Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2021-0082

Vol. **30**, No. 5, pp. 1269-1289
November 2021

Ensemble Gradient for Learning Turbulence Models from Indirect Observations

Carlos A. Michelén Ströfer[1,†], Xin-Lei Zhang[1,2,†] and Heng Xiao[1,*]

[1] *Kevin T. Crofton Department of Aerospace and Ocean Engineering, Virginia Tech, Blacksburg, Virginia, USA.*
[2] *Arts et Métiers ParisTech, 8 Boulevard Louis XIV, 59046 Lille cedex, France.*

Received 16 April 2021; Accepted (in revised version) 18 August 2021

Abstract. Training data-driven turbulence models with high fidelity Reynolds stress can be impractical and recently such models have been trained with velocity and pressure measurements. For gradient-based optimization, such as training deep learning models, this requires evaluating the sensitivities of the RANS equations. This paper explores the use of an ensemble approximation of the sensitivities of the RANS equations in training data-driven turbulence models with indirect observations. A deep neural network representing the turbulence model is trained using the network's gradients obtained by backpropagation and the ensemble approximation of the RANS sensitivities. Different ensemble approximations are explored and a method based on explicit projection onto the sample space is presented. As validation, the gradient approximations from the different methods are compared to that from the continuous adjoint equations. The ensemble approximation is then used to learn different turbulence models from velocity observations. In all cases, the learned model predicts improved velocities. However, it was observed that once the sensitivity of the velocity to the underlying model becomes small, the approximate nature of the ensemble gradient hinders further optimization of the underlying model. The benefits and limitations of the ensemble gradient approximation are discussed, in particular as compared to the adjoint equations.

AMS subject classifications: 76F99, 76M99, 65Z05

Key words: Ensemble methods, turbulence modeling, deep learning.

1 Introduction

The Navier-Stokes equations fully describe the instantaneous velocity and pressure fields in fluid flows. However, the resolution required to capture the range of turbulence scales

[†]These authors contributed equally to this work.
[*]Corresponding author. *Email addresses:* `cmich@vt.edu` (C. A. Michelén Ströfer), `xinlei.zhang@ensam.eu` (X.-L. Zhang), `hengxiao@vt.edu` (H. Xiao)

makes solving the Navier-Stokes equations computationally inaccessible for flows with high Reynolds numbers or complex geometries. Instead, the Reynolds-averaged Navier-Stokes equations (RANS) are widely used in practice thanks to the relatively inexpensive computation required for their solution. The RANS equations are a set of coupled partial differential equations (PDE) that describe the mean velocity ($\boldsymbol{u}$) and mean pressure (p) fields. However, the RANS equations contain the unclosed Reynolds stress tensor $\boldsymbol{\tau}$ which captures the effects of turbulence on the mean flow and requires modeling. The incompressible steady RANS equations are

$$\begin{aligned} &\boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}-\nu\boldsymbol{\nabla}^2\boldsymbol{u}+\boldsymbol{\nabla}\cdot\boldsymbol{\tau}+\nabla p-\boldsymbol{s}=\boldsymbol{0},\\ &\boldsymbol{\nabla}\cdot\boldsymbol{u}=0, \end{aligned} \tag{1.1}$$

where $\boldsymbol{s}$ are the external body forces. Compactly, this can be written as $\mathcal{M}(\boldsymbol{u},p;\boldsymbol{\tau})=0$ where the Reynolds stress $\boldsymbol{\tau}$ requires turbulence modeling.

Although widely used, RANS predictions are known to be inaccurate due to the lack of an accurate general turbulence model. In particular, the widely used linear eddy viscosity models (LEVM) are known to be inaccurate even in simple flows of practical interest, including the inability to predict secondary flows in non-circular ducts [1]. Eddy viscosity models are single-point closures that represent the Reynolds stress as a local function of the velocity gradient. Non-linear eddy viscosity models (NLEVM) can capture more complex non-linear relations between the velocity gradient and the Reynolds stress, but existing NLEVM have not resulted in consistent improvement over LEVM. This has led to an interest in developing data-driven turbulence models [2]. Particularly, data-driven NLEVM [3] have recently gained much attention.

Data-driven NLEVM have been typically trained with full field Reynolds stress data from high fidelity simulations. It has been recently recognized, however, that the use of high fidelity Reynolds stress data for training can be impractical, which has led to the use of measurements derived from the velocity and pressure fields as training data [4,5]. This allows for the use of more complex flows for which solutions of the Navier-Stokes equations are not feasible but for which experimental data is available. Training the model using such data has the added complexity of requiring solving the RANS equations at each training step and, for gradient-based optimization, obtaining the gradient of the RANS equations. In this work we explore the use of ensemble-based derivative approximations as an alternative to adjoint models for gradient-based training of data-driven turbulence models from indirect observations.

1.1 Data-driven eddy viscosity models trained with indirect observations

The representation of the turbulence model and the training framework in this work are the same as in [4] except for the use of the ensemble gradient in place of the adjoint. This framework is summarized here.