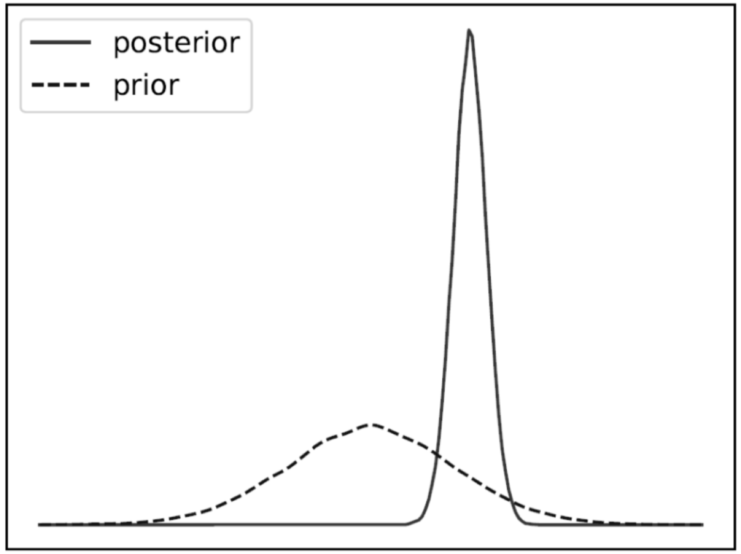


Bayesian inference

- **Bayes'rule** allows to do inference about hypothesis given some data (observations)
- $\theta \rightarrow$ parameters
- $\mathbf{Y} \rightarrow$ observations

Posterior, Likelihood, Prior, Evidence:

$$p(\boldsymbol{\theta}|\mathbf{Y}, M) = \frac{p(\mathbf{Y}, \boldsymbol{\theta}|M)}{p(\mathbf{Y}|M)} = \frac{p(\mathbf{Y}|\boldsymbol{\theta}, M)p(\boldsymbol{\theta})}{p(\mathbf{Y}|M)}$$



- Prior: summarizes hypothesis before we observe data
- Likelihood: probability of observing the data if the hypothesis is true
- Evidence or « Marginal likelihood »: probability that randomly selected parameters from the prior would generate the data
- Posterior: hypothesis updated after observing the data

Bayesian regression vs LMS

- Regression of noisy data:

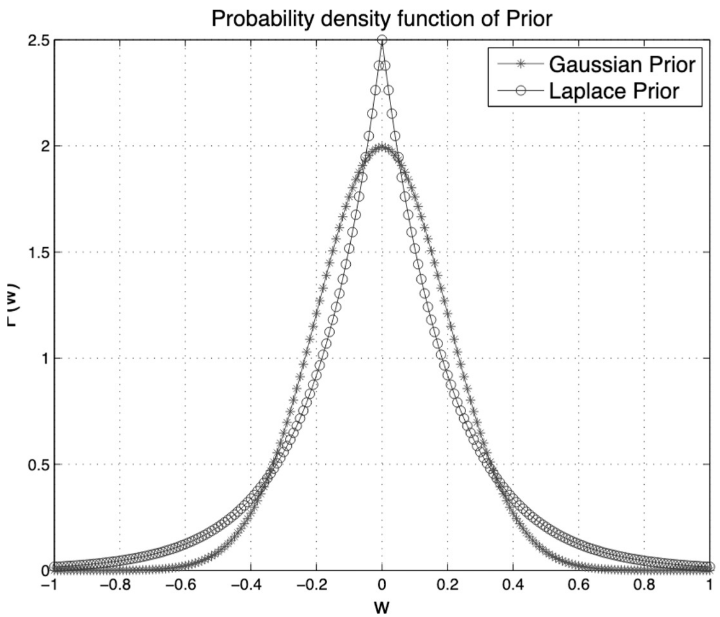
$$Y_n = M(X_n; \boldsymbol{\theta}) + e_n$$

  - e_n i.i.d. Gaussian noise with variance σ^2

$$p(e_n|\sigma^2) = \mathcal{N}(0, \sigma^2)$$

- The maximum likelihood estimate of $\boldsymbol{\theta}$ corresponds to minimizing the MSE to the data
- The maximum a posteriori estimate of $\boldsymbol{\theta}$ corresponds to:
  - **Gaussian** prior → LMS+ridge-type regularization with $\lambda = \sigma^2\alpha$
  - **Laplace** prior → LMS+LASSO-type regularization

More peaked at 0 → sparser



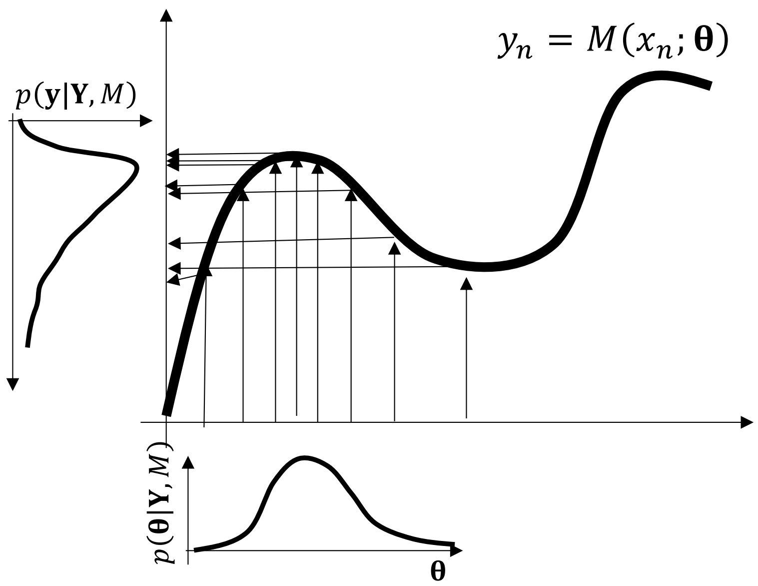
So Bayesian regression equivalent to LMS?

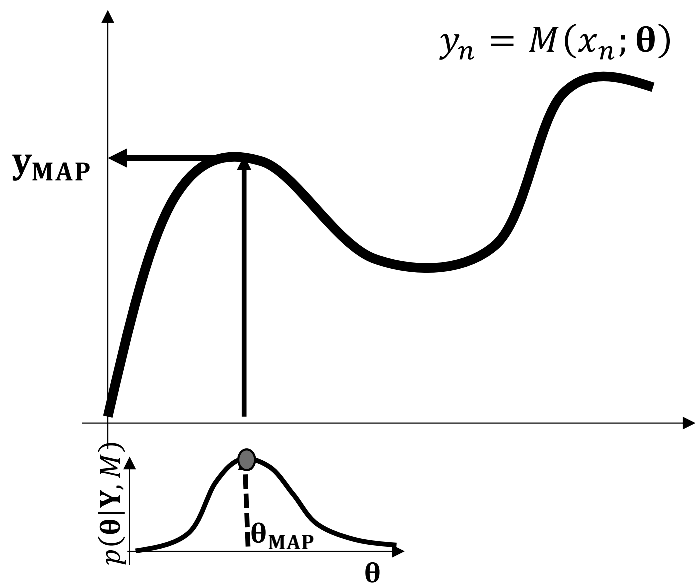
- **No**: now we have **posteriors**!
- Posterior predictive distribution (P.P.D.):

$$p(\mathbf{y}|\mathbf{Y}, M) = \int p(\mathbf{y}|\mathbf{Y}, \boldsymbol{\theta}, \alpha, \sigma^2, M)p(\boldsymbol{\theta}, \alpha, \sigma^2|\mathbf{Y})d\boldsymbol{\theta}d\alpha d\sigma^2$$

$\mathbf{y}$=model output at a new input $\mathbf{x}$

- Prediction with **quantified uncertainty**



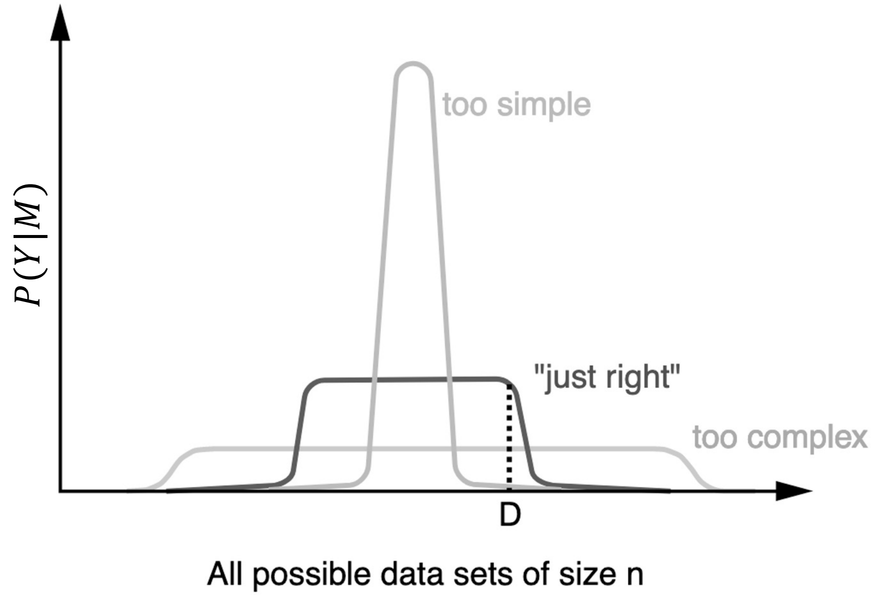


Bayesian Ockham's razor and model comparison

- Automatic model selection through $p(\mathbf{Y}|M)$
  - Compare model classes, e.g. M and M', using posterior model probabilities

$$P(M|\mathbf{Y}) = \frac{p(\mathbf{Y}|M)P(M)}{p(\mathbf{Y})}; \quad P(M'|\mathbf{Y}) = \frac{p(\mathbf{Y}|M')P(M')}{p(\mathbf{Y})}$$

  - Too simple classes unlikely to generate the data set
  - Too complex can generate many possible data sets → unlikely to generate one particular data set at random



From Z. Ghahramani

Pros and cons of Bayesian approach

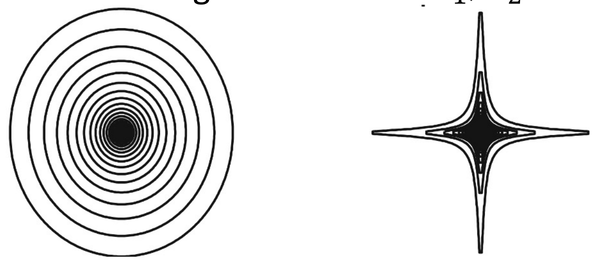
- Pros
  - ☺ Expert knowledge or preferences of the models easily incorporated into the model by prior distribution.
  - ☺ Probabilistic outputs with confidence intervals
  - ☺ Sparsity promoting via automatic implementation of « Ockham's razor »

- Cons
  - ☹ Analytically intractable multidimensional integrals

Model evidence: $p(\mathbf{Y}|M) = \int p(\mathbf{Y}|\boldsymbol{\theta}, M) p(\boldsymbol{\theta}|\alpha) p(\sigma^2) p(\alpha) d\boldsymbol{\theta} d\alpha d\sigma^2$

Posterior predictive distribution: $p(\mathbf{y}|\mathbf{Y}, M) = \int p(\mathbf{y}|\mathbf{Y}, \boldsymbol{\theta}, \alpha, \sigma^2) p(\boldsymbol{\theta}, \alpha, \sigma^2|\mathbf{Y}, M) d\boldsymbol{\theta} d\alpha d\sigma^2$

Sparse Bayesian Learning [Tipping 2001]



- Redundant dictionary of functions
- Gaussian priors with individual precisions α_m
- Objective: find $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma^2|\mathbf{Y}) = \underbrace{p(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\alpha}, \sigma^2)}_{\text{Analytically computable}} \times \underbrace{p(\boldsymbol{\alpha}, \sigma^2|\mathbf{Y})}_{\text{Approximate}}$
- Do iterative procedure:

1. Initialise prior hyperparameters $\boldsymbol{\alpha}, \sigma^2$
2. Posterior maximization: analytically update the optimal weight vector $\boldsymbol{\theta}$ by maximizing the posterior of weights $p(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\alpha}, \sigma^2)$ given $\boldsymbol{\alpha}, \sigma^2$
3. Evidence maximization: update $\boldsymbol{\alpha}, \sigma^2$ by approximately maximizing the evidence $p(\mathbf{Y}|\boldsymbol{\alpha}, \sigma^2)$ →
   $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{MAP}, \sigma^2 = \sigma^2_{MAP}$
   if $\alpha_m \to \infty$, the component is discarded (automatic selection)
4. Loop steps (2) and (3) until converged

Data-driven discovery of turbulence models

[Schmelzer et al., 2020]

SpaRTA = Sparse Regression of Turbulent-stress Anisotropy

- Start with linear eddy viscosity model (here, Menter's $k-\omega$ SST)

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

+ transport equations for k and ω

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

$$b_{ij} = -\frac{\nu_t}{k}S_{ij} + b_{ij}^{\Delta} \qquad \frac{Dk}{Dt} = P + P^{\Delta} + D + T + R \qquad \frac{D\omega}{Dt} = P_{\omega} + P_{\omega,\Delta} + P_{\omega,R} + D + T$$

- **Learn $\boldsymbol{b_{ij}^{\Delta}}$ and $\boldsymbol{R}$ from high-fidelity data**

SPARSE SYMBOLIC IDENTIFICATION

Open-box (symbolic identification) → dictionary of explicit operators

Data-driven discovery of turbulence models

- Ansatz for b_{ij}^{Δ} [inspired from Pope, JFM, 1975]:

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

- Ansatz for R: $$R = 2k b_{ij}^{R} \frac{\partial \bar{u}_i}{\partial x_j}$$

- Dictionary of function libraries $\mathcal{B}_l = [f_{0l}, f_{1l}, \dots f_{Nl}]$ with $f_{kl} = f_{kl}(I_1, \dots, I_5)$

$$b_{ij}^{\square} = \mathbf{C}_{\square} \cdot \mathbf{\Theta}_{\square} \qquad (\square = \Delta, R)$$

$\mathbf{C}_{\square}= \mathcal{B}_1 T_{ij}^{1} \sqcup \mathcal{B}_2 T_{ij}^{2} \sqcup \cdots \sqcup \mathcal{B}_{10}\, T_{ij}^{10}$ → candidate tensor functions

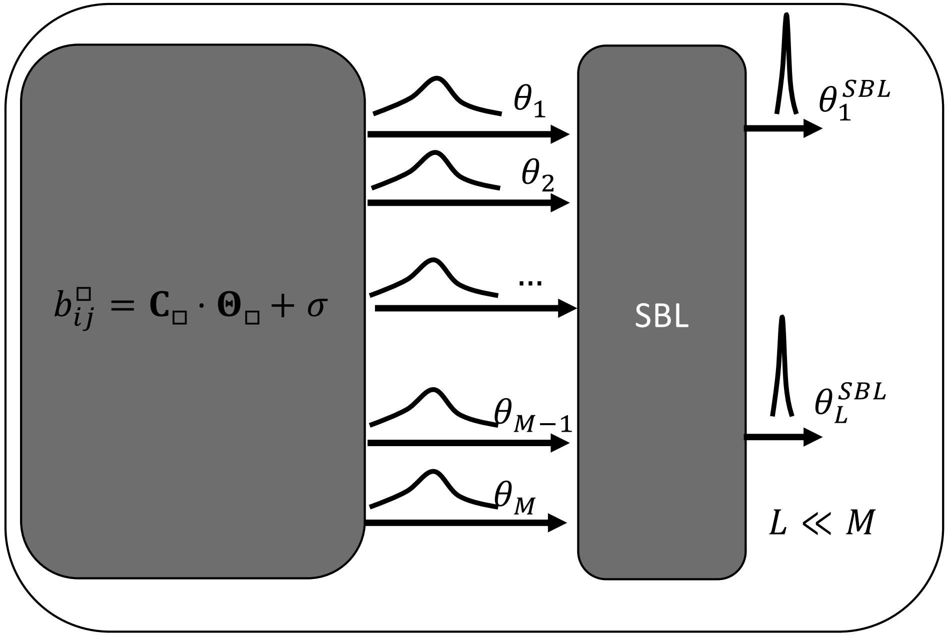
$\mathbf{\Theta}_{\square} = \Theta_1 \sqcup \Theta_2 \sqcup \dots \sqcup \Theta_{10}$ → vector of coefficients

- Find Θ by solving a **regularized regression problem** so that most coefficients are zero

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

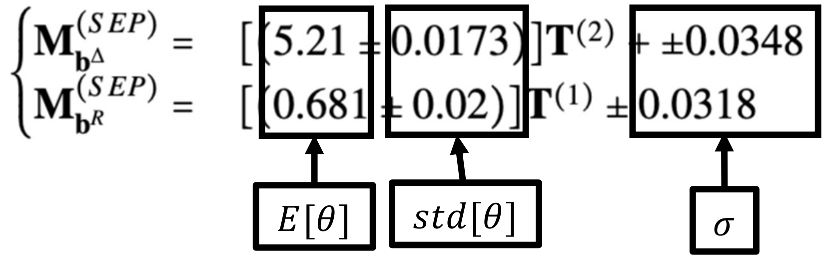
SBL-SpaRTA [Cherroud et al., 2022]

- Find $p\left(\boldsymbol{\Theta}_{\square}, \boldsymbol{\alpha}, \sigma^2 \middle| b_{ij}^{\square}\right)$ using the SBL algorithm
  - Select functions in $\mathbf{C}_{\square}$ dictionary and infer posteriors



Ref	Abr.	Data
D_1	ZPG	DNS of turbulent boundary layer, $670 \leq Re_\theta \leq 4060$[1]
D_2	FDC	DNS of turbulent channel flow, $180 \leq Re_\tau \leq 590$[2]
D_3	ANSJ	PIV of near sonic axisymmetric jet [3]
D_4	APG	LES of adverse pressure-gradient TBL [4] $Re_\theta \leq 4000$, $\beta = 4$, 5 different pressure gradients
D_5	SEP	LES of Periodic Hills at Re=10595 [5] DNS of converging-diverging channel at Re=13600 [6] LES of curved backward facing step at Re = 13700 [7]

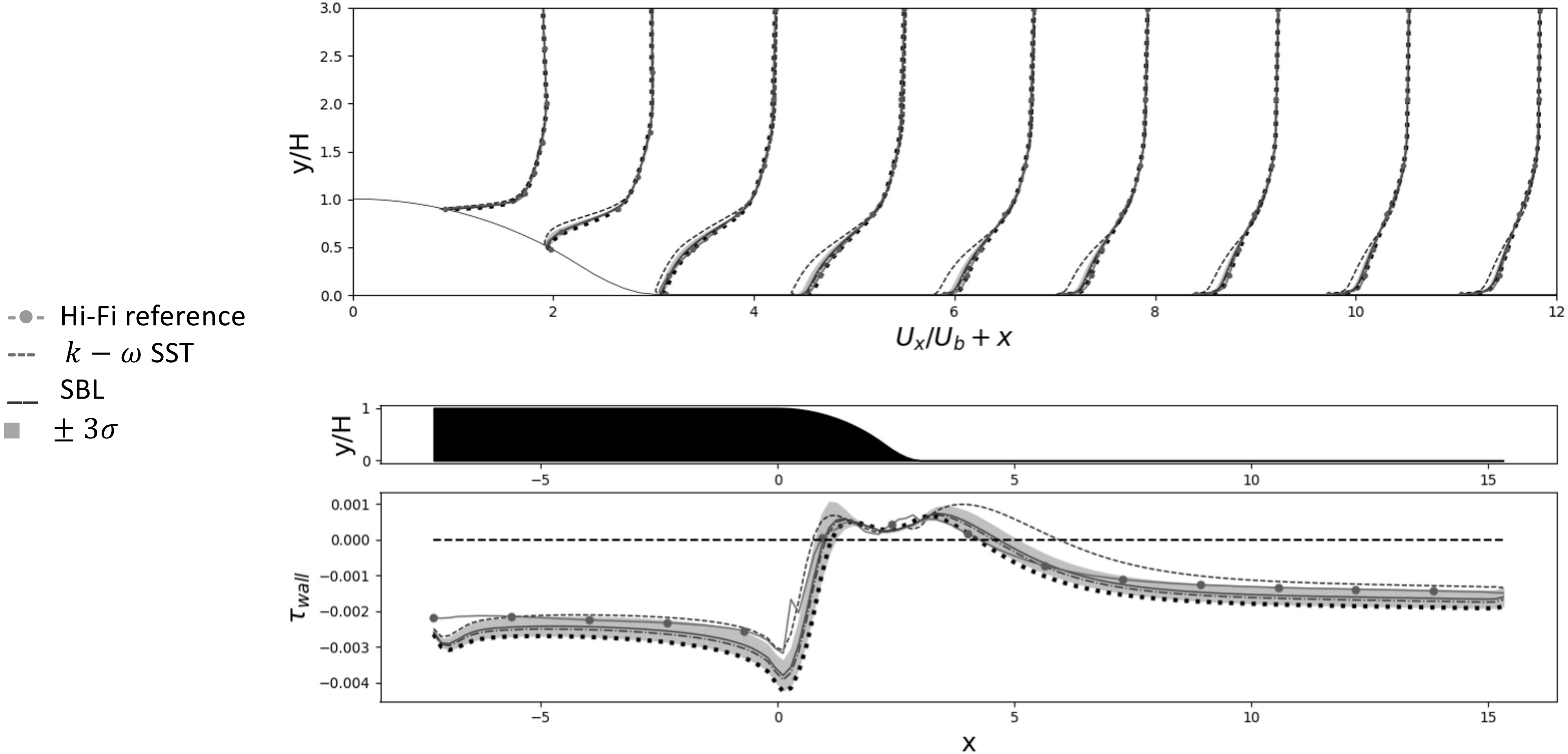
Example of discovered model:



FDC : the discovered model correction is 0!

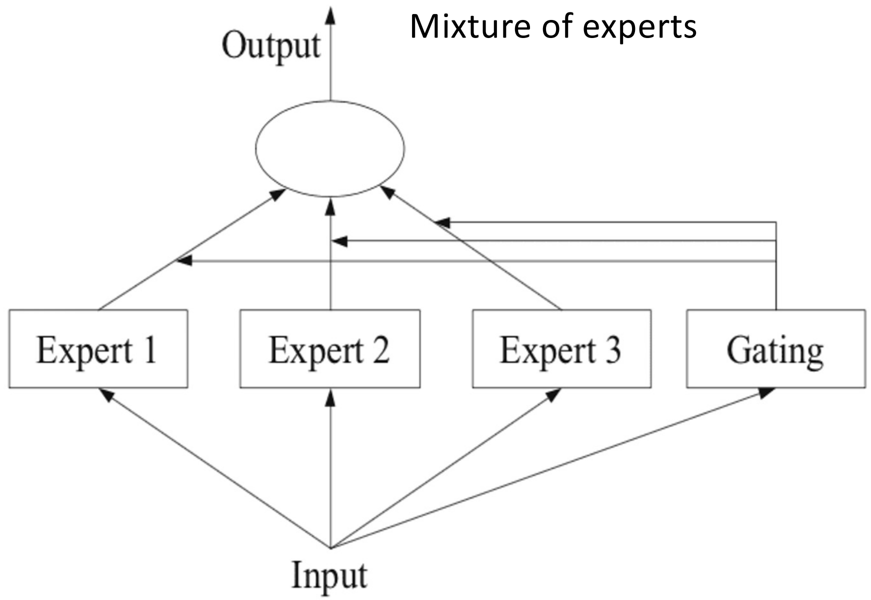
SBL-SpaRTA

- Curved backward-facing step flow at Re=13700



Mixtures of experts

- How good is our model at predicting another data sets?
- Large data sets: combining models trained on subsets better than single model trained over all data
- Out of sample predictions: uncertainty on which model (among those at hand) is better
- Generate hypermodels by combining component models through some gating functions (regional weights)



Mixture of experts

Spatial model aggregation (X-MA) of turbulence models [de Zordo et al., 2021]

- Consider a set of K competing models $\boldsymbol{\mathcal{M}} = \{\boldsymbol{M_1}, \boldsymbol{M_2}, \dots, \boldsymbol{M_K}\}$
- « Hypermodel »: $M_{hyp}(\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k M_K(\mathbf{x}; \boldsymbol{\theta})$, with

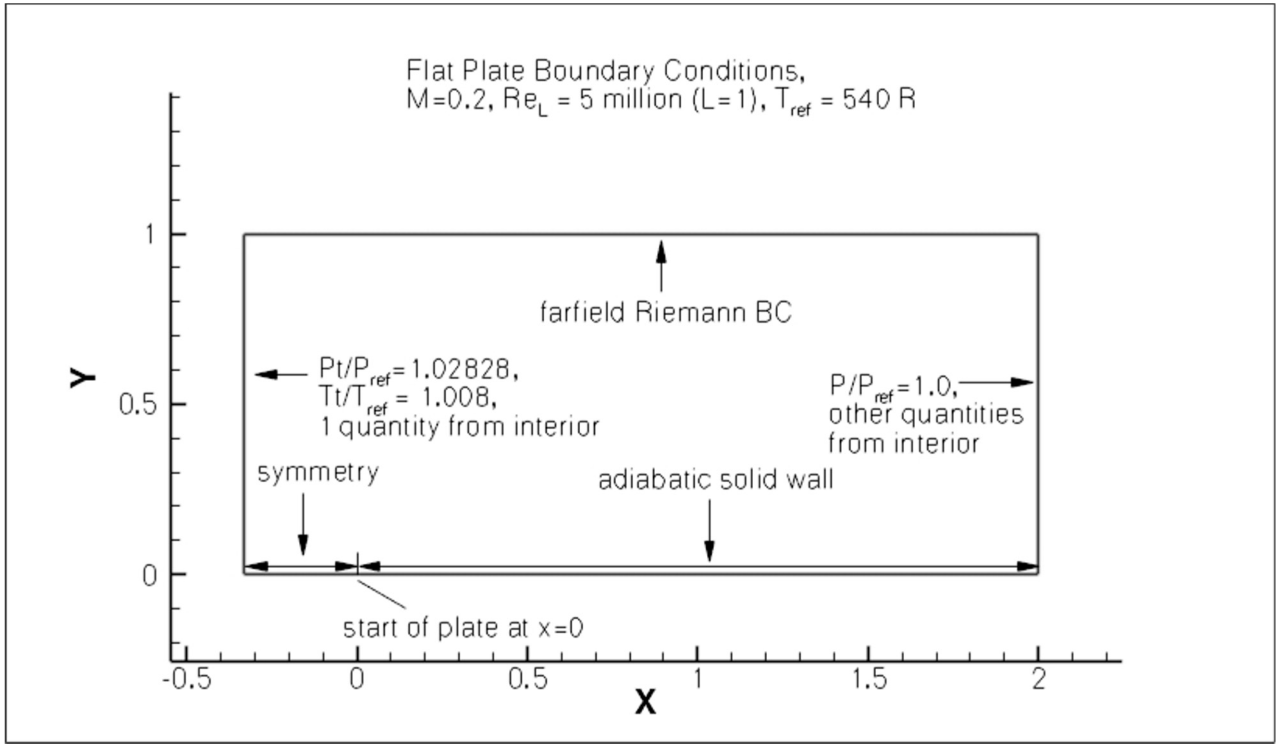
$$w_k = w_k(\mathbf{x}) = w_k(\boldsymbol{\eta}(\mathbf{x}))$$

- Regress $w_k(\boldsymbol{\eta}(\mathbf{x})|\mathbf{Y})$ from data as a function of features → **Random Forests**
  - Predict local model weights for a new case $w_k(\boldsymbol{\eta}(\mathbf{x})|\mathbf{Y})$ and use them to aggregate individual model predictions

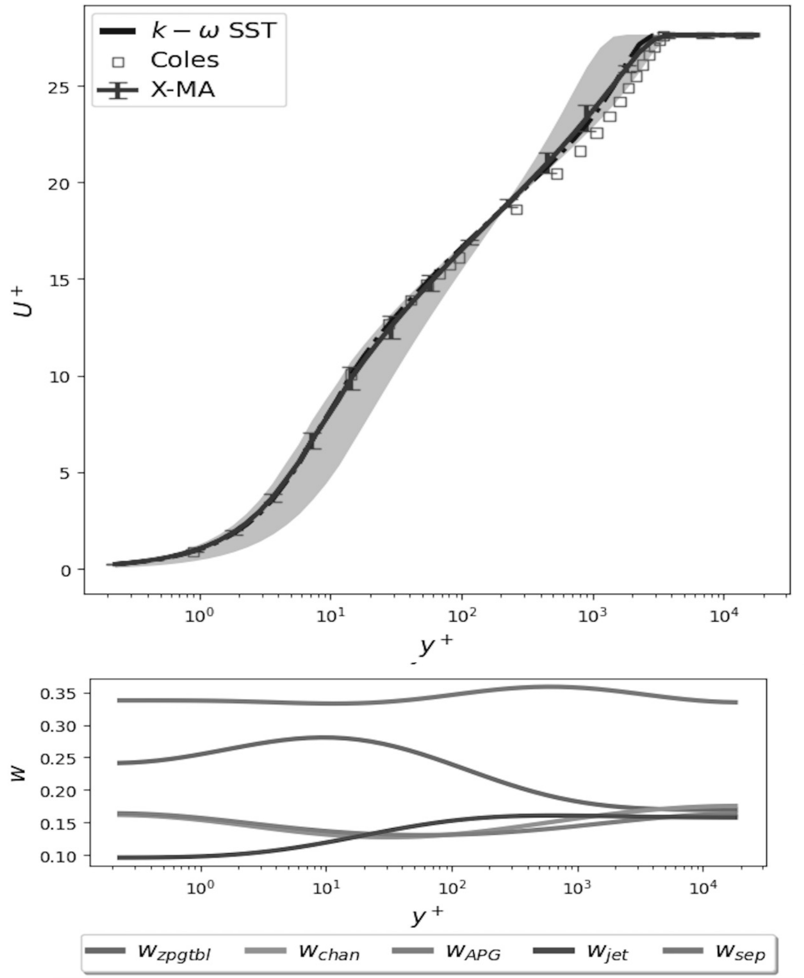
Feature	Description	Formula	Feature	Description	Formula
η_1	Normalized Q criterion	$\dfrac{\|\|\boldsymbol{\Omega}\|\|^2 - \|\|\mathbf{S}\|\|^2}{\|\|\boldsymbol{\Omega}\|\|^2 + \|\|\mathbf{S}\|\|^2}$	η_6	Viscosity ratio	$\dfrac{\nu_T}{100\nu + \nu_T}$
η_2	Turbulence intensity	$\dfrac{k}{0.5U_iU_i + k}$	η_7	Ratio of pressure normal stresses to normal shear stresses	$\dfrac{\sqrt{\frac{\partial P}{\partial x_i}\frac{\partial P}{\partial x_i}}}{\sqrt{\frac{\partial P}{\partial x_j}\frac{\partial P}{\partial x_j} + 0.5\rho\frac{\partial U_k^2}{\partial x_k}}}$
η_3	Turbulent Reynolds number	$\min\left(\dfrac{\sqrt{k}\lambda}{50\nu}, 2\right)$	η_8	Non-orthogonality marker between velocity and its gradient [28]	$\dfrac{\left\|U_kU_l\frac{\partial U_k}{\partial x_l}\right\|}{\sqrt{U_nU_nU_i\frac{\partial U_i}{\partial x_j}U_m\frac{\partial U_m}{\partial x_j}} + \left\|U_iU_j\frac{\partial U_i}{\partial x_j}\right\|}$
η_4	Pressure gradient along streamline	$\dfrac{U_k\frac{\partial P}{\partial x_k}}{\sqrt{\frac{\partial P}{\partial x_j}\frac{\partial P}{\partial x_j}U_iU_i} + \left\|U_l\frac{\partial P}{\partial x_l}\right\|}$	η_9	Ratio of convection to production of k	$\dfrac{U_i\frac{\partial k}{\partial x_i}}{\|\overline{u_j'u_l'}S_{jl}\| + U_l\frac{\partial k}{\partial x_l}}$
η_5	Ratio of turbulent time scale to mean strain time scale	$\dfrac{\|\|\mathbf{S}\|\|k}{\|\|\mathbf{S}\|\|k + \varepsilon}$	η_{10}	Ratio of total Reynolds stresses to normal Reynolds stresses	$\dfrac{\|\|\overline{u_i'u_j'}\|\|}{k + \|\|\overline{u_i'u_j'}\|\|}$

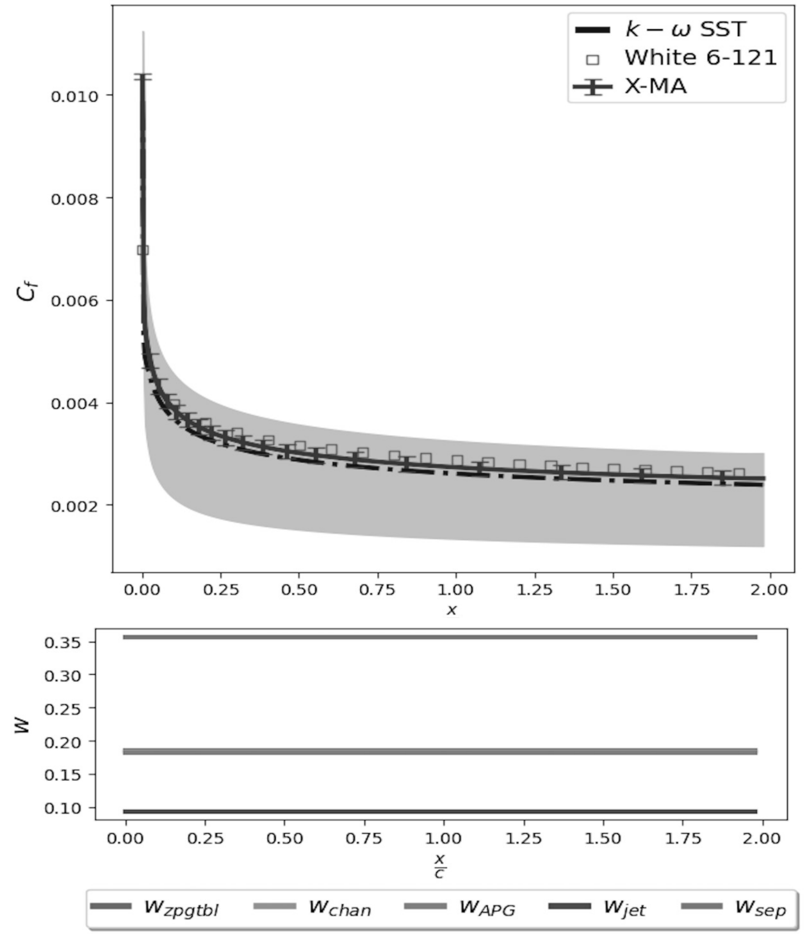
NASA Turbulence Modeling Testing Challenge

- Application to Test Case 1 2DZP: 2D Zero Pressure Gradient Flat Plate
  - Show (1) Cf vs. x and (2) u+ vs. log(y+) at x=0.97; compare with theory



NASA Turbulence Modeling Testing Challenge

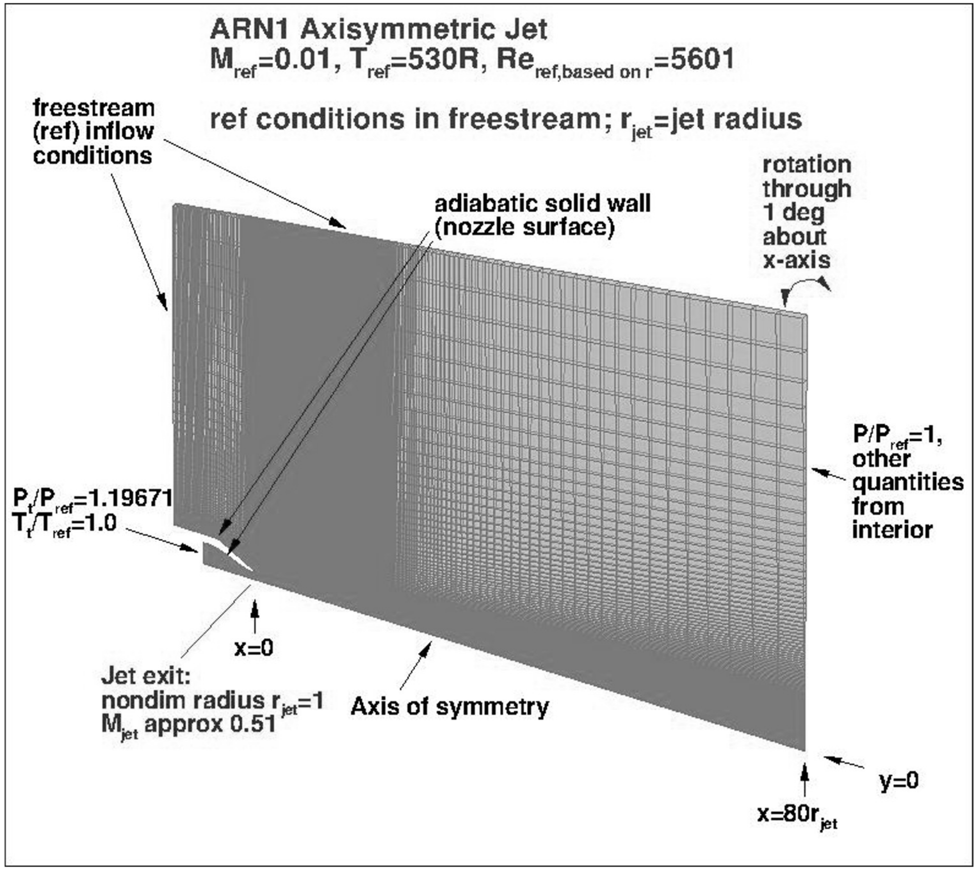




Accurate prediction of the 2DZP

NASA Turbulence Modeling Testing Challenge

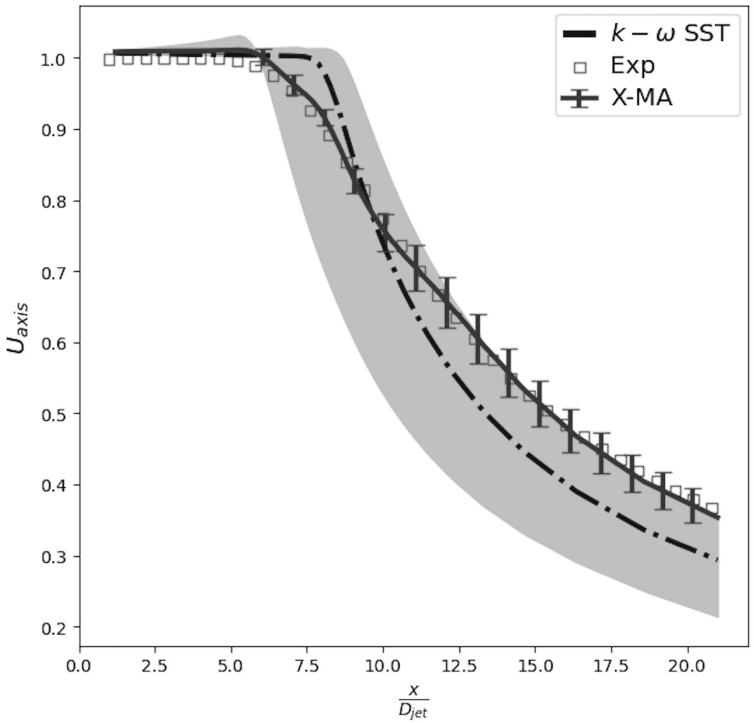
- Application to Test Case 3 ASJ: Axisymmetric Subsonic Jet



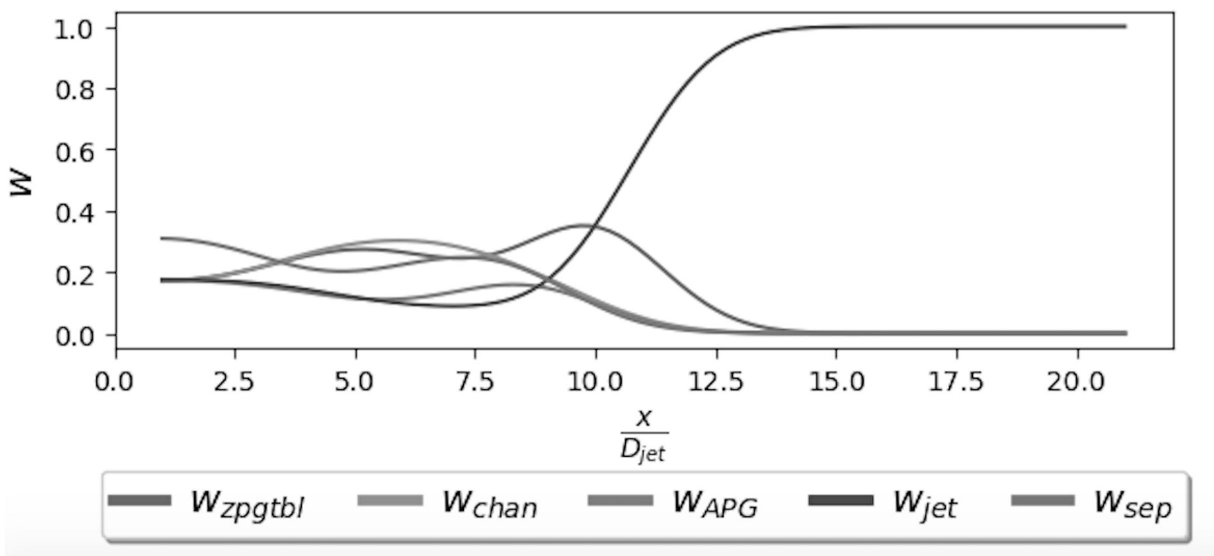
y+

NASA Turbulence Modeling Testing Challenge

- Application to Test Case 3 ASJ: Axisymmetric Subsonic Jet



$\mathbf{k}-\boldsymbol{\omega}$



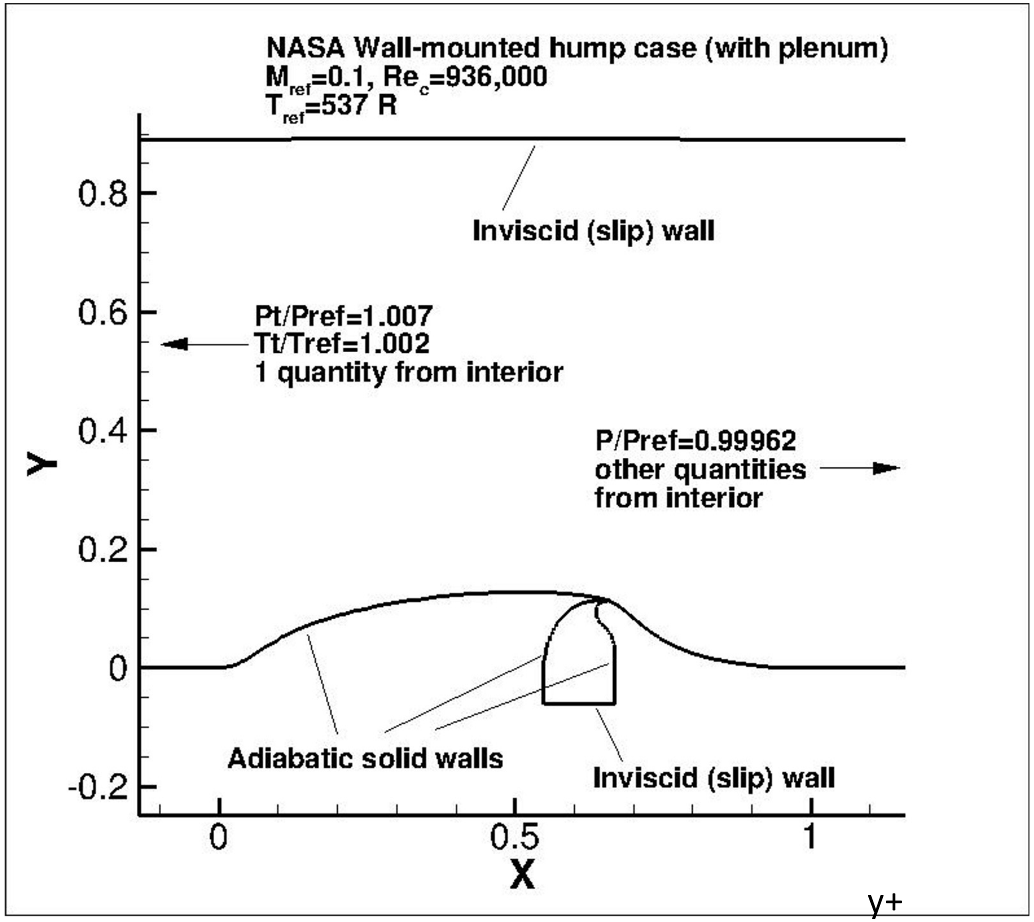
$\mathbf{k}-\boldsymbol{l}$

X-MA puts higher weight on the 'jet model' in the far jet region

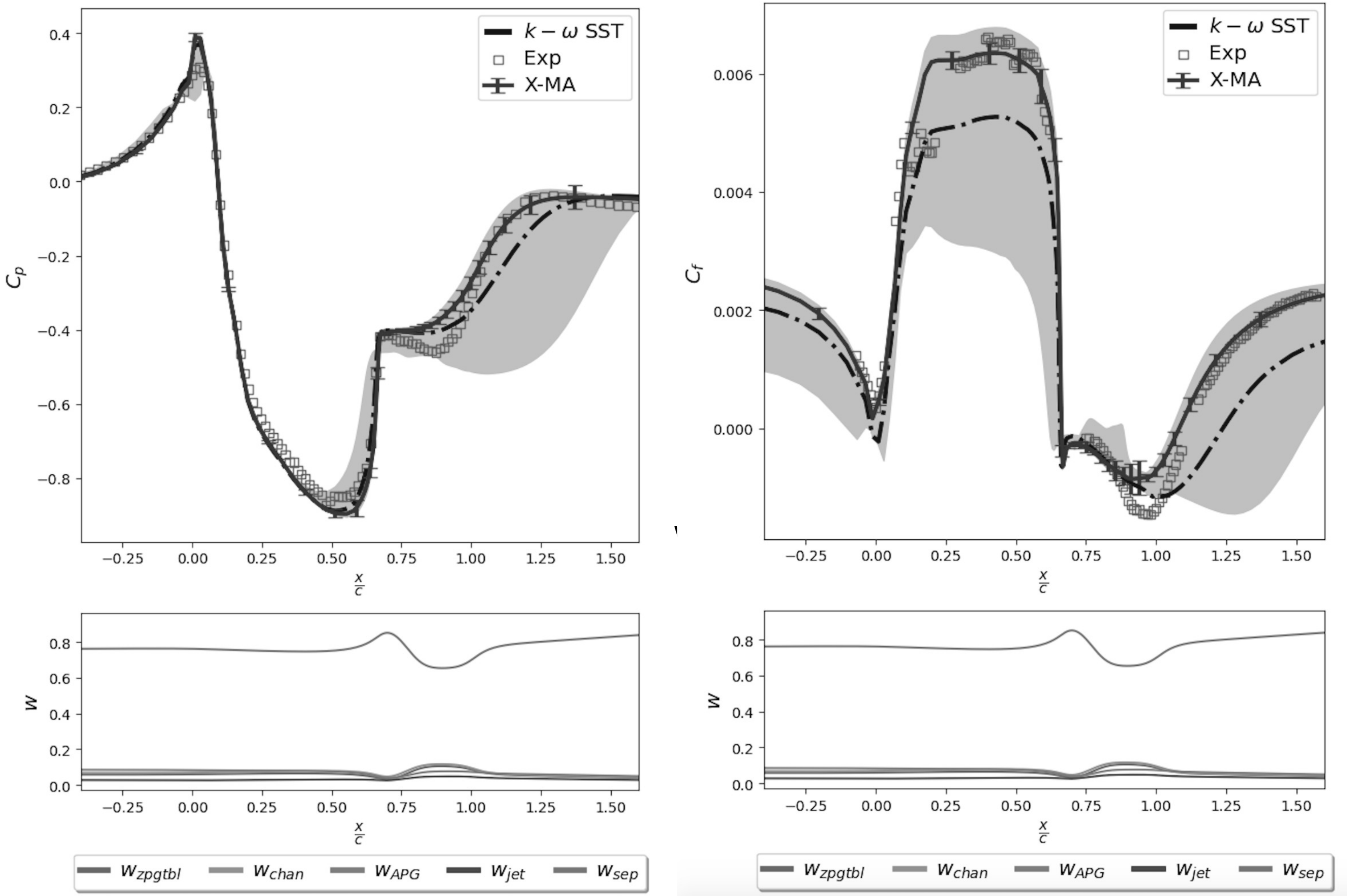
y+

NASA Turbulence Modeling Testing Challenge

- Application to Test Case 4 2DWMH: 2D NASA Wall-Mounted Hump Separated Flow Validation Case



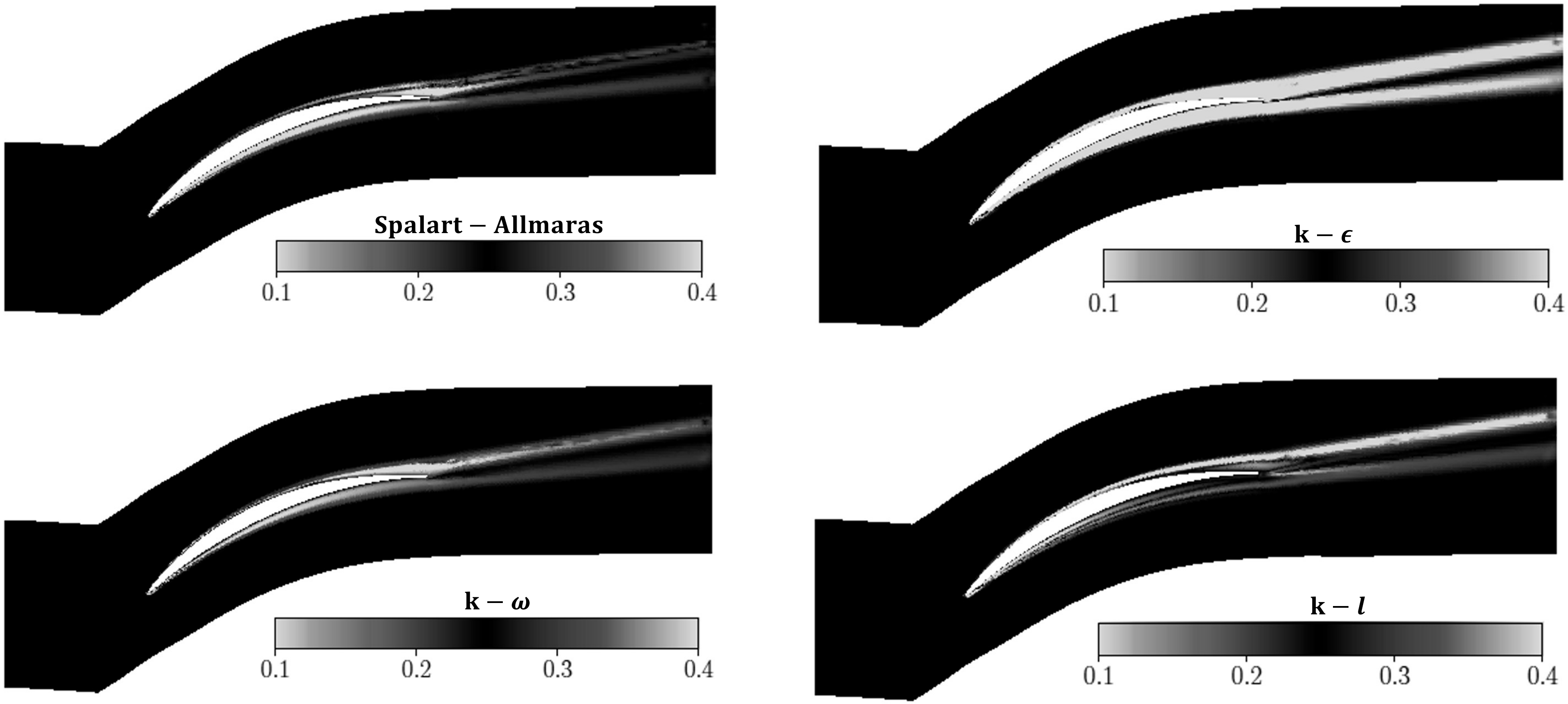
NASA Turbulence Modeling Testing Challenge



X-MA better than individual models for all cases + improvement over the baseline

Application to the NACA 65 V103 cascade

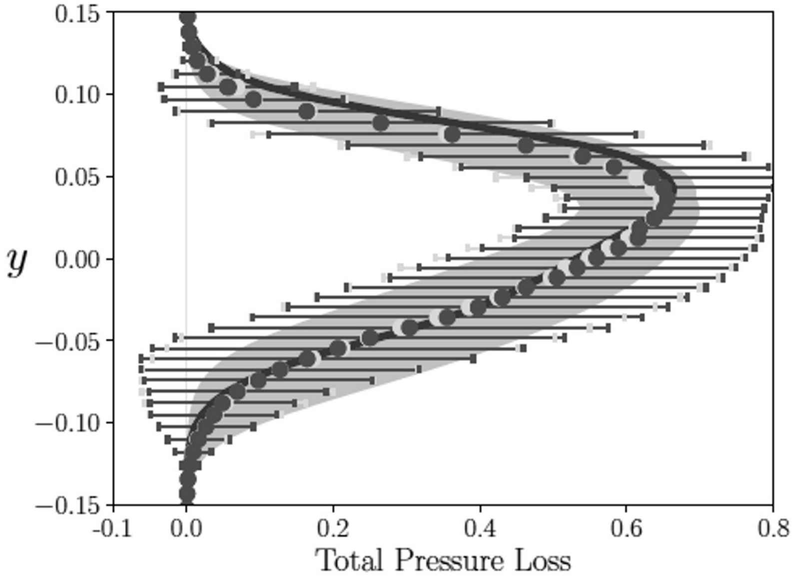
- **Experts**: Spalart-Allmaras, $\mathbf{k}-\boldsymbol{\epsilon}$, $\mathbf{k}-\boldsymbol{\omega}$ and $\mathbf{k}-\mathbf{l}$ models of turbulence
- Training: synthetic total pressure data generated from k-kl EARSM model for S1,S3 and S4; prediction for S2



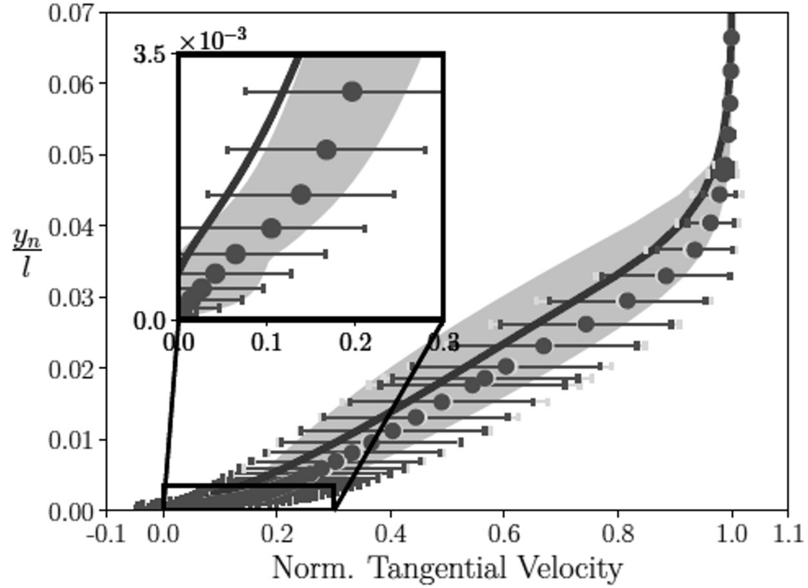
Prediction of selected QoIs for an unseen scenario

- Expectancy and standard deviation computed from local probability mass distribution (model weights)

Total Pressure loss in the wake



Velocity profile at suction side



X-MA consistently favors the best model at each prediction point

Conclusions

- Bayesian data-driven turbulence framework
  - Sparse, interpretable and stochastic models
  - Generalization through **model averaging**
  - Delivers uncertainty estimates
- Encouraging proof-of-concept obtained for data-driven turbulence modeling
  - application to the NASA turbulence modeling testing challenge
- Future work:
  - Further investigate SBL and other Bayesian machine learning algorithms
  - Optimal choice of features?
  - 3D flows?
  - How to perform efficient model-in-the-loop learning?

Acknowledgements

- Bayesian learning

Soufiane Cherroud
DynFluid, ENSAM

Maximilien De Zordo-Banliat
DynFluid, ENSAM

Cécile Roques
D'Alembert, SU

Xavier Merle
DynFluid, ENSAM

Grégory Dergham
Safran Tech

- High-fidelity simulations

Xavier Gloerfelt
DynFluid, ENSAM

Aurélien Bienner
DynFluid, ENSAM

Camille Matar
D'Alembert, SU

Donatella Passiatore
CTR Stanford

Luca Sciacovelli
DynFluid, ENSAM

Advertizing

- Two postdoc positions and a PhD position available in my LearnFluidS group:
  - Physics-constrained deep network augmentation of turbulence models
  - Model-consistent Bayesian learning of turbulence models from sparse data

- One PhD position:
  - Machine-learning-assisted wall-modelled large eddy simulations of transitional flows in turbomachinery

Contact: paola.cinnella@sorbonne-universite.fr

Questions?

1. Xiao H., Cinnella P., «Quantification of Model Uncertainty in RANS Simulations: A Review», Progress in Aerospace Sciences, Vol. 108, pp. 1-31, 2019
2. Hoarau J.-Ch., Cinnella P., Gloerfelt X., « Large Eddy Simulations of strongly non ideal compressible flows through a transonic cascade », Energies 14(3):772, 2021
3. Bienner A., Gloerfelt X., Cinnella P., Leading-edge effects in freestream-turbulence-induced transition in a dense gas flow. Workshop Direct and Large-Eddy Simulation (DLES13), 26-20 October 2022, Udine, Italy
4. Schmelzer M., Dwight R., Cinnella P., «Discovery of Algebraic Reynolds-stress Models using Sparse Symbolic Regression», Flow, Turbulence and Combustion,Vol. 104, No. 2–3, pp. 579–603, 2019
5. Cherroud S., Merle X., Cinnella P., Gloerfelt X., "Sparse Bayesian Learning of Explicit Algebraic Reynolds-Stress models for turbulent separated flows", International Journal of Heat and Fluid Flow, 98: 109047, December 2022
6. De Zordo-Banliat M., Dergham G., Merle X., Cinnella P., « Space-dependent turbulence model aggregation using machine learning », submitted to the Journal of Computational Physics, December 2022. ArXiv:2301.09013
7. Cherroud S., Merle X., Cinnella P., Gloerfelt X., Spatial Model Aggregation (X-MA) of stochastic Explicit Algebraic Reynolds Stress models, 2022 Symposium on Turbulence Modeling: Roadblocks, and the Potential for Machine Learning. July 27-29 2022, Suffolk (VI), USA

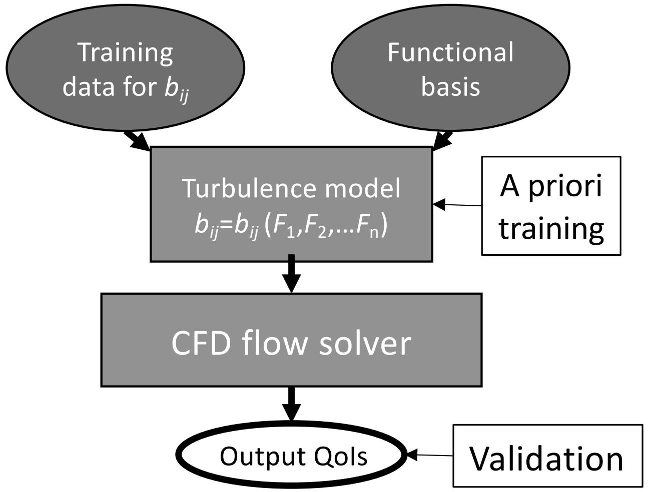
Garry Kasparov vs. Deep Blue
By Feanny, Najlah,1997

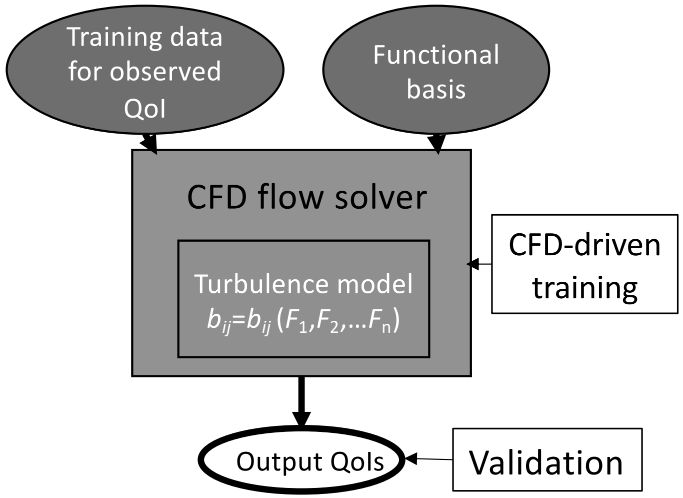
Some properties of Bayesian inference

- Asymptotic certainty:
  - Under suitable assumption on the prior, $\lim_{n\to\infty} p(\boldsymbol{\theta}|\mathbf{Y}) = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_{true})$
  - If the data cannot be captured by the model for any \theta (unrealizable case), $\lim_{n\to\infty} p(\boldsymbol{\theta}|\mathbf{Y}) = \delta\big(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\big)$, with $\widehat{\boldsymbol{\theta}}$ minimizing the distance between the predictive posterior and the true distribution in the Kullbach-Leibler norm
- Asymtotic consensus:
  - Given two different priors of the parameters, the corresponding posteriors tend to converge as $n \to \infty$

Data-augmented turbulence modeling

- Two training strategies
  - **A priori ("CFD-free") training**
    - ☺ Inexpensive (manipulate analytical expressions)
    - ☹ Requires high-fidelity, low noise data for turbulent quantities
    - ☹ Does not warrant model consistency
    - ☹ May lead to non robust models and conditioning problems
  - **Model-in-the-loop ("CFD-driven") training**
    - ☺ May use virtually any data (mean flow and turbulent quantities)
    - ☺ Model-consistent
    - ☺ Produces numerically robust models
    - ☹ Requires the solution of a costly multidimensional optimization problem





Bayesian regression vs LMS

- Regression of noisy data:

$$Y_n = M(X_n; \boldsymbol{\theta}) + e_n$$

  - e_n modelled as zero mean, independent and identically distributed Gaussian noise with variance σ^2

$$p(e_n|\sigma^2) = \mathcal{N}(0, \sigma^2)$$

- Then:

$$p(Y_n|x_n; \boldsymbol{\theta}, \sigma^2) = N(M(X_n; \boldsymbol{\theta}), \sigma^2)$$

  and

$$p(\mathbf{Y}|\boldsymbol{\theta}, \sigma^2) = \prod_{n=1}^{N} N(M(X_n, \boldsymbol{\theta}), \sigma^2) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(Y_n - M(X_n, \boldsymbol{\theta})\right)^2}{2\sigma^2}\right)$$

- The maximum likelihood estimate of $\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta}} p(\mathbf{Y}|\boldsymbol{\theta}, \sigma^2)$, or **min of the negative log-likelihood**:

$$-\log\left(p(\mathbf{Y}|\boldsymbol{\theta}, \sigma^2)\right) = \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(Y_n - M(X_n, \boldsymbol{\theta})\right)^2$$

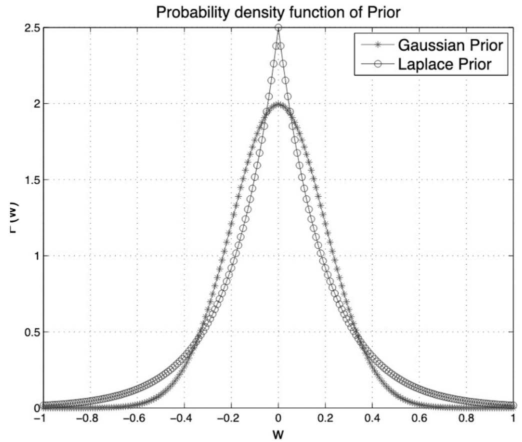
Which corresponds to minimizing the MSE to the data

Bayesian regularization via priors

- A prior expresses the degree of belief about values parameters can take:
  - E.g.: zero mean Gaussian prior with α the inverse variance (hyperparameter)

$$p(\boldsymbol{\theta}|\alpha) = \prod_{m=1}^{M} \sqrt{\frac{\alpha}{2\pi}} \exp\left(-\frac{\alpha}{2}\theta_m^2\right)$$

  - Encodes the belief that parameters are mostly zero (low complexity model)



- Combine beliefs about the prior and the likelihood via Bayes' rule => posterior:

$$p(\boldsymbol{\theta}|\mathbf{Y}, \alpha, \sigma^2) \propto p(\mathbf{Y}|\boldsymbol{\theta}, \sigma^2)p(\boldsymbol{\theta}|\alpha) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(Y_n - M(X_n, \boldsymbol{\theta})\right)^2}{2\sigma^2}\right) \prod_{m=1}^{M} \sqrt{\frac{\alpha}{2\pi}} \exp\left(-\frac{\alpha}{2}\theta_m^2\right)$$

- Maximum A Posteriori (MAP) approximation ($\boldsymbol{\theta}$ most probable a posteriori):

$$\text{MAP}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{2\sigma^2} \sum_{n=1}^{N} \left(Y_n - M(X_n, \boldsymbol{\theta})\right)^2 + \frac{\alpha}{2} \sum_{m=1}^{M} \theta_m^2\right)$$

- **Gaussian** prior → ridge-type regularization with $\lambda = \sigma^2\alpha$
- **Laplace** prior → LASSO-type regularization

More peaked at 0 → sparser