Creating Data-Driven Turbulence Models Using PIV

Steven J. Beresh,* Nathan E. Miller, Eric J. Parish, Matthew F. Barone, and Jaideep Ray

Sandia National Laboratories, Albuquerque, NM, 87185, USA
*: Corresponding author: P.O. Box 5800, Mailstop 0825, (505) 844-4618, email: sjberes@sandia.gov

Keywords: PIV, Data Assimilation

ABSTRACT

Data-driven turbulence models have been created for a $k$-$\varepsilon$ RANS model based on PIV data in a jet in crossflow. In the simpler of two implementations, the nominal value of the model coefficient $C_\mu$ was replaced with an optimized value calibrated to the PIV results. Despite being based on only four flow cases of a canonical configuration, the optimized model demonstrated superior performance over 48 flow cases of increasing complexity. A second, more sophisticated data-driven model has been created by mapping a spatially variable $C_\mu$ to flow state variables using machine learning of experimentally measured flow field properties of the turbulence. This second model has been implemented in a production RANS code but requires further improvements before it can return results deviating from either the nominal or calibrated $C_\mu$ models.

1. Introduction

Most practical Computational Fluid Dynamics (CFD) simulations continue to rely on Reynolds-Averaged Navier Stokes (RANS) approaches, including the jet interaction applications upon which the present effort is focused. Although Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS) produce greatly superior results in terms of accuracy (e.g., [1]), RANS remains the predominant means of providing engineering predictions for the simple reason that it is much less computationally expensive. For engineering campaigns that may require simulation of scores or even hundreds of cases, no alternative to RANS is presently feasible. Even as hybrid RANS/LES methods mature, turbulence closure models remain an inherent component of the simulation.

One of the most utilized RANS turbulence models is the $k$-$\varepsilon$ model, which introduces coupled conservation equations for the turbulent kinetic energy ($k$) and the dissipation rate ($\varepsilon$) of that energy [2, 3]. This model then uses the turbulent-viscosity hypothesis to approximate the unresolved Reynolds stresses ($\overline{u_i u_j}$) using the mean rate-of-strain tensor via the Boussinesq approximation. This typically takes a linear form as

$$2\nu_t S_{ij} = -\overline{u_i' u_j'} + \frac{2}{3} k \delta_{ij}$$  \hspace{1cm} (1)
where the overbars represent Reynolds averages, \( \nu_t \) is the scalar eddy viscosity, \( S_{ij} = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \) is the mean rate-of-strain tensor, and \( \delta_{ij} \) is the Kronecker delta, though non-linear methods have also been proposed and used (e.g., [4]). One of the primary parameters in any \( k \)-based RANS model, including \( k-\varepsilon \), is the model coefficient \( C_\mu \), which is used to define the eddy viscosity based on \( k \) and \( \varepsilon \) as

\[
\nu_t = \frac{C_\mu k^2}{\varepsilon}
\]  

(2)

This relationship is a key link between the evolution of the turbulent kinetic energy and the conservation equations for momentum and energy, and thus the specification of \( C_\mu \) is pivotal in the fidelity and robustness of any RANS simulation. The value of \( C_\mu \) and the other coefficients of the \( k-\varepsilon \) model ordinarily are presumed to be universal for a wide variety of canonical flows.

It may be recognized that all the terms to determine the eddy viscosity in Eqs. (1) and (2) are, in principle, accessible via modern PIV technology, and in fact eddy viscosity has been measured experimentally [5-8]. The dissipation rate, \( \varepsilon \), is more challenging to estimate given that it occurs at small scales that may lie below the spatial resolution of PIV, but nonetheless several methods have successfully estimated values of \( \varepsilon \) (e.g., [9-12]). Therefore, a well-crafted PIV experiment can measure the properties that underlie turbulence models for at least a portion of the flow field, and thereby produce means of adjusting the value of \( C_\mu \) or other modeling parameters to potentially improve performance in a selection of flows.

Meanwhile, the advancement of PIV capability, especially volumetric measurements, has leveraged fluid dynamics governing equations and numerical methods at the heart of CFD, an approach typically labelled as data assimilation. The rich data fidelity of time-resolved PIV in concert with volumetric methods, usually Lagrangian Particle Tracking (LPT), provides higher-fidelity velocity fields determined using fits to governing equations such that the results must reflect what is physically permitted. Moreover, the pressure field is determined through solution of the momentum equation as an inherent part of solving for the flow field that best fits the measured velocity data. Such methods were initiated from the seminal vortex-in-cell (VIC) technique based on a vorticity transport equation [13], expanded to include minimizing material derivative disparities in VIC+ [14] or additional continuity constraints in VIC# [15]. The related code FlowFit, tailored to LPT volumetric methods, minimizes incompressible velocity field divergence in both space and time to obtain reconstructions of the velocity and pressure fields onto regular grids [16, 17].

These data assimilation methods as applied to PIV greatly boost the fidelity of the information returned, but they suffer from a fundamental restriction: they apply only to the test case measured. They cannot return data predictive of the flow as governing parameters or geometries are altered. In contrast, RANS models offer a generalized predictive capability for a class of flows – but they suffer from the well-known
20th International Symposium on the Application of Laser and Imaging Techniques to Fluid Mechanics • LISBON | PORTUGAL • JULY 11 – 14, 2022

Can the performance of RANS models be improved through use of PIV for determination of the turbulence model parameters based on the real physics of an experiment? The present insights suggest that PIV data may be used to determine the eddy viscosity and the dissipation rate and thus also $C_\mu$. This maintains the architecture of existing RANS closure models, thereby expanding their functionality to data-driven models without necessitating creation of a new computational structure. With $C_\mu$ calculated directly from relevant experimental data, RANS models developed on simpler canonical flows may be adapted to offer improved predictive accuracy on a targeted class of flows.

2. Previous PIV Experiments

The PIV data are of a jet-in-crossflow configuration which has been studied and reported on for many years [18-21] and is depicted in Fig. 1. The reader is referred to the cited works for finer details on the experiments while only a brief synopsis is presented here.

Experiments were conducted in a Mach 0.8 flow through the $305 \times 305$ mm test section of Sandia National Laboratories’ Trisonic Wind Tunnel (TWT). A supersonic jet was exhausted transversely into the flow from a Mach 3.73 nozzle positioned upstream of the region from which data were collected and mounted into one wall of the tunnel’s test section. The conical nozzle was designed with an expansion half-angle of $15^\circ$ with an exit diameter of 9.53 mm. The ratio of the dynamic pressure of the jet to that of the freestream...
was nominally $J = 10.2$ but varied as a test parameter. The freestream Reynolds number based on the jet exit diameter was $2 \times 10^5$.

Stereoscopic PIV data were collected on two orthogonal planes downstream of the jet nozzle. The first plane was the wall-normal $x$-$y$ plane oriented along the jet centerline ($z = 0$), perpendicular to the wall. The second plane was a $y$-$z$ plane oriented orthogonally to the freestream, positioned downwind of the nozzle at $x = 219$ mm in one experiment and $x = 321$ mm in another.

The PIV data were processed using LaVision’s DaVis v8.4. A total of 1500 vector fields were acquired for the $x$-$y$ plane, while 9600 independent fields were used for the $y$-$z$ plane. Optimal PIV image interrogation techniques were necessary to determine eddy viscosity and dissipation rates with sufficient precision, which are detailed in Beresh et al [22] and Miller and Beresh [23].

3. Calibrated RANS Model

Due to the complex flow topology of a turbulent jet in crossflow, RANS models built upon simpler canonical flows have struggled to quantitatively reproduce experimental measurements of the richer physics of the jet in crossflow. The most straightforward approach to incorporating PIV data into a RANS model is to recast the value of $C_\mu$ from its nominal value of 0.09 to create better performance for this targeted class of flows. With this in mind, Ray et al [24] used data from four cases of the transonic jet-in-crossflow experiment studied here, along with a Bayesian optimization, to determine refined values for the $k$-$\epsilon$ model coefficients that better aligned the primary vortex pair in the CFD results to that measured in the PIV. The improvement in the induced vortical velocity along the centerline of the interaction is evident in Fig. 2. This was made possible by use of a specific value for $C_\mu$ of 0.1025 rather than the broadly accepted 0.09.

In a follow-up paper, Ray et al [25] applied the same calibration procedure to a wider range of experimental jet-in-crossflow data and again determined that a value of $C_\mu = 0.1025$ was appropriate for these flows.

Miller et al [26] examined the performance of the calibrated RANS model over a much larger parameter space. A total of 48 test cases from various Beresh et al publications were tested, involving a variety of jet interaction strengths, nozzle inclinations, and measurement stations. A flight vehicle configuration incorporating spin rockets was included as well. Figure 3 shows an example of the results, here for a metric of the mean-squared error of the simulation with respect to the PIV velocity field, which is essentially an overall picture of the agreement of the RANS computation and the PIV. Nearly every case shows a dramatic reduction in the error when using the optimal calibrated model ($C_{opt}$) as opposed to the nominal model ($C_{nom}$). Five additional metrics were considered in Miller et al and all but one showed similarly strong improvement.
Remarkably, this recalibration of $C_\mu$ based on only four cases of a transverse jet yielded accuracy improvements in nearly every case among the 48. Even simulations of a flight vehicle configuration with substantial complexity beyond the canonical jet-in-crossflow improved markedly in its comparison to PIV data [27] (the pink five-pointed stars in the figure).

These efforts demonstrate that if a universal value for $C_\mu$ exists and is applicable for this jet-in-crossflow interaction, small changes in its value produce large variation in the behavior of the simulations, and that the resulting improvements are sufficiently robust to aid cases well beyond the canonical geometries from which it was calibrated. This indicates that a concerted effort to refine the value of $C_\mu$ may pay great dividends. The Ray et al approach to recalibrating $C_\mu$ merely uses the mean velocity field from PIV. But of course modern PIV techniques are capable of delivering much higher fidelity data, and therefore more sophisticated approaches may be considered.
4. Variable $C_\mu$ RANS Model

A more sophisticated approach to determining $C_\mu$ is possible using the spatial fidelity of PIV. If all the values in Eqns. (1) and (2) are available from the PIV, $C_\mu$ may be calculated directly from the PIV data. In this manner, $C_\mu$ can be optimized locally in a flow rather than just an optimal constant throughout the flow as in the calibrated model of Section 3.

The details of this approach are lengthy and provided in Miller and Beresh [23], but they are briefly described here. The stereoscopic planar PIV data provided all six unique components of the Reynolds stress tensor and thus all components of the anisotropy tensor (right side of Eq. 1). The data only provided three of the unique components of the mean rate-of-strain tensor (left side of Eq. 1) because out-of-plane gradients are not resolved by planar PIV. The absent terms could be approximated by leveraging the orthogonal data planes as shown in Fig. 1 and employing reasonable assumptions such as divergence-free flow and negligible out-of-plane gradients along the centerline. With all terms in Eq. (1) thusly present, the scalar eddy viscosity $\nu_s$ was found via a least-squares fit across all components of the tensors. It is shown in Fig. 4 on both $x$-$y$ and $y$-$z$ planes.

Direct calculation of $C_\mu$ also requires the determination of $k$ and $\varepsilon$ from the PIV data, as per Eq. (2). The turbulent kinetic energy $k$ was readily available from the Reynolds stress data and its determination did not require any complex methods. It is given in Fig. 5. The dissipation rate $\varepsilon$ was estimated using its relationship to Kolmogorov’s similarity hypotheses and their application via the second-order longitudinal structure function [28]. These hypotheses can be simplified for application to a single axis in which the mean-square differences of velocities separated by a variable distance $r$ along a coordinate axis are expected to vary as $r^{2/3}$ when $r$ is within the inertial subrange. This relationship remains robust even when approaching the inertial subrange. Conveniently for present purposes, the scaling parameter is a known constant multiplied with the dissipation rate $\varepsilon$. This relationship has been exploited to calculate $\varepsilon$ in a variety of applications (e.g. [29, 30]) and has been shown to be a stable and accurate way for estimating dissipation with stereoscopic PIV data [9]. The present $\varepsilon$ fields are shown in Fig. 6.

At this point, a variable $C_\mu$ is now available at all spatial locations at which PIV data were acquired. However, simulation of an entire flow field requires $C_\mu$ at every point. To bridge this gap, a functional mapping between $C_\mu$ and the flow state variables is achieved using deep neural networks to train the algorithm via machine learning based on the available PIV data. The desired state variables that serve as model inputs are chosen from those identified by Pope [3] and include parameters such as mean strain rates and mean vorticity. The desired output is the three-dimensional field of $C_\mu$, which then provides the Reynolds stresses needed to close the Reynolds-averaged conservation equations at any point in the simulation.
Fig. 4 The eddy viscosity $\nu_t$ calculated from PIV data; (a) $y-z$ plane at $x = 219$ mm; (b) $x-y$ plane at $z=0$.

Fig. 5 The turbulent kinetic energy $k$ calculated from PIV data; (a) $y-z$ plane at $x = 219$ mm; (b) $x-y$ plane at $z=0$.

Fig. 6 The dissipation rate $\varepsilon$ calculated from PIV data; (a) $y-z$ plane at $x = 219$ mm; (b) $x-y$ plane at $z=0$. 
domain, not just those at PIV measurement locations. These state variables all are found in the PIV data and span across all test cases and spatial locations, freeing the mapping from dependance on any one particular flow condition. Once this mapping is complete, $C_\mu$ is obtainable over the entire computational domain, as shown in Fig. 7 for the case of a transverse jet. Other flow configurations are derived similarly by training to the same underlying PIV data sets but the resulting algorithm adapts to the local flow behavior such as to map $C_\mu$ to each flow topology. In this manner, a predictive capability may be generated that is generalized to jet interactions however they may be simulated rather than restricted to a single test case.

The process of determining an appropriate $C_\mu$ field for an individualized flow simulation must be incorporated into the RANS computational code itself. It is a substantial and specialized task to recraft a production $k$-$\epsilon$ simulation code to function with the present variable $C_\mu$ model, which is not described here. As this document is written, this task has been completed and verified and shown to yield a realistic jet-in-crossflow flow field, as seen in Fig. 8. However, early results demonstrate that the flow field simulated with the variable $C_\mu$ model is nearly identical to that produced by the original $k$-$\epsilon$ model without any $C_\mu$ modification, shown in Fig. 9. This disappointing result has been traced to three challenges.

The first challenge is that the PIV data comprises a limited portion of the simulation domain, and thus training data is not available for large portions of the domain containing important physics (e.g., the jet
nozzle). In these locations $C_\mu$ cannot be found reliably from the data and a default $C_\mu$ is employed to avoid excessive model extrapolation in unmeasured regions or those in which the calculation “blows up” due to small values of one parameter or another. This default value, though predominantly in locations away from the core of the jet interaction, appears to wield a strong impact due to its prevalence throughout the computational domain. Preliminary results suggest that if the default $C_\mu$ is altered from the nominal value of 0.09 to the calibrated value of 0.1025 (section 3), the simulated flow field instead closely resembles the flow field predicted by the calibrated model. This supports the view that the choice of a default $C_\mu$ when a machine-learned value is unavailable tends to dominate the result.

The second challenge is that the data-driven model is designed to optimally predict an intrinsic model parameter ($C_\mu$) as opposed to an extrinsic output quantity-of-interest (e.g., velocity fields). As a result, improved predictions for output quantities-of-interest rely on the assumption that the “truth” $C_\mu$ field computed from the PIV dataset will propagate through to improved quantity-of-interest predictions. This is in direct contrast to most calibration approaches such as the earlier calibrated $C_\mu$ that was directly tuned to optimally predict the counter-rotating vortex-pair.
The final challenge pertains to data consistency. In the present approach, turbulent kinetic energy \( k \) and dissipation rate \( \epsilon \) are used to non-dimensionalize input features to the data-driven model. It is well appreciated, however, that the \( k \) and \( \epsilon \) fields predicted by RANS models differ significantly from their truth values. This inadequacy introduces a “data-inconsistency” to the formulation: the data-driven model is trained with “truth values” of \( k \) and \( \epsilon \) as extracted from PIV, but then relies on RANS-generated values of these fields when deployed in the RANS code. This inconsistency can result in erroneous predictions of \( C_\mu \) that limit the accuracy of the model (see Section 5.4 of [31]).

Novel techniques, including inverse modeling [32] and extrapolation detection [33, 34], are under development to address the above issues. These techniques directly couple training the data-driven model to the underlying RANS solver and aim to robustly classify extrapolation of the model. Thereby, more sensible choices of \( C_\mu \) may be implemented beyond the limited bounds of the experimental data. Regardless, the preliminary results such as those displayed in Fig. 9 indicate that some additional digestion of the PIV data is required before it may be incorporated into an altered RANS model with improved predictive capacity.

However, the spatially variable \( C_\mu \) model evolves, it represents a blending of CFD with modern experimental techniques to craft a more successful predictive capability. This variable \( C_\mu \) approach may be characterized as a data-driven turbulence model because the modeling constant \( C_\mu \) has been reshaped as a spatially dependent variable determined by experimentally measured flow field properties. It also may be viewed as a data assimilation technique, but unlike those that have become common for volumetric TR-PIV measurements, it yields a generalized model capable of making flow predictions as in the ordinary application of RANS, as opposed to only serving the test case actually measured.

5. Conclusions and Future Work

The present effort seeks to improve the predictive capability of RANS models by incorporating PIV data through the use of deep learning in data-driven models. The key parameters in a \( k-\epsilon \) turbulence model are the turbulent kinetic energy, the dissipation rate, and the eddy viscosity, all of which may be directly measured in a well-crafted PIV experiment. These flow field measurements allow determination of the modeling constant \( C_\mu \) in a spatially varying manner rather than assuming it a universal constant as in a nominal \( k-\epsilon \) model. This approach is anticipated to introduce improved physical fidelity into the RANS model for the class of flows that the training data represent, without requiring alteration to the existing RANS closure models or computational structure.
In a simple initial implementation, the nominal value of the model coefficient $C_\mu$ was replaced with an optimized value calibrated to the PIV results from only four flow cases of a configuration of a transverse jet exhausting from a flat plate at varied jet strength. Yet the optimized model demonstrated superior performance over 48 flow cases of increasing complexity, including a flight vehicle configuration substantially different from the canonical training data.

A second, more sophisticated data-driven model was created by mapping a spatially variable $C_\mu$ to flow state variables using machine learning of experimentally measured flow field properties of the turbulence. This permits local optimization of the model parameter rather than simply selecting an optimal constant throughout the flow field. This second model has been implemented in a production RANS code and has been shown to return physically plausible results, but these predictions closely match the results from either the nominal or calibrated $C_\mu$ models depending upon how values of $C_\mu$ are extrapolated beyond the core of the flow where PIV measurements are available. Continued development of the spatially variable $C_\mu$ model will incorporate more advanced deep learning techniques such as extrapolation detection and discrepancy models using a RANS feedback loop to more broadly estimate $C_\mu$ from the experimental training data. Then, in the same manner that the calibrated $C_\mu$ model was validated against a suite of 48 cases plus a flight vehicle configuration, the new variable $C_\mu$ model may be tested for its predictive capability of jet interactions. Whereas data assimilation techniques that have become common for volumetric TR-PIV measurements can improve the physical fidelity only of the test case actually measured, the approach under development here yields a generalized RANS model with improved physical fidelity that can be applied to predict any relevant flow case.

Acknowledgments

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References


