

Turbulence Modeling in the Age of Data

Karthik Duraisamy, Gianluca Iaccarino, Heng Xiao

<https://arxiv.org/abs/1804.00183>

1. Introduction

over the past half century, starting from the pioneering theoretical studies by *Prandtl*, *Kolmogorov*, and *von Karman*.

RANS(Reynolds Averaged Navier-Stokes) and LES(Large Eddy Simulation) approaches are the most common.

RANS techniques rely completely on modeling assumptions to represent turbulent characteristics and, therefore, lead to considerably lower computational requirements than DNS.

RANS models require closures to represent the turbulent stresses and scalar fluxes emerging from the averaging process.

$$\begin{aligned}\overline{u}(x,t) &\triangleq \frac{1}{T} \lim_{T \rightarrow \infty} \int_0^T u(x,t) dt, & u' &\triangleq u - \overline{u} \\ \implies & u = \overline{u} + u'\end{aligned}$$

then will lead to some terms ($u'_i u'_j$)

Single-point RANS models implicitly assume an equilibrium spectrum and locally-defined constitutive relationships to close the averaged governing equations and express unclosed terms as a function of averaged, local flow quantities.

LES techniques directly represent a portion of the active scales and only require modeling to account for the unresolved turbulent motions. LES is gaining popularity in relatively small Reynolds numbers.

Experimental observation + Statistical inference approaches + modern machine learning.

2. Turbulence closures and uncertainties

L1 : Uncertainties introduced by ensemble-averaging, which are fundamentally irrecoverable. (取雷诺平均带来的不封闭问题)

time- or ensemble-averaging operators + nonlinearity of the NS equation

$$\langle \mathcal{N}(\cdot) \rangle \neq \mathcal{N}(\langle \cdot \rangle)$$

L2 : Uncertainties in the functional and operational representation of Reynolds stress (雷诺应力模型, 对脉动项进行估计)

need to model the $\mathcal{M}(\cdot)$

$$\langle \mathcal{N}(\cdot) \rangle = \mathcal{N}(\langle \cdot \rangle) + \mathcal{M}(\cdot)$$

for an incompressible fluid, $\mathcal{M}(\cdot) = \nabla \cdot \tau$, τ is the Reynolds stress tensor.

L3 : Uncertainties in functional forms within a model (提出不同与雷诺应力的其他模型)

$$\mathcal{M}(\mathbf{w}; \mathcal{P}(\mathbf{w}))$$

One and two-equation models are the most popular.

$\mathcal{P}(\cdot)$ mimic the terms in the Navier-Stokes equation such as convection and diffusion

L4 : Uncertainties in the coefficients within a model (系数误差)

$$\mathcal{M}(\mathbf{w}; \mathcal{P}(\mathbf{w}); \mathbf{c})$$

The choice of the C_μ coefficient in two-equation linear eddy viscosity models is a classical **L4** closure issue

interest \mathbf{q}

$$\mathbf{q} = \mathbf{q}(\mathcal{N}(\langle \cdot \rangle); \mathcal{M}(\mathbf{w}; \mathcal{P}(\mathbf{w}); \mathbf{c}))$$

Uncertainty Quantification

1. aleatory uncertainties(偶然误差)
2. epistemic and model-form uncertainties (the limitations intrinsic in the physics models) (认知上的误差)

3. Models, data and calibration

the famous isotropic turbulence experiments of Comte-Bellot and Corrsin (1966)

The Summer Program of the Center for Turbulence Research at Stanford University

3.1. Naive calibration (朴素校准)

the measured data may be the same as the quantity of interest \mathbf{q} , i.e. Uncertainties in the measurements are typically ignored.

model coefficients(\mathbf{c}) are the dominant source of uncertainty

$$\widetilde{\mathcal{M}} = \mathcal{M}(\mathbf{w}; \mathcal{P}(\mathbf{w}); \tilde{\mathbf{c}}_{\mathbf{q}})$$

3.2. Statistical inference

$$\widetilde{\mathcal{M}} = \mathcal{M}(\mathbf{w}; \mathcal{P}(\mathbf{w}); \tilde{\mathbf{c}}_{\theta}) + \delta + \epsilon_{\theta}$$

posterior probability (maximum a posteriori estimation, MAP)

Statistical inversion

statistical inversion aims to identify parameters \mathbf{c} of a model $\mathcal{M}(\mathbf{c})$ given data θ with uncertainty ϵ_{θ}

probabilities : $p(\text{data} | \text{distribution})$

likelihoods : $p(\text{distribution} | \text{data})$

3.3. Data-driven modeling

the emphasis is on δ rather than on \mathcal{M} ,

represent the discrepancy δ in terms of features η , such as the mean velocity gradients.

$$\widetilde{\mathcal{M}} = \mathcal{M}(\mathbf{w}; \mathcal{P}(\mathbf{w}); \mathbf{c}(\theta); \delta(\theta, \eta); \epsilon_{\theta})$$

3.4. Calibration and prediction

model-form uncertainty is introduced to the prediction by adding a discrepancy term to the model output $\mathbf{o}(\mathcal{M}(\mathbf{c}))$

4. Quantifying uncertainties in RANS models

predictions based on RANS models are affected by the assumptions invoked in the construction of the closure (L1–L4) and by the calibration process

closure problem + calibration process

4.1. Uncertainties in the Reynolds stress tensor

an interval, probabilistic description

RANS modeling

- traditionally aim to approximate unclosed model in the averaged equations
- an alternative idea is to replace these unclosed terms with bounds that are based on theoretical arguments
 - background flow method (decomposes the quantity of interest, for example the energy dissipation rate, into a background profile and a fluctuating component)
 - Reynolds stress realizability (introduce realizable, physics-constrained perturbations to the Reynolds stress tensor) (eigenvalues + Barycentric coordinates)
 - eigen-decomposition, three representative limiting states: one-component(1C,2C,3C)
 - random matrix approach, a probabilistic description of the Reynolds stress uncertainty $\mathbb{E}[\mathbf{T}] = \boldsymbol{\tau}^{\text{RANS}}$

bounding aims to construct prediction intervals that can be proven to contain the true answers, as opposed to explicit estimates that might be inaccurate.

Both approaches have focused on the error bound of the Reynolds stress at a single point.

4.2. Uncertainty in model parameters

optimal parameters for several typical free shear flows

classical uncertainty propagation techniques

4.3. Identifying regions of uncertainty

formulated the evaluation of potential adequacy of the RANS model as a classification problem in machine learning.

5. Predictive modeling using data-driven techniques

above aims at providing confidence in the application of RANS closures by identifying and quantifying uncertainties in the models at various levels.

focus on improving the overall prediction accuracy

5.1. Embedding inference-based discrepancy

inferre the discrepancy as a spatially varying field

The approaches listed above use statistical inference to construct posterior probability distribution of a quantity of interest based on data

An alternative viewpoint, is that this models aims at producing models in which the operators $\mathcal{P}(w)$ used in the (L3) modeling step

5.2. Generalizing the embedded discrepancy

construct discrepancy functions that can be employed within a class of flows sharing similar features(separation, shock/boundary layer interaction, jet/boundary layer interaction)

Machine Learning

5.3. Modeling using machine learning

ML can be applied directly as a black-box tool, or in combination with existing models to provide a posteriori corrections.

use DNS data to enhance the RANS simulation is not encouraging.

use machine learning with DNS data to model the RANS terms is much accurate.

and the linear eddy viscosity model not take in account that the discrepancy between the linear model and the real model. i.e. the rotation

An important aspect of applying machine learning techniques is to ensure the objectivity and the rotational invariance of the learned Reynolds stress models.

earlier a strategy to develop closures for Reynolds stresses is based on the formulation of a generalized expansion of the Reynolds stress tensor

stresses only depend on the mean velocity gradient

$$\tau = \sum_{n=1}^{10} c^{(n)} \tau^{(n)} \tag{11}$$

In a machine learning framework, one rewrites the expansion as

$$\tilde{\tau} = \sum_{n=1}^{10} c^{(n)}(\theta, \eta) \tau^{(n)} \tag{12}$$

an L2 level assumption, and a different set of features (η) and data (θ).

5.4. Combining inference and machine learning

If machine learning is applied directly to high-fidelity data, inconsistencies may arise between the training environment and the prediction environment.

the role of the dissipation rate in a model is only to provide scale information

If the function represents mass, then **the first moment is the center of the mass(质心), and the second moment is the rotational inertia(转动惯量).**

Challenges and perspective

- What data to use?
 - Ideally, the process of calibration should provide direct indication of the need for additional data or potential overfitting.
 - the uncertainty present in the data must be accounted for during the inference and learning stages, and eventually propagated to the final prediction to set reasonable expectations in terms of prediction accuracy.
- How to enforce data-model consistency?
 - It is well known that even if DNS-computed quantities are used to completely replace specific terms in the RANS closure, the overall predictions will remain unsatisfactory
 - The addition of the inference step before the learning phase enforces consistency between the learning and prediction environment.
- What to learn?
 - A more realistic goal is to focus on learning discrepancy functions and an appropriate set of features that satisfy physical and mathematical constraints.
 - An alternative, promising approach is to focus on a specific component of a closure and introduce correction terms that can be learned from data
- What is the confidence in the predictions?
 - unavoidable dependency on the data
 - a lack of interpretability
- What is the right balance between data and models
 - the decision to leverage existing model structures and incorporate/enforce prior knowledge and physics constraints is a modeling choice
 - Therefore, the modeler’s choice is dictated by the relative faith in the available data and prior model structures, physical constraints, and the purpose of the model itself(whether the model will be used in reconstruction, or parametric prediction, or true prediction)

$$\widetilde{\mathcal{M}} = \mathcal{M}(\mathbf{w}; \mathcal{P}(\mathbf{w}; \theta); \mathbf{c}(\theta); \delta(\theta, \eta); \epsilon_{\theta})$$