DISSERTATION

BAYESIAN DATA ASSIMILATION FOR CFD MODELING OF TURBULENT COMBUSTION

Submitted by

Yijun Wang

Department of Mechanical Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2022

Doctoral Committee:

Advisor: Xinfeng Gao

Milija Zupanski
Stephen Guzik
Bret Windom
Matthew Koslovsky

ABSTRACT

BAYESIAN DATA ASSIMILATION FOR CFD MODELING OF TURBULENT COMBUSTION

Achieving accurate CFD prediction of turbulent combustion is challenging due to the multiscale nature of the dynamical system and the need to understand the effect of the small-scale physical features. Since direct numerical simulation (DNS) is still not feasible even for today's computing power, Reynolds-averaged Navier-Stokes (RANS) or large-eddy simulation (LES) is commonly used as the practical approach for turbulent combustion modeling. Nevertheless, physical models employed by RANS or LES for describing the interactions between the turbulence, chemical kinetics, and thermodynamic properties of the fluid are often inadequate because of the uncertainties in the dynamical system, including those in the model parameters, initial and boundary conditions, and numerical methods. Understanding and reducing these uncertainties are critical to the CFD prediction of turbulence and chemical reactions. To achieve this, this dissertation is focused on the development of a Bayesian computational framework for the uncertainty estimation of the dynamical system. In the framework, a data assimilation (DA) algorithm is integrated to obtain a more accurate solution by combining the CFD model and available data.

This research details the development, verification, and validation of a multi-algorithm system (referred to as DA+CFD system) that aims to increase the predictability of CFD modeling of turbulent and combusting flows. Specifically, in this research, we develop and apply a Bayesian computational framework by integrating our high-order CFD algorithm, Chord, with the maximum likelihood ensemble filter to improve the CFD prediction of turbulent combustion in complex geometry. The verified and validated system is applied to a time-evolving, reacting shear-layer mixing problem and turbulent flows in a bluff-body combustor with and without C_3H_8-air combustion.

Results demonstrate the powerful capability of the DA+CFD system in improving our understanding of the uncertainties in model and data and the impact of data on the model.

This research makes novel contributions, including (i) the development of a new alternative approach to improve the predictability of CFD modeling of turbulent combustion by applying data assimilation, (ii) the derivation of new insights on factors, such as where, what, and when data should be assimilated and thus providing potential guidance to experimental design, and (iii) the demonstration of data assimilation as a potentially powerful approach to improve CFD modeling of turbulent combustion in engineering applications and reduce the uncertainties with data. Future work will focus on a performance study of the present DA+CFD system for turbulent combustion of high Reynolds numbers and understanding the uncertainty in model parameters for developing and assessing physical models based on available information.

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Dr. Xinfeng Gao, for the guidance she has given me and all that I have learned from her. Through her feedback, discussions, and passion for research, I have grown as a student, researcher, and person.

I want to thank Dr. Milija Zupanski for his valuable guidance throughout my DA studies. Your insightful feedback pushed me to sharpen my thinking on data assimilation and brought my research to a higher level.

I want to thank Dr. Stephen Guzik for his excellent help and guidance on high-performance computing.

I want to thank my doctoral committee, Dr. Bret Windom and Dr. Mathew Koslovsky, for their feedback and suggestions during the course of my studies and research.

This research was supported by Department of Defense United States Air Force (DOD-USAF-AirForce) under the award number FA9550-18-1-0057 (Dr. Chiping Li, program manager) and by National Science Foundation under the award number 1723191. I would also like to thank the Department of Defense High Performance Computing Modernization Program for their compute time and support.

I am forever indebted for the love and support from my parents and sister Keren with their encouragement and understanding along the way.

I would like to thank my colleagues Landon, Jordan, Nate S., Tyler, Justin, Noah, Scott, Nate O., Josh, Sean, Chris, Erik, and Andy for the discussions and friendship.

Last but not least, I could not have completed this research without the support of all my friends Mo and Haiming for providing happy distractions to rest my mind outside of my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background

In computational fluid dynamics (CFD) modeling of highly-turbulent combustion, the difficult question to answer is: how is the approximate solution of the Navier-Stokes equations augmented with chemical reactions related to the "true" solution over long time intervals? This is difficult because the problem at hand is unstable and chaotic. That is, for example, if the exact solution converges to an unstable stationary solution, it is most unlikely the numerical solution would converge to an approximation of this unstable stationary solution [3]. Hence, turbulence and combustion modeling in CFD is difficult and intricate. One major issue is the accuracy of the physical models, which are often employed in modeling turbulent combusting flows in engineering because direct numerical simulation (DNS) of turbulence is still not feasible even for today's computing power. In Reynolds-averaged Navier-Stokes (RANS) or large-eddy simulation (LES), the practical engineering approaches for turbulence modeling, models are used and their accuracy is often significantly influenced by model assumptions and model parameters which are further dependent on flow conditions (e.g., boundary conditions, operating conditions, geometrical configurations, and flow regimes, etc.). While models free of coefficients are desirable and recent progress has been made in this effort (e.g., Wiart et al. [4]), the development of a model that is truly coefficient free is still challenging because it is paramount to know the validity and sensitivity of the assumptions employed in the dynamic procedure to determine the free coefficient and the optimal values of the coefficient. After all, the choice of the coefficient will inevitably have an intricate nonlinear relationship with the physical model itself. A common practice in many CFD applications is to use reduced order models (ROMs) to simulate the statistical turbulence evolution in a natural dynamical system. However, these ROMs are often sensitive to flow conditions. For instance, Spalart and Rumsey [5] reported that the ratio of eddy viscosity to laminar viscosity used by the Spalart-

Allmaras turbulence model "has a huge dynamic range that makes it unwieldy" and they show that the tuning of this parameter from 5 to 5000 results in a significant difference in the predictions of Mach number. The need to tune model parameters dynamically becomes apparent when the models would drift away from the known state as time evolves due to the uncertainties in the models. Unfortunately, tuning is not a well-defined term, and it is often accomplished by hard-coded inputs or highly sensitive calibrating procedures during each simulation; considerable ambiguity exists in the choice of the parameter and how it is tuned.

We believe that sensible tuning should be based on the statistical error analysis and mathematical compatibility of the dynamical system by using all information available. This is a natural course of using data assimilation (DA) [6–9]. The uncertainty estimate is an intrinsic part of the solution in the DA algorithm. Data assimilation is unique in that it rigorously combines probability, statistics, control theory, information theory, linear algebra, and functional analysis to maximize the utility of information from prediction models and data. Since the 1960s, data assimilation has been widely used in atmospheric science and meteorology and has dramatically improved the predictability of the atmospheric and meteorological modeling systems [10–16]. By melding the simulation model (e.g., CFD model) and available data (e.g., physical experiments or high-fidelity computational simulations), data assimilation helps assess, validate, and develop physical models to improve prediction. In many cases, when there is a lack of data in regimes where physical testing is limited or does not exist, data assimilation can still help quantify the sensitivity and applicability of models and nonlinear error correlations by assimilating data from other regimes. Many existing engineering data are often not effectively exploited to quantify uncertainties and assess models, although some well-founded Kriging-based approaches are available for merging such data [17–19]. In response to assist in addressing these numerical challenges in CFD modeling of turbulent combustion for engineering problems, this dissertation aims to employ a DA algorithm to improve our fundamental understanding of the uncertainty and the CFD prediction by symbiotically integrating model and data. This dissertation details the development, verification, and validation of a multi-algorithm system that combines the CFD algorithm and observation data through Bayesian data

assimilation to increase the predictability of CFD modeling of turbulent and combusting fluid flow problems in engineering. The verified and validated system is then applied to the turbulent combustion simulations with realistic geometries to gain a better understanding of the uncertainties in the model and data, the impact of experimental data on the model for improved prediction, and provide guidance on data measurements and experimental design.

1.2 Motivations and Objectives

While the fields of atmospheric science and meteorology exercise data assimilation routinely with great success, the field of engineering is relatively new to data assimilation. A pioneering study by Gao et al. [20] first introduced DA to CFD for engineering. In recent years, with the availability of different observations (experimental and high-fidelity simulation data) and the increasing power of the computational ability, how to combine the existing simplified turbulence model and the perfect or noisy data to improve the uncertainty estimation in turbulence modeling has become an active research area in the CFD community. There are some recent works using machine learning techniques for turbulence modeling (e.g., [21–23]). Oliver and Moser [24] and Cheung et al. [25] used the Monte Carlo sampling methods for the posterior probability density functions (PDF) to quantify the uncertainty in the turbulence models. Markov-chain Monte Carlo (McMC) methods were used in Edeling et al. [26] for the statistical inference for their RANS model. A brief overview of recent studies mainly focused on DA applications to flow, turbulence, and combustion demonstrates effective use of the family of ensemble Kalman filter (EnKF) and variational DA method to improve the CFD prediction. Colburn et al. [27] applied the EnKF to estimate a 3D incompressible turbulent channel flow using the experimental data along the wall, and their DA results have at least an order of magnitude less error at the location about 20 viscous units from the wall. Their numerical method is a second-order finite-difference scheme. Using particle tracking velocimetry, Suzuki [28] employed a reduced-order Kalman filter to a 2D incompressible flow with DNS algorithm and found that the increasing feedback gain improves the replicability of the simulation to the experimental data. The DNS algorithm adopts the fractional-step method

with an overall second-order scheme. Gronskis et al. [29] explored a variational DA method by using experimental data to improve initial and inflow boundary conditions for 2D incompressible flows. Their numerical method is a sixth-order compact finite-difference scheme. Kato [30] used an ensemble transform Kalman filter to reconstruct transonic flows around a 2D airfoil flow and a 3D flow around a wing to show the effectiveness in combining experimental fluid dynamics and a second-order finite-volume method for representing complex turbulent flows. Mons et al. [31] investigated, analyzed, and assessed three DA methods, such as four-dimensional variational (4D-Var), the ensemble Kalman smoother (EnKS), and ensemble-based four-dimensional variational (4DEnVar), using examples in the construction of the unsteady 2D compressible flow past a cylinder, and provided a review and useful information on DA application to CFD. Gao et al. [20] applied the EnKF to estimate burning speed in a 2D freely propagating flame. The EnKF method has been used to examine local extinction events in turbulent flame with high-speed measurements and models in Labahn et al. [32]. Iterative EnKF (IEnKF) was used in Xiao et al. [33] to handle the nonlinearity of the applications. In He et al. [34], a continuous adjoint method was used to minimize the cost function, which is similar to the variational DA methods [9], surveys recent developments in improving the RANS modeling with the experimental observations and/or the DNS data sets. Nevertheless, the DA application to turbulent combustion flows with realistic geometry, a high-order forward model, and experimental data is still a new area. This motivates the thesis research to be dedicated to this new application.

The present work aims to develop and apply a Bayesian computational framework by integrating a high-order CFD algorithm with a DA method for improving the CFD prediction of turbulent combustion flow in complex geometry and help obtain a better understanding of the uncertainty of the dynamical system with data. Since nonlinear problems are common in engineering, particularly for turbulent combustion applications, one of the main questions of DA applications is how a DA method is appropriately chosen and assessed. Motivated by the general advantages of the ensemble-based DA methods, allowing higher computational efficiency and providing more statistical information on both analysis and uncertainty of states than the variational ones, this

dissertation focuses on applying an ensemble-based DA method to address the nonlinearity in the data assimilation for CFD simulations. Specifically, the maximum likelihood ensemble filter (MLEF), one of the deterministic ensemble filters, is implemented and adapted into the computational framework since the method is suitable for high-dimensional nonlinear applications. By minimizing the cost function, MLEF finds the maximum a-posteriori (MAP) point of the posterior probability distribution that produces a good estimate of the analysis and the uncertainty about the DA intervals. In addition, under the concept of using nonlinear iterative minimization, MLEF has its advantage on the nonlinear DA application over other standard ensemble DA algorithms [35].

The following objectives are established in this thesis research:

1. Choose an appropriate DA method by comparing EnKF, iterative EnKF, and MLEF using a convection-diffusion-reaction problem and the chaotic Lorenz 96 model.

2. Develop a DA+CFD system by implementing the MLEF to interface with our high-order inhouse CFD algorithm, Chord.

3. Validate the performance of the DA+CFD system by turbulence Couette flow based on various quality assessment methods.

4. Apply the DA+CFD system to improve the predictions of the time-evolving shear-layer mixing with CH_4-air combustion, correct the boundary conditions of turbulent flow in a bluff-body combustor, and understand the uncertainties of the turbulent flows with and without C_3H_8-air combustion in the same combustor.

The DA+CFD system consists of three major components: the forward model (i.e. CFD model), the observation data, and the DA algorithm, as shown in Fig. 1.1. The remainder of this chapter is a brief review of each component of this system.

1.3 Forward Model

Chord [36–41], a fourth-order finite-volume code, is used as the forward model. It is developed by the CFD & Propulsion Laboratory at Colorado State University and designed for achieving

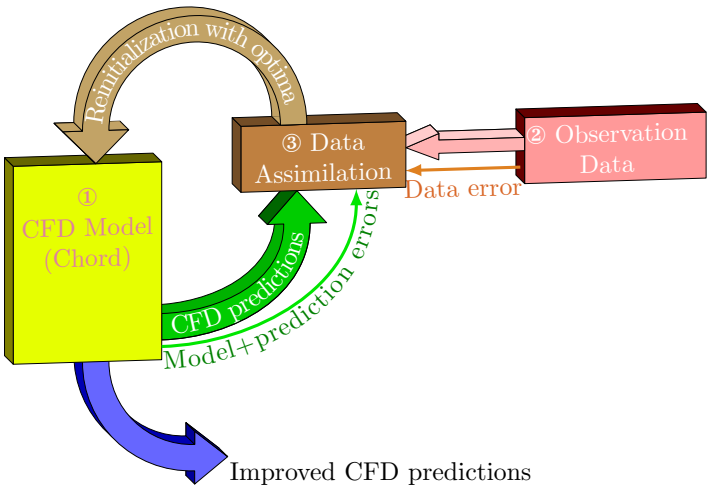


Figure 1.1: Illustration of the CFD+DA system.

high-order accuracy in time and space on modern high-performance computing architecture for modeling fluid flows with turbulence and combustion. For smooth flows, it is at least fourth-order in space and time. Chord features adaptive mesh refinement (AMR) in space and time, accommodates complex geometry while preserving free-stream conditions with mapped multiblock (MMB) technique, and scales to at least 1×10^5 cores. Chapter 3 will describe the detailed information of Chord.

1.4 Observation

In the present DA+CFD system, the errors of observation and forecast are assumed to be Gaussian. The discretization error introduced by the CFD forward model is neglected, which is considered as a "perfect" model scenario. When assuming that there is no uncertainty in the initial and boundary conditions, the physical models, and model parameters, the numerical simulation obtained by the perfect CFD model is regarded as the "truth". In the present research, observations are either synthesized from the "truth" or the literature data for the verification, validation, and DA application to the time-evolving shear-layer mixing with combustion. The physical experimental data is used for DA application to the turbulent combustion simulation in the bluff-body combustor. Chapter 9 will describe the detailed information.

1.5 Data Assimilation

Practical DA methods can be broadly categorized into variational and ensemble approaches, along with a variety of hybrid approaches that combine the variational and ensemble methods. Variational DA method, such as three-dimensional variational (3D-Var) and 4D-Var (e.g., [42,43]), utilizes optimal control theory to find the analysis state that minimizes a cost function, which is a function attempting to balance the model prediction and the data. In contrast to the variational method, the ensemble DA method, such as EnKF, EnKS, square-root EnKF, and MLEF (e.g., [35, 44–49]), is based on the statistical estimation and from the Bayesian point of view, striving to obtain the information about the mean and the covariance of the posterior probability density function (PDF) conditioned on the data. In practice, it is infeasible to determine the complete information, and instead, statistical parameters, such as the mean/mode and/or the variance of the distribution, are sought.

Regardless of classification, all DA methods can, in principle, involve one of the two different types of correction processes: filtering and smoothing. Both processes use observations from the past up to the current time to achieve the optimal estimate of a state. In filtering, the optimal analysis is achieved by directly estimating the state at the current time, while in smoothing, the state at some moment in the past is changed/optimized, which is then used to predict the state up to the current time. Mathematically speaking, in the analysis step, the filtering process only involves the observation operator, but the smoothing process involves both the prediction model and observation operators.

The core operation of data assimilation is the estimation of PDF parameters in the analysis step, which, in practice, reduces to an estimate of the mean and error covariance. Over the years in atmospheric and meteorological research, many types of variational and ensemble DA methods along with their hybrids have been developed and applied to improve the computational efficiency of this operation. Further details can be found in work by Bannister [43], which provides a thorough review of these methods, including their developments, performances, strengths, and limitations. Since ensemble DA methods, in general, are more advantageous than variational methods due to

estimating the flow-dependent forecast uncertainty, we focus on understanding the ensemble filters in the present dissertation.

Another important aspect of data assimilation is its ability to address nonlinearity when it comes to practical engineering applications. A practical approach to handle the nonlinearity in data assimilation is to apply nonlinear optimization such as gradient-based unconstrained minimization algorithms. Numerical optimization has been a major component of variational data assimilation for decades (e.g., [50–54]) and later was applied to ensemble data assimilation (e.g., [35, 49]) as well as in hybrid variational-ensemble DA methods (e.g., [55, 56]). It is well-known that one can iteratively use the Kalman filter equations to form a Gauss-Newton minimization algorithm [57]. Alternative ways to address the nonlinearity in the ensemble Kalman filters have been developed as IEnKF methods (e.g., [58–60]). These IEnKF often combine empirical arguments with particular numerical optimization algorithms, such as Newton and Gauss-Newton methods. The main difference between MLEF and IEnKF methods in addressing nonlinearity is the methodology used to find the optimal analysis solution. MLEF [35] is a straightforward application of standard optimization algorithms developed in control theory, while IEnKF methods are generally empirical methods with a limited mathematical foundation to assess their optimization characteristics. A consequence is that MLEF can benefit from the mathematical foundation of control theory that assures a minimization convergence with favorable properties [61]. For the computational cost, IEnKF methods [60, 62] are often implying the use of a smoother which requires running an ensemble of prediction models in each optimization iteration. The use of a prediction model in the optimization iterations greatly increases the cost of minimization as well as the nonlinearity, implying even more optimization iterations are needed. While it is straightforward to introduce a smoother into MLEF, the MLEF formulation used in this dissertation is based on the filtering approach. The smoother or similar approaches are being by ongoing studies in the group.

Like other standard ensemble DA methods, MLEF computes the analysis error covariance to update the forecast error covariance. In addition, similar to other standard variational DA methods, MLEF obtains the optimal analysis by maximizing the posterior PDF (e.g., MAP), practically

accomplished by an iterative minimization of a cost function. Hence, it has theoretical advantages in producing more accurate and efficient solutions in nonlinear DA problems over standard ensemble methods and updating error covariance over standard variational methods. Moreover, it does not need any adjoint calculation and uses Hessian as a preconditioner to accelerate the convergence for finding the MAP point. This present work focuses on employing the MLEF algorithm with a fourth-order accurate finite-volume CFD model to improve the CFD modeling of turbulent combustion flows with bluff-body geometry.

1.6 Outline of Dissertation

This dissertation is structured as follows. Chapter 2 describes the mathematical formulations of the problem. Chapter 3 reviews the numerical solvers used for the forward model. Chapter 4 describes and compares the differences in the mathematical formulations and numerical algorithms among the three ensemble DA methods. The integrated new DA+CFD system is summarized in Chapter 5, along with the observation operator, the observation quality control, and the initialization strategy. The quality measures of DA performance are introduced in Chapters 6. Chapter 7 presents a performance study for supporting the decision of choosing MLEF as the primary DA method in this dissertation. In Chapter 8, the verification and validation of the system are performed. The verified and validated system is then applied to canonical fluid dynamics problems. Chapter 9 delineates the configurations of the problems, procedures, analysis, and discussion of the results. Finally, the conclusions, original contributions, and future work are detailed in Chapter 10.

Chapter 2

Mathematical Modeling

The present work solves the fully-coupled, compressible, thermally-perfect, reacting Navier-Stokes (NS) equations on a structured grid to model turbulent combustion. The Navier-Stokes equations are used to model fluid flow phenomena' physics. A set of species transport equations is coupled with the NS equations for reaction modeling when combustion is considered. The computational domain is Cartesian. To model the reacting flow in realistic, complex geometries, the MMB technique is employed to represent complex geometry from the physical space onto a Cartesian grid with AMR in the computational space using the general curvilinear coordinate transformation. The present research assumes that the grid will not deform over time. The remainder of this chapter provides a review of the important elements in the system of governing equations.

2.1 Governing Equations

On a rectangular physical domain which is the same as the computational domain without applying the coordinate transformation, the system of partial differential equations (PDEs) for a compressible gas consisting of the continuity, momentum, energy, and a set of species transport equations is given by

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{u}) = 0 \,, \tag{2.1}$$

$$\frac{\partial}{\partial t}(\rho \vec{u}) + \vec{\nabla} \cdot \left(\rho \vec{u}\vec{u} + p\vec{\vec{I}} \right) = \vec{\nabla} \cdot \vec{\vec{\tau}} + \rho \vec{f} \,, \tag{2.2}$$

$$\frac{\partial}{\partial t}\left(\rho e\right) + \vec{\nabla} \cdot \left[\rho \vec{u} \left(e + \frac{p}{\rho} \right) \right] = \vec{\nabla} \cdot \left(\vec{\vec{\tau}} \cdot \vec{u} \right) - \vec{\nabla} \cdot \vec{q} + \rho \vec{f} \cdot \vec{u} \,, \tag{2.3}$$

$$\frac{\partial}{\partial t}(\rho c_n) + \vec{\nabla} \cdot (\rho c_n \vec{u}) = - \vec{\nabla} \cdot \vec{\mathcal{J}}_n + \rho \dot{w}_n \,, \quad n = 1, \ldots, N_g \,, \tag{2.4}$$

where ρ is the density, $\vec{u}$ is the velocity vector, p is pressure, $\vec{\vec{I}}$ is the identity tensor, and $e = |\vec{u}|^2/2 + \sum c_{\mathrm{n}} h_{\mathrm{n}} - p/\rho$ is the total specific energy with c_n and h_n being the mass fraction and the

specific enthalpy of species n, respectively. The pressure is given by the ideal gas law $p = \rho R_u T$, where R_u is the universal gas constant and T is the temperature. $\vec{f}$ is a body force per unit volume acting on the gaseous mixture, but its impact is negligible in the present work. $\vec{q}$ is the molecular heat flux vector. The molecular fluid stress tensor, $\vec{\vec{\tau}}$, is defined as

$$\vec{\vec{\tau}} = 2\mu \left(\vec{\vec{S}} - \frac{1}{3} \vec{\vec{I}} \vec{\nabla} \cdot \vec{u} \right) , \tag{2.5}$$

where $\vec{\vec{S}}$ is the strain rate tensor and μ is the fluid molecular viscosity. In addition, a total of N_g species comprise the gaseous mixture, with N_g transport equations. $\vec{\mathcal{J}}_n$ and $\dot{w}_n$ are the species diffusive flux and the production/destruction rate of species n, respectively. The thermally-perfect reactive mixture of gases neglects Dufour, Soret, and radiative heat transfer effects in the present work. To describe the production/destruction rate of species n, the finite-rate chemistry modeling is briefly reviewed.

2.2 Finite-rate Chemistry Modeling

For reacting flows, the finite-rate chemistry model is used for calculating the reacting source term, which is described in detail in works by Gao et al. [63–66] and Owen et al. [41,67]. For completeness and convenience, the calculation of the production/destruction rate is briefly described here. The production/destruction rate for species n is calculated from the general form of the law of mass action [68] with

$$\dot{w}_n = \frac{M_n}{\rho} \sum_{r=1}^{N_r} \left(\upsilon''_{n,r} - \upsilon'_{n,r} \right) \left(\sum_{j=1}^{N_g} \alpha_{j,r} \left[X_j \right] \right) \left[k_{f_r} \prod_{i=1}^{N_g} \left(\left[X_i \right] \right)^{\upsilon'_{n,r}} - k_{f_b} \prod_{i=1}^{N_g} \left(\left[X_i \right] \right)^{\upsilon''_{n,r}} \right] , \tag{2.6}$$

with M_n being the molar mass of species n, $[X_n] = \rho c_n M_n$ being the molar concentration of the n^{th} species, and N_r being the number of chemical reaction steps. υ' and υ'' are defined as the stoichiometric coefficient for individual products and the stoichiometric coefficient for individual reactants in each chemical reaction step, respectively. α is the third-body coefficient specified in the

chemical mechanism. The forward reaction rate, k_{f_r}, for each species is modeled by the Arrhenius equation [69] in form of $k_{f_r} = A_r T^{\beta_r} \exp(-E_{a,r}/R_u T)$, where A_r is the pre-exponential factor in the rate constant for the r^{th} reaction, β_r is temperature exponent in the rate constant for the r^{th} reaction, and $E_{a,r}$ is the activation energy for the r^{th} reaction. For reversible reactions, the backward reaction rate can be calculated by $k_{f_b} = k_{f_r}/K_{\text{eq},r}$, where the equilibrium rate is given by

$$K_{\text{eq},r} = \exp\left[\sum_{n=1}^{N_g} \upsilon_{n,r}\left(\frac{-G_n}{R_u T}\right)\right]\left(\frac{p_{\text{atm}}}{R_u T}\right)^{\sum_{n=1}^{N_g}\upsilon_{n,r}}, \tag{2.7}$$

where the net of the stoichiometric coefficients is calculated by $\upsilon_{n,r} = \upsilon''_{n,r} - \upsilon'_{n,r}$ and the Gibb's free energy, G_n, is defined by

$$\frac{G_n}{R_u T} = \frac{H_n}{R_u T} - T\frac{S_n}{R_u}. \tag{2.8}$$

In the equation, H_n and S_n are the molar specific enthalpy and the molar specific entropy of the n^{th} species, respectively. Specific reaction kinetics are employed in this dissertation for modeling the combustion of hydrocarbon fuels, i.e., methane and propane.

2.3 Chemical Kinetics

In this dissertation, the combustion of two hydrocarbon fuels – methane (CH_4) and propane (C_3H_8) are considered. The non-premixed CH_4-air combustion is modeled by a 12-species, 38-steps reaction mechanism developed by Xu et al. [2]. The C_3H_8-air combustion is premixed with an equivalence ratio of ϕ = 0.65 and modeled by the chemical mechanism developed by Zettervall et al. [1] ("Z66"), including 25 species and 66 reactions. For convenience, both reaction mechanisms are listed in Appendix A.

2.4 Grid Mapping

This thesis focuses on applying data assimilation to improve the predictions of turbulent combustion flows in realistic geometry. In order to model the reacting flow in complex geometry, the MMB technique accommodates complex geometry to work with structured AMR grids [70]. The transformation grid metrics, N^{T}, and metric Jacobian, J, are introduced into the system of governing equations to transform from physical space, $\vec{x}$, into computational space, $\vec{\xi}$, where $\vec{\xi} = \vec{\xi}(\vec{x})$. They are defined as

$$\mathrm{N} = J\left(\vec{\nabla}_x\vec{\xi}\right)^{\mathrm{T}}, \quad \mathrm{N}^{\mathrm{T}} = J\vec{\nabla}_x\vec{\xi}, \quad J \equiv \det\left(\vec{\nabla}_\xi\vec{x}\right). \tag{2.9}$$

After applying the coordinate transformation to Eqs. (2.1) - (2.4), the governing equations on a mapped domain can be written into the vector form, given by

$$\frac{\partial(J\mathbf{U})}{\partial t} + \vec{\nabla}_\xi \cdot \left(\mathrm{N}^{\mathrm{T}}\vec{\mathbf{F}} - \mathrm{N}^{\mathrm{T}}\vec{\boldsymbol{\mathcal{G}}}\right) = J\mathbf{S}, \tag{2.10}$$

where $\mathbf{U}$ is the solution vector, $\vec{\mathbf{F}}$ is the inviscid (hyperbolic) flux vector, $\vec{\boldsymbol{\mathcal{G}}}$ is the viscous (elliptic) flux vector, and $\mathbf{S}$ is the reaction source vector. They are defined as

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho\vec{u} \\ \rho e \\ \rho c_n \end{pmatrix}, \vec{\mathbf{F}} = \begin{pmatrix} \rho\vec{u} \\ \rho\vec{u}\vec{u} + p\vec{\vec{I}} \\ \rho\vec{u}\left(e + \frac{p}{\rho}\right) \\ \rho c_n\vec{u} \end{pmatrix}, \vec{\boldsymbol{\mathcal{G}}} = \begin{pmatrix} 0 \\ \vec{\vec{\mathcal{T}}} \\ \vec{\vec{\mathcal{T}}}\cdot\vec{u} - \vec{\mathcal{Q}} \\ -\vec{\mathcal{J}}_n \end{pmatrix}, \mathbf{S} = \begin{pmatrix} 0 \\ \rho\vec{f} \\ \rho\vec{f}\cdot\vec{u} \\ \rho\dot{w}_n \end{pmatrix}. \tag{2.11}$$

The mapped stress tensor, $\vec{\vec{\mathcal{T}}}$, is given by

$$\vec{\vec{\mathcal{T}}} = 2\mu\left(\vec{\vec{\mathcal{S}}} - \frac{1}{3}J^{-1}\vec{\vec{I}}\vec{\nabla}_\xi \cdot \left(\mathrm{N}^{\mathrm{T}}\vec{u}\right)\right), \tag{2.12}$$

where the mapped strain rate tensor, $\vec{\vec{\mathcal{S}}}$, is updated by

$$\vec{\vec{\mathcal{S}}} = \frac{1}{2}\left((\vec{\nabla}_\xi \vec{u}) \left(\frac{\mathrm{N}^{\mathrm{T}}}{J} \right) + \left(\frac{\mathrm{N}^{\mathrm{T}}}{J} \right)^{\mathrm{T}} (\vec{\nabla}_\xi \vec{u})^{\mathrm{T}} \right) . \tag{2.13}$$

The mapped molecular heat flux, $\vec{\mathcal{Q}}$, is modeled by

$$\vec{\mathcal{Q}} = -\left(\kappa \frac{\mathrm{N}}{J} \vec{\nabla}_\xi T - \sum_{n=1}^{N_g} \left(h_n \vec{\mathcal{J}}_n \right) \right) , \tag{2.14}$$

where κ is the thermal conductivity coefficient and $\vec{\mathcal{J}}_n$ is the mapped mass diffusion of species n modeled by the Fourier's law, given by

$$\vec{\mathcal{J}}_n = -\rho D_n \frac{\mathrm{N}}{J} \vec{\nabla}_\xi \vec{u} \, . \tag{2.15}$$

In the equation, D_n is defined as the molecular diffusivity for species n.

Chapter 3

Forward Model: Chord

The Bayesian computational framework developed in the present work consists of three important components: the forward model, the observation data, and the DA algorithm. In this chapter, we are going to discuss the forward model, Chord. Chord [36–41] is our inhouse high-order finite-volume CFD code. Chord achieves high-order accuracy in time and space for modeling fluid flows with turbulence and combustion, features AMR in space and time, accommodates complex geometry while preserving free stream conditions with MMB technique, and performs efficient computations on high-performance computing (HPC) architectures. These essential features of Chord are described briefly in the remainder of this chapter.

3.1 Spatial Discretization

Traditionally, finite-volume methods (FVMs) have been constrained to second-order accuracy [71], and high-order methods are more complicated for the implementation and stability study. However, increasing the order of accuracy of an algorithm from second-order to fourth-order increases the convergence rate of solution error as the grid is refined in a smooth flow [72]. In this research, our forward model is based on the FVM with a fourth-order accurate approximation in space and time for smooth flows. Recall from the previous chapter, the system of governing equation on a Cartesian/computational domain in the vector form is given by Eq. (2.10). For FVMs, this general divergence form is first integrated over control volume as

$$\frac{\partial}{\partial t}\int_{V_i} J\mathbf{U}\,\mathrm{d}V + \int_{V_i}\left(\vec{\nabla}_\xi\cdot\left(\mathrm{N}^{\mathrm{T}}\vec{\mathbf{F}} - \mathrm{N}^{\mathrm{T}}\vec{\boldsymbol{\mathcal{G}}}\right) - J\mathbf{S}\right)\mathrm{d}V = 0\,. \tag{3.1}$$

On a rectangular computational domain, the cell centers are marked by the points $(i_0, ..., i_{\mathrm{D}-1}) = \boldsymbol{i} \in \mathbb{Z}^{\mathrm{D}}$. Thus, the control volumes $V_{\boldsymbol{i}}$ take the form $V_{\boldsymbol{i}} = [\boldsymbol{x}_0 + (\boldsymbol{i} - \frac{1}{2}\boldsymbol{u})\Delta x, \boldsymbol{x}_0 + (\boldsymbol{i} + \frac{1}{2}\boldsymbol{u})\Delta x]$ where $\boldsymbol{x}_0 \in \mathbb{R}^{\mathrm{D}}$ is some fixed origin of coordinates, Δx is the mesh spacing, and $\boldsymbol{u}$ is the vector

whose components are all equal to one. Representing the integrals as averages and applying the divergence theorem of Gauss, we arrive at a semi-discrete form

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle J\mathbf{U}\rangle_{\boldsymbol{i}} = -\frac{1}{\Delta x^d}\sum_{d=1}^{D}\Bigg(&\left(\langle J\vec{\nabla}_d\vec{\xi}\,\vec{\mathbf{F}}\rangle_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d} - \langle J\vec{\nabla}_d\vec{\xi}\,\vec{\mathbf{F}}\rangle_{\boldsymbol{i}-\frac{1}{2}\boldsymbol{e}^d}\right)\\ &-\left(\langle J\vec{\nabla}_d\vec{\xi}\,\vec{\boldsymbol{\mathcal{G}}}\rangle_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d} - \langle J\vec{\nabla}_d\vec{\xi}\,\vec{\boldsymbol{\mathcal{G}}}\rangle_{\boldsymbol{i}-\frac{1}{2}\boldsymbol{e}^d}\right)\Bigg) + \langle J\mathbf{S}\rangle_{\boldsymbol{i}}\,,\end{aligned} \tag{3.2}$$

where the math symbol $\langle\cdot\rangle$ indicates cell-averaged or face-averaged quantities, and the superscript d is used to denote the direction of the flux vector. A fourth-order center-differencing scheme is used to compute face-averaged quantities and gradients for flux evaluation. Some care is further taken in transforming from conservative to primitive variables to preserve fourth-order accuracy. The stencil for computing the fourth-order, face-averaged primitive variables, $\langle\mathbf{W}\rangle$, uses four neighboring cells in each coordinate direction, given by

$$\langle\mathbf{W}\rangle_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d} = \frac{7}{12}(\langle\mathbf{W}\rangle_{\boldsymbol{i}} + \langle\mathbf{W}\rangle_{\boldsymbol{i}+\boldsymbol{e}^d}) - \frac{1}{12}(\langle\mathbf{W}\rangle_{\boldsymbol{i}-\boldsymbol{e}^d} + \langle\mathbf{W}\rangle_{\boldsymbol{i}+2\boldsymbol{e}^d})\,. \tag{3.3}$$

Then, the hyperbolic (inviscid) face-averaged fluxes $\langle\mathbf{F}\rangle^d_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d}$ can be computed by $\langle\mathbf{F}\rangle^d_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d} = \mathbf{F}^d(\mathbf{W}^d_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d}) + \frac{\Delta x^2}{24}\Delta^{(d,2)}\mathbf{F}^d(\langle\mathbf{W}\rangle^d_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d})$. In addition, the viscous flux vector requires the fourth-order evaluation of the gradients of primitive variables, $\langle\vec{\nabla}\mathbf{W}\rangle_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d}$, in addition to $\langle\mathbf{W}\rangle_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}^d}$. The detail for the gradient evaluation is not duplicated here and can be found in Gao et al. [36]. Additional information for the modified stencils for the hyperbolic spatial discretization near physical boundaries, including issues on limiters and the dissipation mechanisms for strong shocks or detonation waves, can be found in this studies [37, 39, 41]. In the present research, the discretization error in the forward model is neglected, which is considered as a "perfect" model scenario.

3.2 AMR and MMB

Chord features AMR and utilizes the generalized curvilinear coordinate transformation with the MMB technique to accommodate complex geometries to use AMR effectively. As mentioned in

Chapter 1, this thesis research aims to apply the DA+CFD system to improve the CFD predictions of the turbulent flows in a bluff-body combustor (BBC) with or without C_3H_8-air combustion. To demonstrate the application of the DA+CFD system to practical geometry, the bluff-body combustor is used in this research. Figure 3.1 shows a bluff-body-stabilized premixed flame computational configuration, consisting of an equilateral triangle flame holder centered in a rectangular duct and a sudden expansion plenum for the combustor flow exhausts. At the left boundary of the duct, the inflow is the premixed C_3H_8-air mixture at an equivalence ratio of 0.65 under the temperature of 310 K, the pressure of 100 KPa, and the velocity of 16.562 m/s. The Reynolds number, based on the trailing edge length of the bluff body and the inlet air velocity, is Re_D = 38,658. At the left x-boundary of the plenum, the second inflow is specified as air under the temperature of 310 K and the velocity of 2 m/s. The total pressure at the end of the plenum is set to be 100 KPa. No-slip and adiabatic conditions are applied to the solid surfaces of the bluff body. The AMR and MMB features are shown in Fig. 3.2, and the spatial resolution is adaptively refined near the trailing edge of the bluff body where re-circulation, shearing, and mixing occur. AMR enables the computational efficiency of flows by effectively adapting the refinement in regions characterized by sharp gradients in temperature, species mass fractions, and vorticities, such as flame fronts and intense recirculating, shearing, or separation zones. The refinement criteria are based on vorticity and $c_{\mathrm{OH}} \times c_{\mathrm{CH_2O}}$, satisfying some thresholds for the non-reacting and reacting flows, respectively.

3.3 Time-marching Methods

To advance the solution, the fourth-order standard four-stage Runge-Kutta (RK4) method is used to evolve Eq. (3.2) for the cold flow. However, for reacting flow, small time-step constraints can be placed for the simulation if the combustion is stiff, for example, the premixed C_3H_8-air combustion used in the BBC case. To overcome the small chemical time step due to fast chemistry, the implicit-explicit (IMEX) time integration approach is used. Specifically, the fourth-order additive Runge-Kutta (ARK4) method is employed [73–75] in Chord for the solution's accuracy, stability, and efficiency. The ARK_4 method has shown significant computational efficiency by

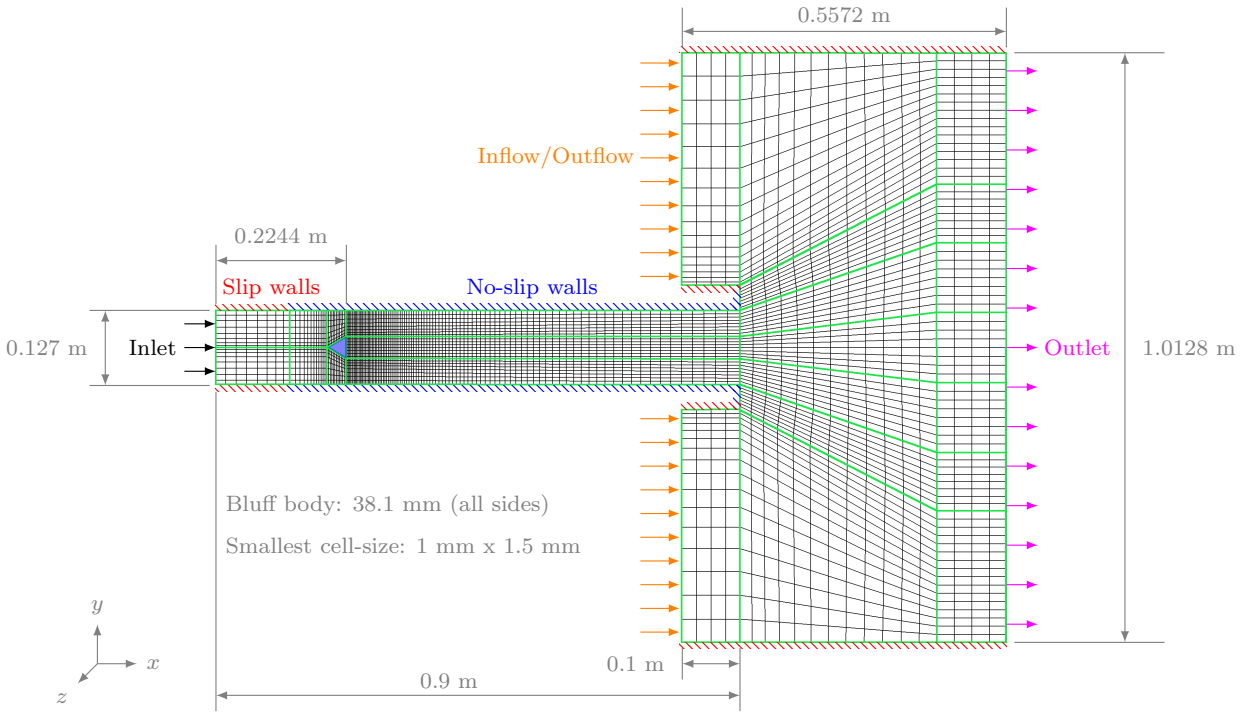
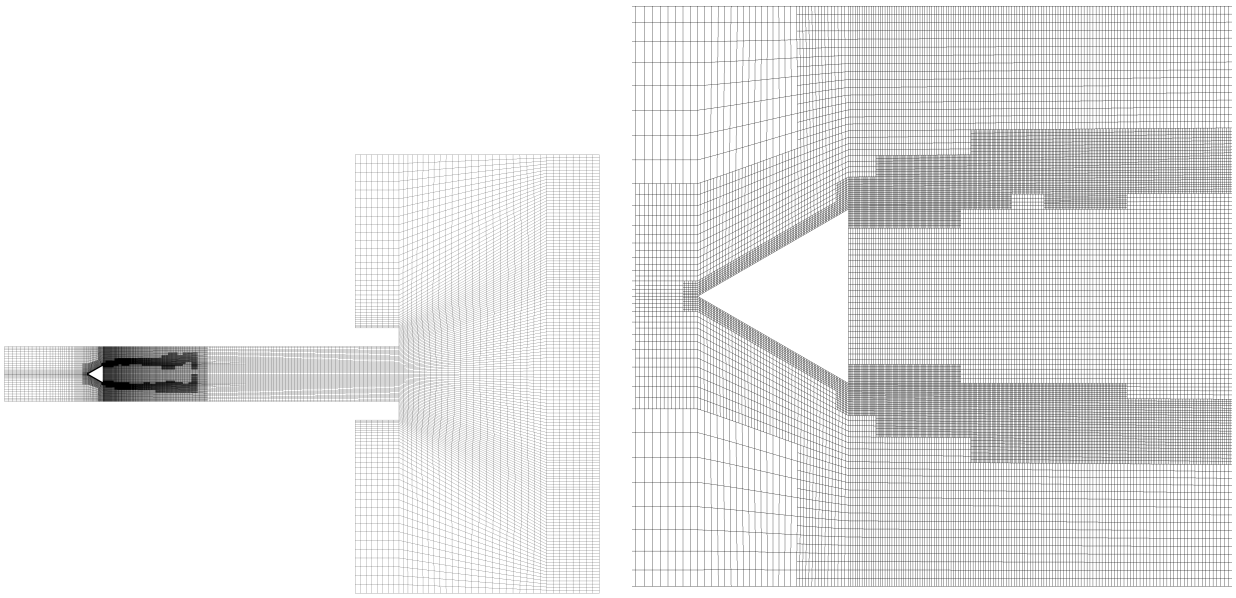


Figure 3.1: The computational domain of the bluff-body combustor. The light green boxes are multiblock in physical domain.



(a) The overview. **(b)** The zoom-in view near the bluff-body region.

Figure 3.2: An MMB grid with 3-level grid (2 levels AMR) in the physical space of BBC case. The refinement ratio of each AMR level is 2.

partitioning the physics into stiff and non-stiff terms, then treating each with implicit and explicit methods, respectively. The convection and diffusion physics are solved explicitly, while chemical reactions are solved implicitly. Christopher et al. [74, 75] reported that ARK_4 could provide a $70\times$ speedup over the standard RK4 in the two-dimensional test case, and three-orders-of-magnitude-larger time step sizes are achieved for the three-dimensional test case. ARK_4 is employed as the time marching method; however, it is not the focus of discussion in this dissertation; interested readers are referred to the above study.

Chapter 4

Data Assimilation Methods

This chapter is focused on the discussion of data assimilation. Data assimilation is the method of melding available data with a model (here, in particular, CFD) to improve the prediction. The mathematical foundation of data assimilation is based on Bayes' theorem, which provides a systematic approach for model state estimate and uncertainty assessment. To better understand the uncertainty of the dynamical system, an appropriate DA method needs to be considered to address the accuracy and efficiency of computing the error covariance correlations ($\mathbf{P}_f$) for the uncertainty estimation. This is the central point of data assimilation. One must incorporate the nonlinear dynamics and the uncertainty of the dynamics in the transient dynamical system. One must be computationally efficient when working with high-dimensional dynamical systems, such as the case of CFD modeling of turbulent combustion.

To obtain an accurate estimation of $\mathbf{P}_f$, one needs a full rank matrix over the model state space (N_s), requiring a sample consisting of N_s independent model states vectors. However, for the high-dimensional ($N_s \approx 10^9$) applications, the cost of evaluating the full rank $\mathbf{P}_f^{1/2}$ matrix becomes unaffordable in data assimilation. In recent years, ensemble-based DA methods have been increasingly used for approximating $\mathbf{P}_f$. Even the high-dimensional data assimilation becomes feasible when using the square root covariance and ensemble DA approach, since the $N_s \times N_s$ covariance matrix then reduces to $N_s \times N_e$ matrix. Given that the ensemble size, N_e, is approximately 10-100, while N_s is approximately 10^9, the dimension reduction can be considerable. Therefore, this dissertation focuses on applying an ensemble-based DA method to improve the predictions of the turbulent combustion simulations. The remainder of this chapter describes the classification of three ensemble-based DA methods, including the MLEF, the EnKF, and the IEnKF, and their characteristics. After that, the similarities and distinctions between the MLEF and EnKF methods are delineated at the concept, mathematical formulation, and numerical algorithm levels.

4.1 Ensemble Kalman Filter

We evaluate the ensemble Kalman filter (EnKF) developed by Evensen [13]. The solution procedures for applying the EnKF method include three steps: initialization, forecast, and analysis. They are described in the following. The forecast state is obtained by conducting a CFD simulation from the current analysis time to the next analysis time when the observation is available, using the current analysis state as the new initial condition for the forward model. For example, to transport uncertainties, the CFD model advances each analysis member, $\mathbf{Q}^a_{h-1}$, to get the forecast state at the h^{th} DA cycle as shown by

$$\begin{aligned}
\mathbf{Q}^f_{h,1} &= \mathcal{M}(\mathbf{Q}^a_{h-1,1})\,, \\
&\cdots \\
\mathbf{Q}^f_{h,i} &= \mathcal{M}(\mathbf{Q}^a_{h-1,i})\,, \\
&\cdots \\
\mathbf{Q}^f_{h,N_e} &= \mathcal{M}(\mathbf{Q}^a_{h-1,N_e})\,,
\end{aligned} \tag{4.1}$$

where $\mathbf{Q}$ is the model state vector, $\mathcal{M}(\cdot)$ is defined as the forward CFD model operator, and the superscripts, f and a, denote the forecast and analysis state, respectively. At the forecast step, the forecast ensemble mean ($\bar{\mathbf{Q}}^f$) and forecast error covariance ($\mathbf{P}_f$) can be calculated in the form of

$$\bar{\mathbf{Q}}^f = \frac{1}{N_e}\sum_{i=1}^{N_e}\mathbf{Q}^f_i\,, \tag{4.2}$$

$$\mathbf{P}_f = \frac{1}{N_e-1}\sum_{i=1}^{N_e}\left(\mathbf{Q}^f_i - \bar{\mathbf{Q}}^f\right)\left(\mathbf{Q}^f_i - \bar{\mathbf{Q}}^f\right)^{\mathrm{T}}. \tag{4.3}$$

The symbol T stands for the transpose. Then, each member in the ensemble is updated at the analysis step with the perturbed observations [76] as

$$\mathbf{Q}^a_i = \mathbf{Q}^f_i + \mathbf{K}\left(\mathbf{O}_i - \mathbf{H}\mathbf{Q}^f_i\right)\,, \tag{4.4}$$

where the Kalman gain matrix, $\mathbf{K}$, is given by

$$\mathbf{K} = \mathbf{P}_f \mathbf{H}^T \left(\mathbf{H} \mathbf{P}_f \mathbf{H}^T + \mathbf{R} \right)^{-1}, \tag{4.5}$$

where $\mathbf{H}$ is the linearized observation operator ($\mathcal{H}(\cdot)$) at the forecast mean ($\bar{\mathbf{Q}}^f$). $\mathbf{O}_i$ is defined as the observation vector. In some of the cases in this research, we create the vector $\mathbf{O}_i$ by synthesizing from the "truth" or the literature for the verification and validation processes, given by

$$\mathbf{O}_i = \bar{\mathbf{O}} + \boldsymbol{\epsilon}_i \,, \quad \bar{\mathbf{O}} = \mathcal{H}\left(\mathbf{Q}^t\right) + \boldsymbol{\lambda} \,, \tag{4.6}$$

$$\boldsymbol{\lambda}, \boldsymbol{\epsilon}_i \in N(0, \mathbf{R}), \quad i = 1, \cdots, N_e \,. \tag{4.7}$$

Finally, the analysis covariance matrix $\mathbf{P}_a$ can be calculated by

$$\mathbf{P}_a = \left(\mathbf{I} - \mathbf{K}\mathbf{H} \right) \mathbf{P}_f \,. \tag{4.8}$$

Evaluating the trace of $\mathbf{P}_a$ is one of the DA performance measures in this dissertation, which will be discussed in Chapter 6. The subsequent forecast is applied for running to next DA cycle,

$$\mathbf{Q}^f_{h+1,i} = \mathcal{M}_h \left(\mathbf{Q}^a_{h,i} \right), \quad i = 1, \cdots, N_e \,. \tag{4.9}$$

4.2 Iterative Ensemble Kalman Filter

The Kalman gain matrix in Eq. (4.5) applies linear update of the ensemble by linearizing the observational operator. However, this linear update process could be problematic when the observation operator is strongly nonlinear since the information in the observation data (e.g., experimental measurements) is not fully utilized during the analysis step. As described in work by Lorentzen et al. [58], this can lead to estimates that are not as accurate as they could be. In this work, we adopt the iterative ensemble Kalman filter (IEnKF) method developed by Lorentzen et al. [58]. It does not need any adjoint calculations. It addresses the nonlinearity by introducing an inner-loop

iterative process when solving Eq. (4.4) at the analysis step. Specifically, we show the concept of the inner-loop iterative solver to distinguish the difference from the EnKF method in the following. The threshold of the tolerance (ζ_{tol}) and maximum iteration integer (n) are set to be 0.02 and 12 in the present work, respectively. At an iteration k,

1. If $k = 0$, then we initialize $\mathbf{Q}_k$ and $\bar{\mathbf{Q}}_k$ as follows

   - Set each member with

$$\mathbf{Q}_k^i = \mathbf{Q}_h^{f,i}, \quad i = 1, \cdots, N_e \,. \tag{4.10}$$

   - Calculate the mean estimation in form of

$$\bar{\mathbf{Q}}_k = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{Q}_k^i \,. \tag{4.11}$$

2. Otherwise, we perform the augmented state approach as follows

   - Estimate $\mathbf{P}_k^f$ based on $\bar{\mathbf{Q}}_{k-1}$ and $\mathbf{Q}_{k-1}$ by using Eq. (4.3).
   - Set

$$\mathbf{H}_k = \frac{\partial}{\partial \bar{\mathbf{Q}}_{k-1}} \left(\mathcal{H} \left(\bar{\mathbf{Q}}_{k-1} \right) \right) \,. \tag{4.12}$$

   - Evaluate $\mathbf{K}_k$ by using Eq. (4.5).
   - Calculate $\left(\mathbf{O} - \mathcal{H}(\mathbf{Q}_k) \right)_i$ for $i = 1, \ldots, N_e$.
   - Update each member, $\mathbf{Q}_k^{1:N_e}$, by using Eq. (4.4).
   - Calculate the mean estimation, $\bar{\mathbf{Q}}_k$, by using Eq. (4.11).
   - Evaluate the L_2-norm of the ensemble mean difference by

$$\zeta_k = \sqrt{\frac{\sum_{j=1}^{N_s} \left(\bar{\mathbf{Q}}_{k,j} - \bar{\mathbf{Q}}_{k-1,j} \right)^2}{N_s}} \,. \tag{4.13}$$

- Check the following conditions

  (a) if $\zeta_k \leq \zeta_{tol}$ or $k > n$, the augmented state vector, $\bar{\mathbf{Q}}_k$, is obtained. Then, exit the loop.

  (b) otherwise, set $k = k + 1$ and repeat the above steps.

Once exiting from the inner-loop iterative solver, the analysis solution can be updated by

$$\mathbf{Q}_h^{a,\{i\}} = \mathbf{Q}_k^{\{i\}}, \quad i = 1, \cdots, N_e,$$

and then used to reinitialize the next forecast step.

4.3 Maximum Likelihood Ensemble Filter

The MLEF method developed by Zupanski [49] is considered. There are three steps in the MLEF method, including initialization, forecast, and analysis steps. The MLEF method follows the main idea of the Kalman filter (KF), in which the forecast state and the forecast uncertainty are obtained by transporting the analysis and the analysis uncertainty through a simulation model. A cost function is defined and minimized for $\hat{\mathbf{Q}}$, a control vector consisting of a vector for the solution variables and a vector for the model parameters if desired. The general derivation of the cost function is based on the Bayes formula, which finds the posterior PDF from the prior, the conditional, and the observation PDFs under Gaussian assumption [77, 78]. The cost function is defined as a negative logarithm of the posterior PDF and takes the form of

$$J(\hat{\mathbf{Q}}) = \frac{1}{2}\left[\hat{\mathbf{Q}} - \hat{\mathbf{Q}}^f\right]^{\mathrm{T}} \mathbf{P}_f^{-1}\left[\hat{\mathbf{Q}} - \hat{\mathbf{Q}}^f\right] + \frac{1}{2}\left[\mathbf{O} - \mathcal{H}(\hat{\mathbf{Q}})\right]^{\mathrm{T}} \mathbf{R}^{-1}\left[\mathbf{O} - \mathcal{H}(\hat{\mathbf{Q}})\right], \tag{4.14}$$

where $\mathbf{P}_f$ is the forecast error covariance, $\mathbf{R}$ is the observation error covariance, $\mathbf{O}$ is the observation vector, and $\mathcal{H}(\cdot)$ is again the observation operator mapping state space to observation space.

The forecast state can be obtained by conducting a CFD simulation from the current analysis time to the next analysis time, using the current analysis state as the initial condition. Similar to the

EnKF method, we define the initial state for each member using the analysis control vector ($\hat{\mathbf{Q}}^a$) perturbed by a column of square-root error covariance $\mathbf{P}_a^{1/2}$. To transport uncertainties, the CFD model advances the deterministic state ($\hat{\mathbf{Q}}^f$) and the ensemble forecast ($\mathbf{Q}^f$) using $\mathbf{P}_{a,i}^{1/2}$ (i.e., the i^{th} column of $\mathbf{P}_a^{1/2}$) at the h^{th} DA cycle as

$$\begin{aligned}
\hat{\mathbf{Q}}_h^f &= \mathcal{M}\left(\hat{\mathbf{Q}}_{h-1}^a\right), \\
\mathbf{Q}_{h,1}^f &= \mathcal{M}\left(\hat{\mathbf{Q}}_{h-1}^a + \mathbf{P}_{a,1}^{1/2}\right), \\
&\cdots \\
\mathbf{Q}_{h,N_e}^f &= \mathcal{M}\left(\hat{\mathbf{Q}}_{h-1}^a + \mathbf{P}_{a,N_e}^{1/2}\right).
\end{aligned} \tag{4.15}$$

Since the state vector $\hat{\mathbf{Q}}$ is generally multivariate, it can include variables with different orders of magnitude. It is not feasible to invert the forecast error covariance in high-dimensional problems. The MLEF algorithm starts by introducing the variable change in the cost function and the minimization process for computational efficiency. A typical change of variable is applied as

$$\hat{\mathbf{Q}} = \hat{\mathbf{Q}}^f + \mathbf{P}_f^{1/2}\boldsymbol{\psi}\,. \tag{4.16}$$

Then, we can get $\left[\hat{\mathbf{Q}} - \hat{\mathbf{Q}}^f\right] = \mathbf{P}_f^{1/2}\boldsymbol{\psi}$. The new control variable $\boldsymbol{\psi}$ is dimensionless. Correspondingly, the cost function, Eq. (4.14), is reformulated based on $\boldsymbol{\psi}$ as

$$J(\boldsymbol{\psi}) = \frac{1}{2}\boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\psi} + \frac{1}{2}\left[\mathbf{O} - \mathcal{H}(\hat{\mathbf{Q}}^f + \mathbf{P}_f^{1/2}\boldsymbol{\psi})\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{O} - \mathcal{H}(\hat{\mathbf{Q}}^f + \mathbf{P}_f^{1/2}\boldsymbol{\psi})\right]. \tag{4.17}$$

Taking the second derivative with respect to $\boldsymbol{\psi}$, we obtain

$$\nabla^2 J(\boldsymbol{\psi}) = \mathbf{I} + \left(\mathbf{P}_f^{1/2}\right)^{\mathrm{T}}\mathbf{H}_{\boldsymbol{\psi}}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_{\boldsymbol{\psi}}\mathbf{P}_f^{1/2}\,, \tag{4.18}$$

where

$$\mathbf{H}_{\boldsymbol{\psi}} = \frac{\partial \mathcal{H}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \,. \tag{4.19}$$

Refer to the Hessian matrix as

$$\nabla^2 J(\boldsymbol{\psi}) \equiv \mathbf{A}_{\boldsymbol{\psi}} \,. \tag{4.20}$$

The condition number, $K(\mathbf{A}_{\boldsymbol{\psi}})$, can be calculated by

$$K(\mathbf{A}_{\boldsymbol{\psi}}) = \frac{|\lambda_{\text{max}}|}{|\lambda_{\text{min}}|} \,, \tag{4.21}$$

where λ_{min} and λ_{max} are the smallest and largest eigenvalues of the Hessian matrix, respectively. In general, when $K(\mathbf{A}_{\boldsymbol{\psi}})$ becomes larger, more iterations are needed to minimize the cost function. To achieve a fast convergence of minimization and a reasonable estimate of analysis error covariance in the system, we introduce another variable change, referred to as the Hessian preconditioning

$$\boldsymbol{\psi} = \left[\mathbf{I} + \left(\mathbf{Z}(\hat{\mathbf{Q}})\right)^{\mathrm{T}} \left(\mathbf{Z}(\hat{\mathbf{Q}})\right)\right]^{-1/2} \boldsymbol{\xi} \,, \tag{4.22}$$

where $\boldsymbol{\xi}$ is the final control variable, and $\mathbf{Z}$ is a matrix defined as

$$\mathbf{Z}(\hat{\mathbf{Q}}) = \left(\mathbf{z}_1(\hat{\mathbf{Q}}), \mathbf{z}_2(\hat{\mathbf{Q}}), \mathbf{z}_3(\hat{\mathbf{Q}}), \ldots, \mathbf{z}_{N_e}(\hat{\mathbf{Q}})\right)_{N_o \times N_e} \,, \tag{4.23}$$

$$\mathbf{z}_i(\hat{\mathbf{Q}}) = \mathbf{R}^{-1/2}\left[\mathcal{H}(\hat{\mathbf{Q}} + \mathbf{P}_{f,i}^{1/2}) - \mathcal{H}(\hat{\mathbf{Q}})\right], i = 1, 2, \ldots, N_e \,, \tag{4.24}$$

where N_o is the number of available observation or data points. The Hessian preconditioning significantly reduces the condition number, which helps speed up the convergence rate of the iterative minimization methods.

In the minimization process, both the first derivative and second derivative with respect to $\boldsymbol{\xi}$ are required. In our application, we employ the nonlinear conjugate gradient method since it is one

of the important methods for solving nonlinear unconstrained optimization problems. The method has the advantage of not requiring the storage of large matrices, which is better suited for the high-dimensional problems as those in the present research. For example, in the problem of minimizing a function of n variables,

$$\min f(\mathbf{x}) , \quad \mathbf{x} \in \mathbb{R}^{n\times 1} , \tag{4.25}$$

where $f(\mathbf{x})$ is assumed to be a smooth function so that its gradient ∇f exists. The problem can then be solved iteratively in form of

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k , \tag{4.26}$$

where the descent direction, $\mathbf{d}_k$, can be calculated by

$$\mathbf{d}_k = -\mathbf{B}_k \nabla f(\mathbf{x}_k) . \tag{4.27}$$

In the method, k indicates the number of iterations, α_k is defined as the step-length value that can vary in each iteration and is determined by the methods, such as the conjugate gradient, quasi-Newton, or Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) methods, and $\mathbf{B}_k$ is either a scalar or a square matrix depending on the chosen method.

In our system, following Eq. (4.26), the control variable $\boldsymbol{\xi}$ is iteratively updated by

$$\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \alpha_k \mathbf{d}_k . \tag{4.28}$$

The descent direction ($\mathbf{d}_k$) can be computed by using the quasi-Newton method as

$$\mathbf{d}_k = -\mathbf{S}_k \nabla_{\boldsymbol{\xi}} J(\hat{\mathbf{Q}}_k) , \tag{4.29}$$

where $\mathbf{S}_k$ is defined as the inverse Hessian matrix. At $k = 1$, it can be calculated as

$$\mathbf{S}_1 = \left(\nabla^2_{\boldsymbol{\xi}} J(\hat{\mathbf{Q}}_1) \right)^{-1} . \tag{4.30}$$

For $k > 1$, $\mathbf{S}_k$ can be updated in form of

$$\mathbf{S}_k = \left(\mathbf{I} - \boldsymbol{\lambda}_{k-1} \boldsymbol{\beta}_{k-1} \boldsymbol{\omega}^{\mathrm{T}}_{k-1} \right)^{\mathrm{T}} \mathbf{S}_{k-1} \left(\mathbf{I} - \boldsymbol{\lambda}_{k-1} \boldsymbol{\beta}_{k-1} \boldsymbol{\omega}^{\mathrm{T}}_{k-1} \right) + \boldsymbol{\lambda}_{k-1} \boldsymbol{\omega}_{k-1} \boldsymbol{\omega}^{\mathrm{T}}_{k-1} , \tag{4.31}$$

where

$$\boldsymbol{\lambda}_{k-1} = \nabla_{\boldsymbol{\xi}} J(\hat{\mathbf{Q}}_k) - \nabla_{\boldsymbol{\xi}} J(\hat{\mathbf{Q}}_{k-1}) , \quad \boldsymbol{\beta}_{k-1} = \alpha_{k-1} \mathbf{d}_{k-1} , \quad \text{and} \quad \boldsymbol{\omega}_{k-1} = \frac{1}{\boldsymbol{\lambda}_{k-1} \boldsymbol{\beta}^{\mathrm{T}}_{k-1}} . \tag{4.32}$$

Or, the descent direction ($\mathbf{d}_k$) can also be computed by using the nonlinear conjugate gradient method in form of

$$\mathbf{d}_k = \begin{cases} -\nabla_{\boldsymbol{\xi}} J(\hat{\mathbf{Q}}_k) & k = 1 \\ -\nabla_{\boldsymbol{\xi}} J(\hat{\mathbf{Q}}_k) + \beta_k \mathbf{d}_{k-1} & k \geq 2 \end{cases} , \tag{4.33}$$

where the scalar β_k is approximated by using the Fletcher–Reeves formula as

$$\beta_k = \frac{\nabla^{\mathrm{T}}_{\boldsymbol{\psi}} J(\hat{\mathbf{Q}}_k) \nabla_{\boldsymbol{\xi}} J(\hat{\mathbf{Q}}_k)}{\nabla^{\mathrm{T}}_{\boldsymbol{\psi}} J(\hat{\mathbf{Q}}_{k-1}) \nabla_{\boldsymbol{\xi}} J(\hat{\mathbf{Q}}_{k-1})} . \tag{4.34}$$

The minimization process starts with the initial step-length $\alpha = 1$ and the initial guess $\boldsymbol{\xi}_{k=0} = 0$, which means

$$\hat{\mathbf{Q}}_{k=0} = \hat{\mathbf{Q}}^{f} . \tag{4.35}$$

After the optimal value of $\boldsymbol{\xi}_k$ is obtained, the analysis deterministic state vector $\hat{\mathbf{Q}}^a$ is then calculated by

$$\hat{\mathbf{Q}}^a = \hat{\mathbf{Q}}^f + \mathbf{P}_f^{1/2} \left[\mathbf{I} + \left(\mathbf{Z}(\hat{\mathbf{Q}}^f) \right)^{\mathrm{T}} \left(\mathbf{Z}(\hat{\mathbf{Q}}^f) \right) \right]^{-1/2} \boldsymbol{\xi}_k \,. \tag{4.36}$$

Then, $\hat{\mathbf{Q}}^a$ is used to update the square-root analysis error covariance,

$$\mathbf{P}_a^{1/2} = \mathbf{P}_f^{1/2} \left[\mathbf{I} + \left(\mathbf{Z}(\hat{\mathbf{Q}}^a) \right)^{\mathrm{T}} \left(\mathbf{Z}(\hat{\mathbf{Q}}^a) \right) \right]^{-1/2} . \tag{4.37}$$

Finally, a Gaussian perturbation to $\hat{\mathbf{Q}}^a$ is made to prepare for each member in the ensemble as initial conditions for the next DA cycle,

$$\mathbf{Q}_i^a = \hat{\mathbf{Q}}^a + \mathbf{P}_{a,i}^{1/2} , \quad i = 1, \ldots, N_e \,. \tag{4.38}$$

4.4 Comparison of the Filters

The MLEF, EnKF, and IEnKF methods follow the main idea of the Kalman Filter (KF); that is that the forecast state and its uncertainty are obtained by transporting the analysis and analysis uncertainty through a simulation model (aka. the forward model). All methods involve three significant steps: (i) initialization step, (ii) forecast step, and (iii) analysis step. The main difference is in the analysis step. In the EnKF method, both the simulation model and observation operators have to be linear. The Kalman filter equation is used in the linear update process for solving the optimized state. To address the nonlinearity of the observation operator, IEnKF uses the augmented state approach and iteratively utilizes the Kalman filter equation for optimization. Unlike the EnKF and IEnKF, MLEF does not assume that either model or observation has to be linear. The simulation model and observation operator can be nonlinear. The forecast state is obtained by conducting a simulation from the current analysis time to the next analysis time, using the current analysis state as the initial conditions for the next DA cycle. To transport uncertainties, one needs

to define the initial state for each member in the ensemble as the analysis perturbed by a column of square-root analysis error covariance, which makes MLEF more suitable for high-dimensional applications. It does not need any adjoint calculation but uses Hessian as a preconditioner to accelerate the convergence for finding the MAP point.

Algorithms 1-3 present the major steps in each component for three ensemble-based DA methods, respectively. In the EnKF and IEnKF methods, observations have to be perturbed again when evaluating the observation increments with respect to each member in the ensemble. The Kalman gain matrix (**K**) is then introduced and used to update each member in that ensemble sequentially by linearizing the observation impact from the observation space into the model state space. As shown in line 18 of Algorithm 1, the EnKF method solves the optimized state by the linear update process using the Kalman filter equation in Eq. (4.4). Lines 16 and 17 of Algorithm 2 show that the IEnKF method uses the augmented state approach and iteratively utilizes Eq. (4.4) to form a minimization algorithm for addressing the nonlinearity of the observation operator. In the MLEF method, as presented in lines 20-30 of Algorithm 3, an iterative minimization solver is applied to the cost function in Eq. (4.14) for the optimization. In addition, only the optimal deterministic state vector is solved in the ensemble space (N_e), and then it is used to update the analysis error covariance matrix and the members to reinitialize the new DA cycle. However, the potential computational cost of the MLEF method is still comparable to that of the EnKF and IEnKF methods. Based on our experience in lower-dimensional problems ($N_s \leq 10^5$), they are mostly on the same order of magnitude.

Algorithm 1 EnKF Method

Step 1: Initialization

INITIAL STEP ($\mathbf{Q}_0^{1:N_e}$)
  Select the initial conditions of uncertainties.
  if (Uncertainty in model states) **then**
    Use the lagged forecast method.
  end if
  if (Uncertainty in model parameter) **then**
    Use Box–Muller transform method.
  end if
  Construct $\mathbf{Q}_0^{1:N_e}$.
end

Step 2: Forecast Step

FORECAST STEP ($\mathbf{Q}^f$,$\mathbf{P}_f$)
  Propagate $\mathcal{M}(\mathbf{Q}^{1:N_e})$ from t_{h-1} to t_h
  Calculate $\bar{\mathbf{Q}}^f$ and $\mathbf{P}_f$ by Eqs. (4.2) and (4.3)
end

Step 3: Analysis Step

ANALYSIS STEP ($\mathbf{Q}^a$,$\mathbf{P}_a$)
  Calculate $(\mathbf{O} - \mathcal{H}(\mathbf{Q}))_i$ for $i = 1, \ldots, N_e$.
  Evaluate $\mathbf{K}$ by Eq. (4.5).
  Update $\mathbf{Q}^a$ by Eq. (4.4).
  Calculate $\mathbf{P}_a$ by Eq. (4.8).
end
Next DA cycle

Algorithm 2 IEnKF Method

Step 1: Initialization

INITIAL STEP ($\mathbf{Q}_0^{1:N_e}$)
  Select the initial conditions of uncertainties.
  if (Uncertainty in model states) **then**
    Use the lagged forecast method.
  end if
  if (Uncertainty in model parameter) **then**
    Use Box–Muller transform method.
  end if
  Construct $\mathbf{Q}_0^{1:N_e}$.
end

Step 2: Forecast Step

FORECAST STEP ($\mathbf{Q}^f$,$\mathbf{P}_f$)
  Propagate $\mathcal{M}(\mathbf{Q}^{1:N_e})$ from t_{h-1} to t_h.
  Calculate $\bar{\mathbf{Q}}^f$ and $\mathbf{P}_f$ by Eqs. (4.2) and (4.3).
end

Step 3: Analysis Step

ANALYSIS STEP ($\mathbf{Q}^a$,$\mathbf{P}_a$)
  Call inner-loop iterative solver.
  Update $\mathbf{Q}^a$ with the output from last step.
  Calculate $\mathbf{P}_a$ by Eq. (4.8).
end
Next DA cycle

Algorithm 3 MLEF Method

Step 1: Initialization

INITIAL STEP($\mathbf{Q}_0^{1:N_e}$,$\hat{\mathbf{Q}}_0$)
  Select the initial conditions of uncertainties.
  if (Uncertainty in model states) **then**
    Use the lagged forecast method.
  end if
  if (Uncertainty in model parameter) **then**
    Use Box–Muller transform method.
  end if
  Construct $\mathbf{Q}_0^{1:N_e}$ and $\hat{\mathbf{Q}}_0$.
end

Step 2: Forecast

FORECAST STEP($\mathbf{Q}^f$,$\mathbf{P}_f$)
  Propagate $\mathcal{M}(\mathbf{Q}^{1:N_e})$ and $\mathcal{M}(\hat{\mathbf{Q}})$ from t_{h-1} to t_h.
  Obtain $\mathbf{Q}^f$ and $\hat{\mathbf{Q}}^f$.
  Calculate $\mathbf{P}_f^{1/2}$.
end

Step 3: Analysis

ANALYSIS STEP($\hat{\mathbf{Q}}^a$,$\mathbf{P}_a$)
  Set $\hat{\mathbf{Q}}_{k=0} = \hat{\mathbf{Q}}^f$.
  Apply change of variable $\psi_{k=0}$.
  Evaluate the Hessian preconditioner by Eq. (4.22).
  while $1 \leq k \leq n$ **do**
    if (solver == 'Quasi-Newton') **then**
      Calculate $\mathbf{d}_k$ by Eqs. (4.30)-(4.31).
    end if
    if (solver == 'Nonlinear conjugate-gradient') **then**
      Calculate $\mathbf{d}_k$ by Eq. (4.33).
    end if
    Update $\boldsymbol{\xi}_k$ by Eq. (4.28).
    Update $\hat{\mathbf{Q}}_k$ by Eq. (4.36).
    Calculate $\mathbf{Z}(\hat{\mathbf{Q}}_k)$.
  end while
  Update $\hat{\mathbf{Q}}^a$ by Eq. (4.36).
  Calculate $\mathbf{P}_a^{1/2}$ by Eq. (4.37).
  Update $\mathbf{Q}_i^a$ for $i = 1, \ldots, N_e$ by Eq. (4.38).
end

Continue the next DA cycle.
Repeat Steps 1-3.

Chapter 5

DA+CFD Framework

It is critical to couple the CFD and DA algorithms and ensures proper data/information communication between these two components in the framework. This chapter reviews some essential features established in the current DA+CFD system, including the framework structure, the observation operator, the ensemble size, the concept of observation quality control or assessment, and the DA initialization strategy.

5.1 Framework Structure

There are three steps in the DA+CFD (or "DA+Chord") system, as illustrated in Fig. 5.1, including forward model, innovation (aka. observation increment), and analysis. Note that this flow chart uses the symbol of the deterministic state of MLEF for illustration. Firstly, each member

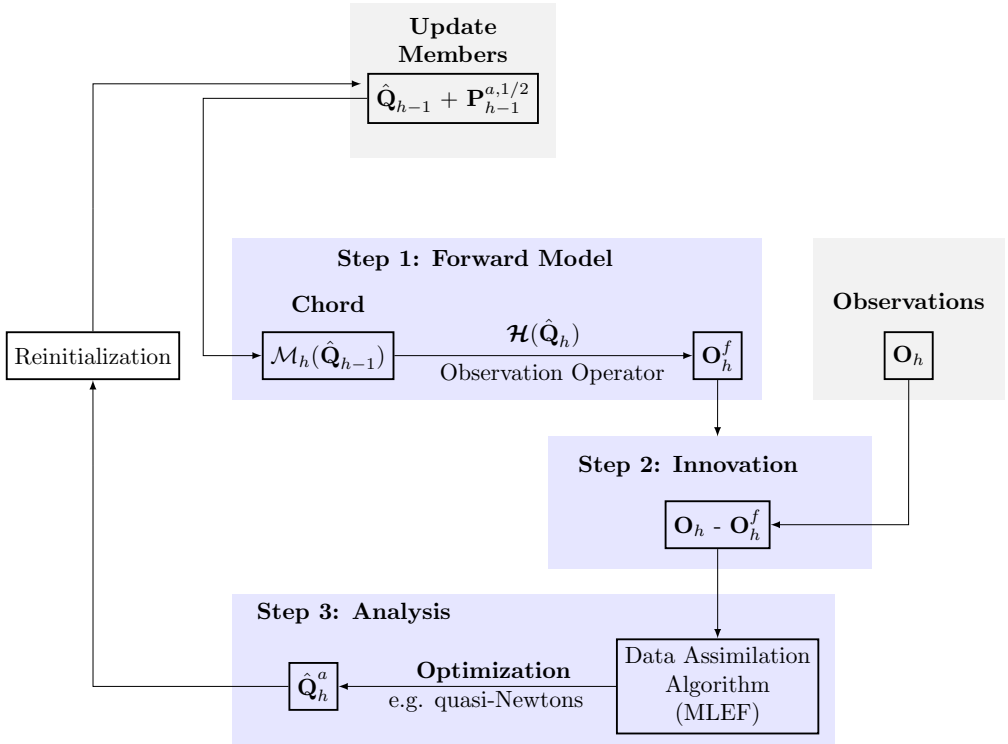


Figure 5.1: Schematic description of elements in the DA+CFD system.

in the ensemble is propagated by the forward CFD model to the time when the observation data is available, and $\mathbf{P}_f$ is evaluated at that time. The CFD model predicts the temporal and spatial evolution of the system, whose spatial and temporal scales depend on the numerical algorithm. Secondly, the observations can be measurements in space and time, whose spatial and temporal scales rely on the technique used to make the measurements in the physical experiments. When assimilating the observation data into the forward model, the innovation is calculated by the difference between the observation and the initial guess of the forecast state. Finally, the analysis combines CFD simulation and experimental data and makes the estimation of the optimal state based on the optimization criteria from statistical error analysis. Finally, the analysis state is used to reinitialize the CFD model for each member in the ensemble to propagate the forward CFD model to the subsequent DA cycles.

5.2 Observation Operator

Assimilating the observation data into the CFD model generally requires a transformation or mapping from the state space (solved by the forward model) to the observation space, accomplished through an observation operator, $\mathcal{H}(\cdot)$, since observations are not necessarily the same as the states. For example, Fig. 5.2 describes a mapping operation process of using 2D observations that are different from states and measured at spatial locations different from the computational mesh points/cells. Introduce the observation operator to be $\mathcal{H}(\cdot) = \boldsymbol{\beta}_B \cdot \boldsymbol{\beta}_G \cdot \boldsymbol{\beta}_S(\cdot)$. Matrix $\boldsymbol{\beta}_S$ is a spatial operator to average a solution variable of interest from 3D to 2D in the state space. Matrix $\boldsymbol{\beta}_G$ converts this variable (marked as black dots) to the observed variable (marked as blue dots) in the 2D computational grid. Matrix $\boldsymbol{\beta}_B$ interpolates the quantity from computational points in the black mesh to the measurement points in the red grid. Then, the innovation can be evaluated in the form of $\left(\mathbf{O}-\mathcal{H}(\hat{\mathbf{Q}})\right)$. Moreover, we use experimental measurements as observations for some of the DA+CFD applications in the current work. For example, a part of the data used for improving our prediction of the turbulent reacting flow is the signals of the OH planar laser-induced fluorescence (OH-PLIF) measurements. The signals are defined as $S_{\mathrm{OH}} = c_{\mathrm{OH}} T^{(\gamma-1)}$, where γ is

a constant and suggested to be 1.0 by Wu et al. [79], c_{OH} is the concentration of species OH, and T denotes the temperature. Since ρc_{OH} and ρ are the variables solved by our forward model, the nonlinear observation operator needs to be constructed in the form of $\mathcal{H}(\hat{\mathbf{Q}}) = \rho c_{\mathrm{OH}}/\rho$ for the c_{OH} increment calculation. If $\gamma \neq 1$, the nonlinear operator accordingly has a more complicated form.

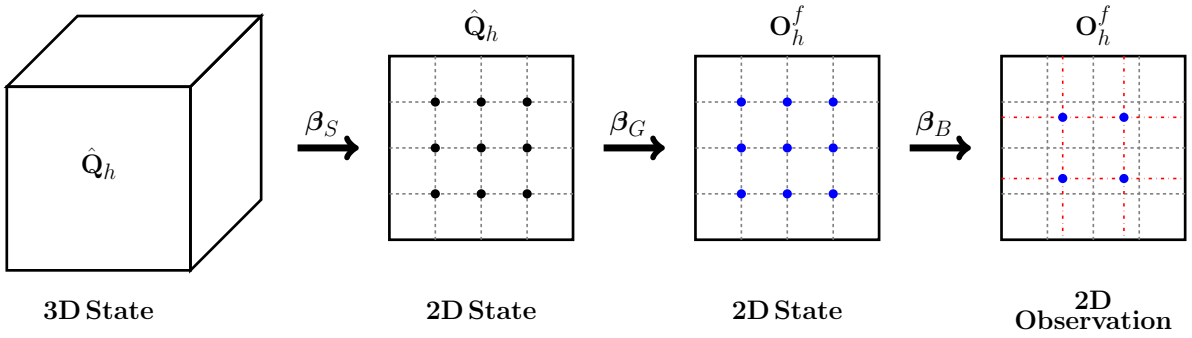


Figure 5.2: Schematic of an observation operator, $\mathcal{H}(\cdot)$.

5.3 Ensemble size

The uncertainty estimate is sensitive to the ensemble size when applying the ensemble-based DA method. In a sensitivity study, it needs to be performed first to consider both the solution accuracy and computational efficiency. A reduction of the ensemble size generally results in a large spurious correlation in the error covariance estimation. To eliminate spurious non-zero correlation, the minimum number of the ensemble required is approximately $1/(k_p\eta^2)$ in theory [80], where k_p is the probability of the non-zero correlation and η is its estimated value. For example, if we want to control the probability of the spurious non-zero correlation (e.g., $\eta \geq 0.03$) to be 5%, the ensemble size needs to be on the order of 20,000, which is infeasible in practice. In the DA community, the ensemble size is typical of the order of 10-100 for the high-dimensional dynamical systems (of the order of 10^9) applications. In addition, various covariance localization schemes have been proposed and can be applied implicitly or explicitly in the DA algorithms to suppress the spurious correlations if needed [80]. In the present DA+CFD system, we focus on applying the ensemble-based DA method to a high-dimensional system. The ensemble is used for the representation of the

PDF, and members in the ensemble are independent. Under the Gaussian assumption, the forecast error is unbiased.

5.4 Quality Control Concept

While assimilating the observations into the CFD model, an important issue is that the quality of the observation data needs to be examined since data need to be inspected for the potential unfavorable impact on the optimization process and the computational cost. A common practice for assessing the quality is calculating an error of the observational increment during data assimilation cycles. As a reminder, the observation increment is measured by $\left(\mathbf{O} - \mathcal{H}(\hat{\mathbf{Q}})\right)$. Under the assumption of Gaussian distribution for the uncertainty perturbation, the error of the observation increment should also be expected to follow the Gaussian distribution. However, the observations may be very distant from the predictions by the forward model at a certain time and space domain. This, consequently, adds difficulty and computational expense to the process of determining the optimal state by the MLEF minimizer. To remedy this, a quality region is specified in this work for accepting or rejecting observations when and where the observation increment error deviates from the thresholds. For example, observations are assimilated when the errors fall inside the quality region by assessing

$$-C_T\sigma_{\text{obs}} \leq \varepsilon_{\text{obs}} \leq C_T\sigma_{\text{obs}}\,, \quad \varepsilon_{\text{obs}} = \left(\mathbf{O} - \mathcal{H}(\hat{\mathbf{Q}})\right)\,, \tag{5.1}$$

with σ_{obs} being the expected observation error for each solution variable and C_T the coefficient of the quality control. Theoretically, in statistics, 95% of a sample that follows the Gaussian distribution falls inside the region of $\pm$ 3 standard deviation (SD). In the present work, when the observations are synthesized from the perfect model for most test cases, the quality control region is set here as $C_T = 5$, ideally permitting 99% of the data to the DA algorithm. When the observations are from the experimental measurements, we set the quality control region with $C_T = 3$, which

ideally permits 95% of the data to the DA algorithm. However, the actual rate of acceptance in each case is always lower than we expect.

5.5 DA Initialization Concept

Before any data assimilation can begin, it is necessary to define an initial state (including solution variables and empirical parameters) and the uncertainty of that initial state. In the present work, we employ the lagged forecast method [12] for the initial conditions of data assimilation, as shown in Fig. 5.3. Given the time for which the initial state and its uncertainty are needed at $t_h = 0$, a deterministic simulation of length $4t_\tau$ is started from a time t_h = -$2t_\tau$ in the past. The time frame t_τ can typically be equal to the assimilation frequency window (τ_{DA}). For an ensemble with N_e members, one runs "$N_e + 1$" CFD simulations in the MLEF method. While N_e in the EnKF family method can be equally distributed over the $4\tau_{\text{DA}}$ interval, in this approach, we use time-lagged simulations as members of the ensemble. The initial deterministic state may be set by the simulation whose start is at $t_h = 0$, or simply the ensemble mean of N_e members. We use random perturbations for the initial uncertainty of empirical parameters, currently based on a Gaussian probability density function. However, the lagged forecast method is not an inherent

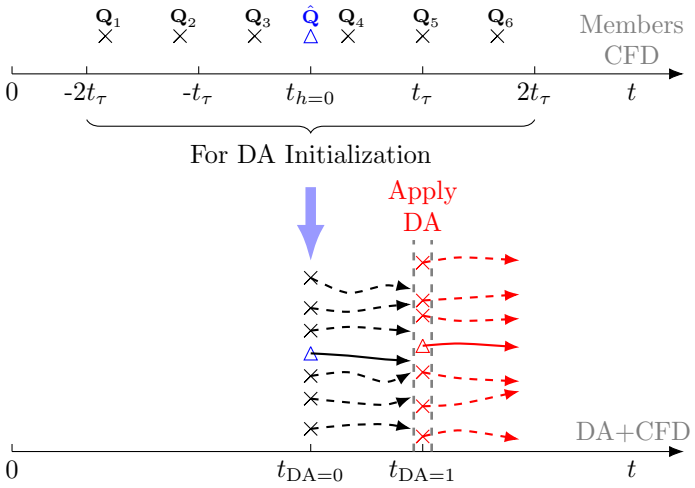


Figure 5.3: The schematic of the lagged forecast method.

component of any of these DA methods. It is chosen here because it has a good record in applications to various time-dependent problems in meteorology. When using the lagged forecast method, some spatial differences and the uncertainty correlation between the model variables and the model parameters are included as dynamical features are advected over time. Based on our experience, it can help reduce the non-physical perturbation noise when forming the initial forecast error covariance matrices in ensemble DA methods, which could benefit these methods to recover correct error covariance after only a couple of DA cycles.

Chapter 6

DA Performance Measures

The performance of the DA+CFD system needs to be assessed in the verification and validation processes and the final results for gaining a better understanding of the uncertainty. Four quality measurement methods are utilized in the present research, including the trace of the error covariance, the true error and root-mean-square (RMS) errors, the χ^2-squared diagnostic, and the normalized cost function estimation. All are prescribed in this chapter.

6.1 True Error and Root-mean-square Error

To verify the DA impact on the numerical predictions of the system, the true and RMS errors are commonly used in the context of the ensemble DA method. The true error value can be evaluated by the total difference between the "truth" and the DA+CFD prediction in the form of

$$\epsilon_t = \left(\mathbf{Q}^t - \tilde{\mathbf{Q}}\right)^{\mathrm{T}} \left(\mathbf{Q}^t - \tilde{\mathbf{Q}}\right) , \tag{6.1}$$

where $\mathbf{Q}^t$ denotes the truth vector and $\tilde{\mathbf{Q}}$ represents the mean predicted solution vector by the data assimilation. However, the "truth" is usually unknown. In some of the cases presented in this dissertation, the numerical simulation obtained by the perfect CFD model is regarded as the "truth" and used to create the observations. When the observations are from the experimental measurements, the $\mathbf{Q}^t$ term is replaced with the observation vector $\mathbf{O}$ for the calculation. The spatially averaged RMS error is calculated for each observed solution variable, labeled as the index j, by

$$\mathrm{RMS}_j = \sqrt{\frac{\sum_{m=1}^{N_{\mathrm{obs},j}} \left(\mathbf{O} - \mathcal{H}(\tilde{\mathbf{Q}}^a)\right)_m^2}{N_{\mathrm{obs},j}}} . \tag{6.2}$$

In Eqs. (6.1) and (6.2), the $\tilde{\mathbf{Q}}$ term is replaced by the analysis ensemble mean vector ($\bar{\mathbf{Q}}^a$) for the EnKF and IEnKF methods and the deterministic analysis vector ($\hat{\mathbf{Q}}^a$) for the MLEF method. Both the estimations of ϵ_t and RMS are expected to decrease as more DA cycles are performed.

6.2 Trace of the Error Covariance Matrix

The other important statistical quality assessment of the ensemble filter algorithm is to calculate the trace of the analysis error covariance matrix, trace($\mathbf{P}_a$), which indicates the filter performance on the uncertainty reduction during the DA process. Since the uncertainty error is expected to be reduced by the available observations, the variance of the members should decrease and converge to a sufficiently small tolerance as more DA cycles are performed. The expected value for trace($\mathbf{P}_a$) should be close to sufficiently small tolerance as possible at the end of the data assimilation.

6.3 χ^2-squared Test

For the validation of the DA performance, the χ^2-diagnostic test is used in ensemble data assimilation. It evaluates the correctness of the innovation covariance matrix based on the observation error matrix $\mathbf{R}$ and the forecast error covariance matrix $\mathbf{P}_f$,

$$\chi^2 = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{N_{\text{obs}}}, \quad \mathbf{X} = \left(\mathbf{R} + \mathbf{H}(\tilde{\mathbf{Q}})\mathbf{P}_f\mathbf{H}^{\mathrm{T}}(\tilde{\mathbf{Q}})\right)^{-1/2}\left(\mathbf{O} - \mathcal{H}(\tilde{\mathbf{Q}})\right), \tag{6.3}$$

where N_{obs} denotes the number of accepted observations after applying the observation quality control method, and $\mathbf{H}$ is the tangent linear model of $\mathcal{H}(\cdot)$. The χ^2-calculation evaluates the correctness of the innovation covariance matrix with respect to $\mathbf{R}$ and $\mathbf{P}_f$. The expected values for χ^2 after analysis step should be close to 1.0 as possible [49].

6.4 Normalized Cost Function

For the deterministic ensemble filter algorithm, such as MLEF, the necessary statistical verification is to estimate the normalized cost function value of the observational correction part, which

indicates the difference between the observation and the solution after the minimization process by DA method. For the MLEF method, the normalized cost function is, in fact, related to the χ^2 test [49]. The idea of using normalized cost function in addition to χ^2 is only for assessing the minimization in a multi-variate problem, where we try to achieve a similar decrease of normalized cost-function for each observation. For each solution variable j, the newly defined cost function is normalized by the number of observation points, $N_{\text{obs},j}$, and calculated by

$$\left(\frac{J}{N_{\text{obs}}}\right)_j = \frac{1}{N_{\text{obs},j}}\left(\frac{1}{2}\left(\mathbf{O}-\mathcal{H}(\hat{\mathbf{Q}})\right)_j^{\mathrm{T}}\mathbf{R}_j^{-1}\left(\mathbf{O}-\mathcal{H}(\hat{\mathbf{Q}})\right)_j\right). \tag{6.4}$$

Based on the equation, it measures the ratio of the expectation of the difference between the observation and prediction ($E\left[\left(\mathbf{O}-\mathcal{H}\left(\hat{\mathbf{Q}}\right)\right)^2\right]$) with respect to the expected observation error. The observation error is under the assumption of following the Gaussian distributed curve. The optimal state is generally determined by using both the background and observation information. At the beginning of the DA process, the background error is bigger than the observation error. The expectation value after the correction is bigger than the variance of observation error, which makes J/N_{obs} slightly bigger than 0.5 (e.g., $0.5 \leq J/N_{obs} \leq 0.8$). As more DA cycles are involved, the optimal state gets pulled closer and closer toward the observation. The expectation value is then closer to the variance of observation error ($J/N_{obs} \approx 0.5$) at the end of the data assimilation [80].

Chapter 7

The Choice of DA method

To choose an ensemble DA method for efficient and effective application to large nonlinear reacting flows, we have investigated and compared the MLEF, the EnKF, and the IEnKF methods. In this chapter, the assimilation performance of three ensemble DA methods is verified and compared by the 1D convection-diffusion-reaction (CDR) problem and the Lorenz 1996 (L96) model regarding the solution accuracy and computational efficiency. The performance study shows that MLEF demonstrates better convergence and higher accuracy. Therefore, MLEF is selected as the primary DA method in this dissertation work.

The CDR problem involves three transient physical processes of convection, diffusion, and source, but the system eventually becomes steady. The L96 is a chaotic model. The rationality of choosing the two representative problems is based on that (i) both have the "truth" which can be used to assess the assimilation performance and predication accuracy easily, and (ii) complex engineering fluid dynamics problems are often characterized by physical processes, including convection, diffusion, and source, in addition to turbulence/chaos. In most engineering applications, solution variables are not always observed, so nonlinear mapping between observation and state space is required. Both linear and nonlinear observation operators are studied using the EnKF, IEnKF, and MLEF methods. The verification process focuses on the capabilities of the three DA methods to address (i) the nonlinearity of the observation operator and (ii) the chaotic dynamics and the nonlinear physics processes. The assimilation performance is assessed based on the quality metrics, such as the squared true error, the trace of $\mathbf{P}^a$, and the RMS error.

7.1 Convection-diffusion-reaction Model

7.1.1 Problem Configuration

We start the investigation of the DA impact on the imperfect initial condition for the one-dimensional transient inhomogeneous CDR problem, which eventually has a steady-state solution. Assume that the model parameters are inaccurate and need to be estimated by data assimilation. In addition, we are interested in assessing how the adverse impact of imperfect initial conditions on predictions can be removed or mitigated by data assimilation. Specifically, the uncertain model parameters to be estimated are γ, μ, and c in the CDR equation,

$$\frac{\partial \phi}{\partial t} + \gamma \frac{\partial \phi}{\partial x} - \mu \frac{\partial^2 \phi}{\partial x^2} + c\phi = f\,, \quad 0 \leq x \leq 1\,, \tag{7.1}$$

where ϕ is the solution variable. For convenience, the true values for the convection speed, γ, the diffusivity, μ, and source term coefficient, c, are known, and they are all unity. The other source term, f, is modeled by $f = 2\,(x-1)\,e^{-t}$. Dirichlet boundary condition is applied for the inlet, and Neumann boundary condition is specified for the outlet domain extent. The computational domain size for the problem is set to be $L_x = 1$ m. A total of 128 cells are used. The analytical solution of the problem is used as the truth and defined in the form of

$$\phi^t(x,t) = x^2 e^{-t}\,. \tag{7.2}$$

Two tests are configured. The first case examines the consequence of an imperfect initial condition on the solution field. The second test investigates the situation in which the errors exist in both the initial condition and the three model parameters simultaneously. The assimilation performance for each case is assessed by comparing the solution differences between the truth and the DA+CFD predictions at the end of the DA cycles.

7.1.2 Imperfect Initial Condition in Transient-mode Problem

The first case is to assess the assimilation performance and make a comparison among the EnKF, IEnKF, and MLEF methods when the observation operator is linear. The observation operator, $\mathcal{H}(\cdot)$, is given by

$$\mathcal{H}(\phi_j) = \phi_j \,, \quad j = 1, 2, \ldots, N_{\text{obs}} \,. \tag{7.3}$$

The observations are synthesized from the perfect model at every other 4 cells by adding noise with the error distribution $\mathrm{N}(0, 0.15)$. The linear observation operator $\mathcal{H}(\cdot)$ can be written into the matrix format, as

$$\mathcal{H}(\cdot) = \begin{pmatrix} \delta & 0 & \ldots & 0 \\ 0 & \delta & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \delta \end{pmatrix}_{N_{\text{obs}} \times N_s} , \quad \delta = \begin{cases} 1 & \text{if observed} \\ 0 & \text{else} \end{cases} , \tag{7.4}$$

where N_{obs} and N_s denote the number of available observation data and the total number of the model states, respectively. The DA frequency is set to be every 5000 time steps, and the time step size is $\Delta t = 1 \times 10^{-5}$ s, that is that a DA cycle occurs every 0.05 s along with the propagation of forward model. An ensemble consisting of 8 members is used in this work. A small amount of perturbations is added into the initial CFD condition for propagating the solution using the lagged forecast method as

$$\phi(x, t = 0) = x^2 \sin(\beta_1 \frac{2\pi x}{L_x} + \beta_2) \,, \quad \beta_1 \in N(0.1, 0) \,, \quad \beta_2 \in N(0.01, 0) \,. \tag{7.5}$$

The initial members in the ensemble are presented and compared to the truth in Fig. 7.1. The DA performance is assessed by comparing the predictions of each DA method with the truth. The

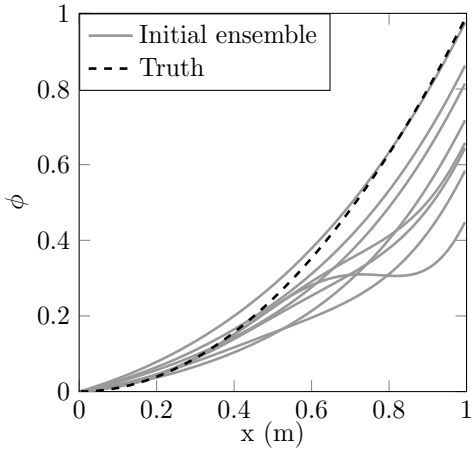


Figure 7.1: The members in the initial ensemble.

results in Fig. 7.2 show the predictions of each member and their mean estimation in 3 columns and 3 rows. For the MLEF method, we use the prediction of the control vector as its mean estimation. In each column, the predictions from different DA methods are compared to the truth at the same DA cycles. Rows show the comparisons at 5^{th}, 10^{th}, and 15^{th} DA cycles, respectively. At the 5^{th} DA cycle, the variance of the members by the MLEF method is the smallest, which indicates that MLEF has better convergence rate of the uncertainty reduction than the other two methods. In addition, as more DA cycles are performed, all the DA predictions are getting closer and closer toward the truth, which is what we expected from a proper DA process.

The performance is further assessed by measures of the true error, the trace of $\mathbf{P}^a$ matrix, and the RMS error along the DA process. By tracking the trajectories of the convergence history in Fig. 7.3a-7.3c, we clearly see that the IEnKF method shows faster convergence at the beginning of the DA process. But, the MLEF method results in a smoother and higher error reduction rate than both the EnKF and IEnKF methods along the DA process. Moreover, we calculate the absolute solution difference, $\|\Delta\phi\| = \left\|\phi^{\text{DA}} - \phi^t\right\|$, over the computational domain after 15 DA cycles. The MLEF prediction shows the smaller difference in the results among the 3 DA methods, as shown in Fig. 7.3d.

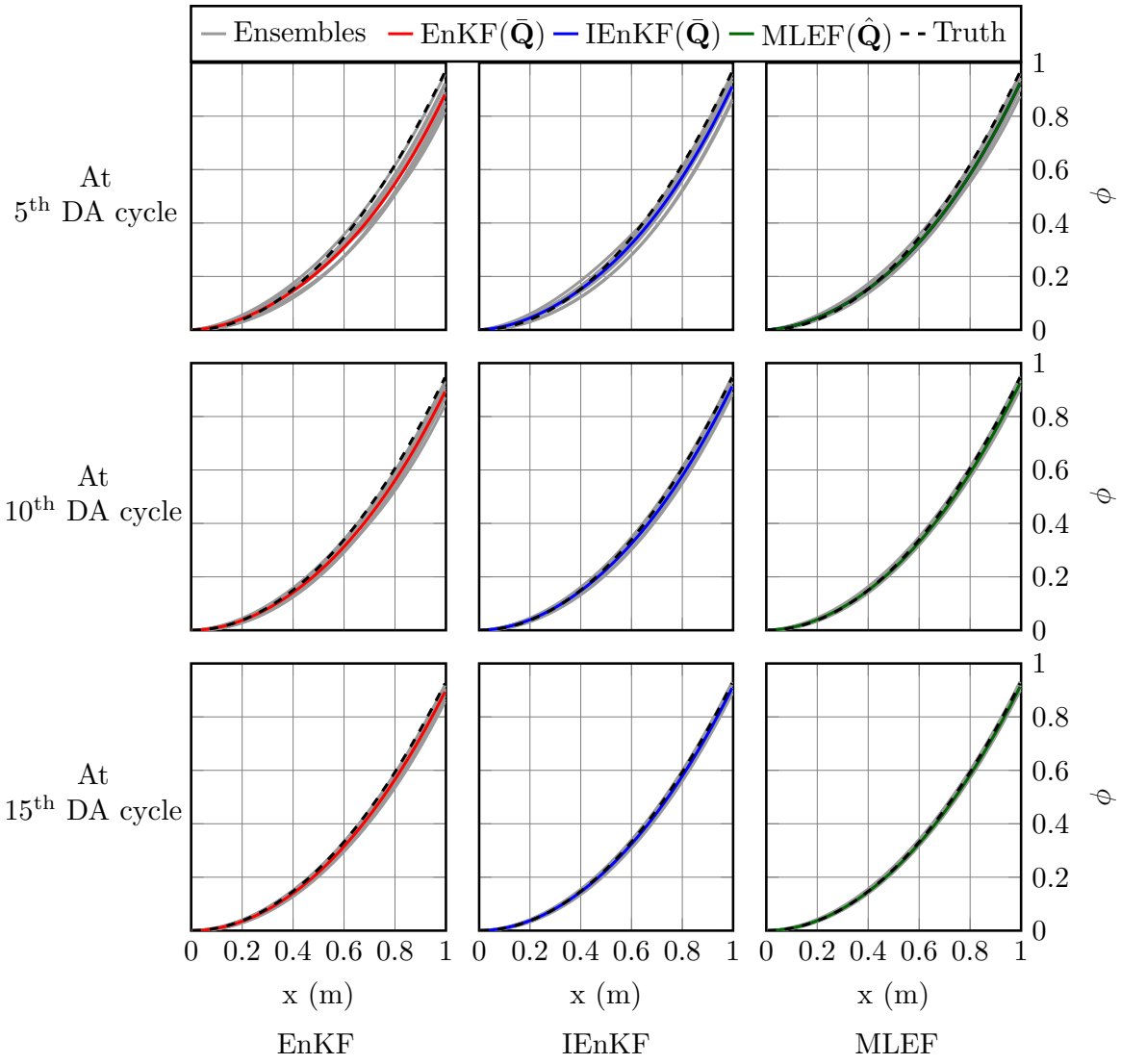
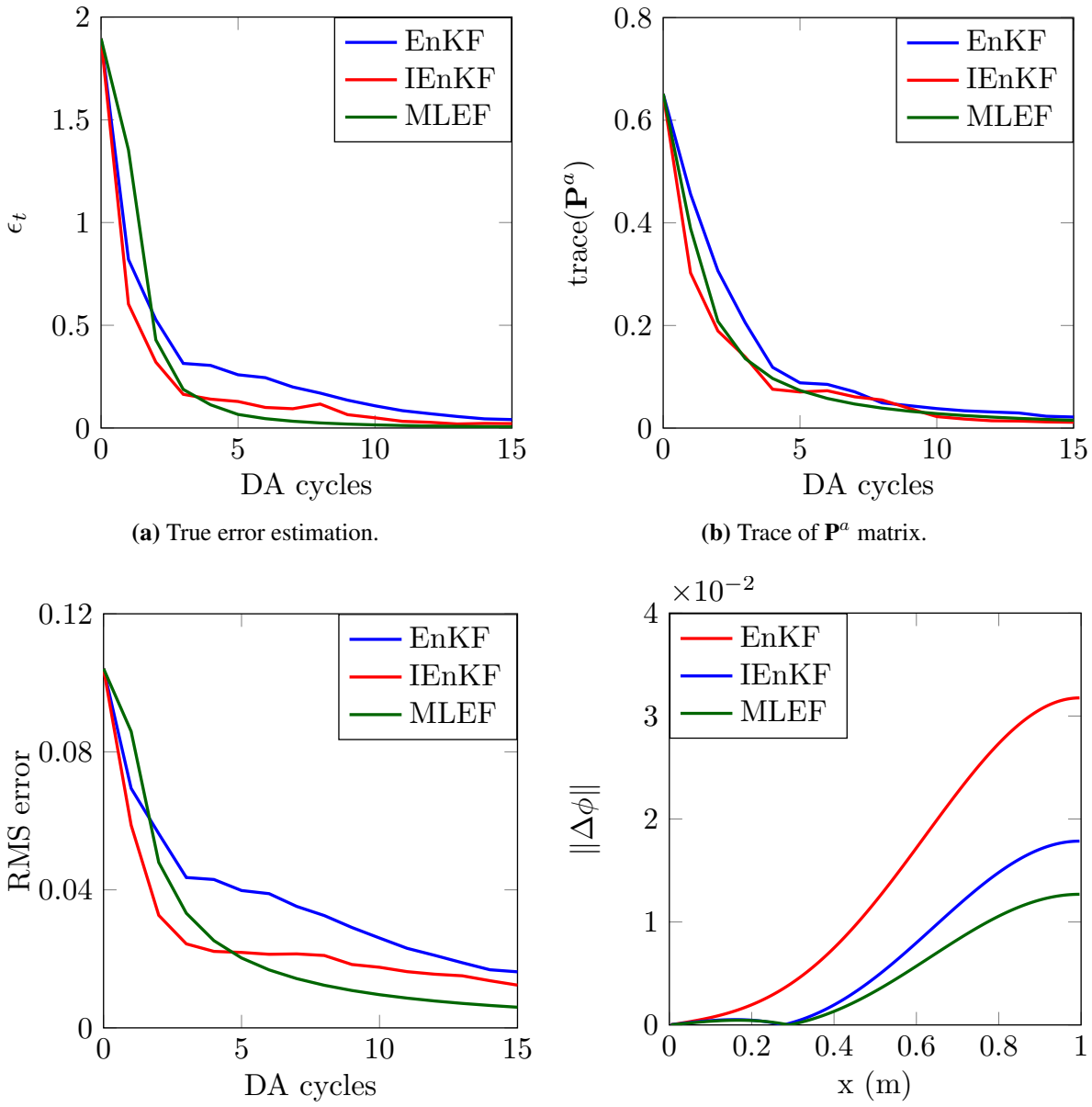


Figure 7.2: Solution comparisons among the EnKF (1st column), IEnKF (2nd column), and MLEF (3rd column) methods at the 5th, 10th, and 15th DA cycles.

Furthermore, we investigate the performance of the MLEF and IEnKF methods for the case of using a nonlinear quadratic observation operator, where $\mathcal{H}(\cdot)$ is given by

$$\mathcal{H}(\phi_j) = 2.5\phi_j^2 + 1.5\phi_j\,, \quad j = 1, 2, \ldots, N_{\text{obs}}\,. \tag{7.6}$$

The DA setups are the same as those in the above case; that is the same in the assimilation frequency and the initial ensemble for the initial condition. In this study, the assessment of the perfor-



(a) True error estimation.

(b) Trace of $\mathbf{P}^a$ matrix.

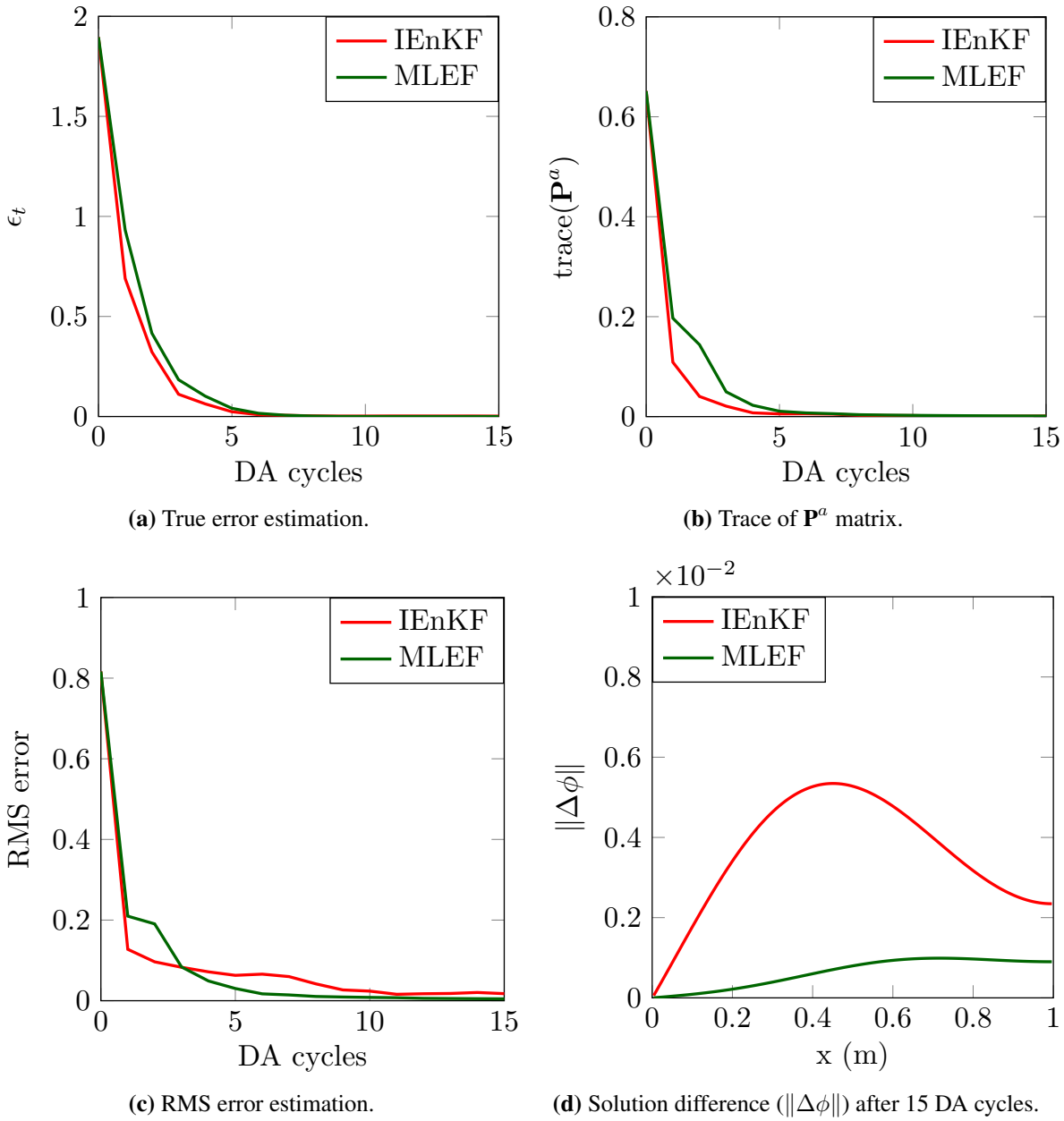
(c) RMS error estimation.

(d) Solution difference ($\|\Delta\phi\|$) after 15 DA cycles.

Figure 7.3: Comparisons of different measures among the EnKF (blue line), IEnKF (red line), and MLEF (green line) methods.

mance depends on tracking the convergence histories of the true error, the trace of $\mathbf{P}^a$ matrix, and the RMS error along the DA process. The trajectories for the true error estimation and the trace of $\mathbf{P}^a$ matrix are shown in Fig. 7.4a and 7.4b. Both methods have achieved an error reduction by more than 95% within 6 DA cycles. As the DA predictions converge toward the truth, the overall impact of data assimilation on the CFD simulations decreases as expected. Similarly, the RMS error can

be considered for the quality assessment of data assimilation. As shown in Fig. 7.4c, the MLEF method drives the error down much lower in the case of the quadratic observation operator based on the RMS error. The solution differences between the truth and the DA+CFD predictions at the last DA cycle are shown in Fig. 7.4d. The final prediction by MLEF shows a better convergence towards the truth than IEnKF.



(a) True error estimation.

(b) Trace of $\mathbf{P}^a$ matrix.

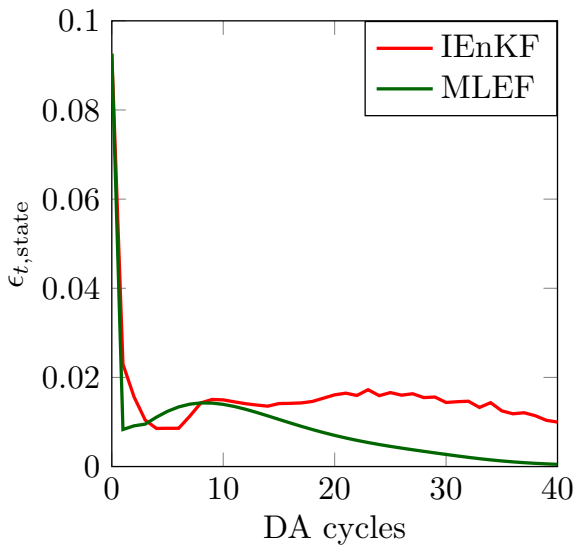
(c) RMS error estimation.

(d) Solution difference ($\|\Delta\phi\|$) after 15 DA cycles.

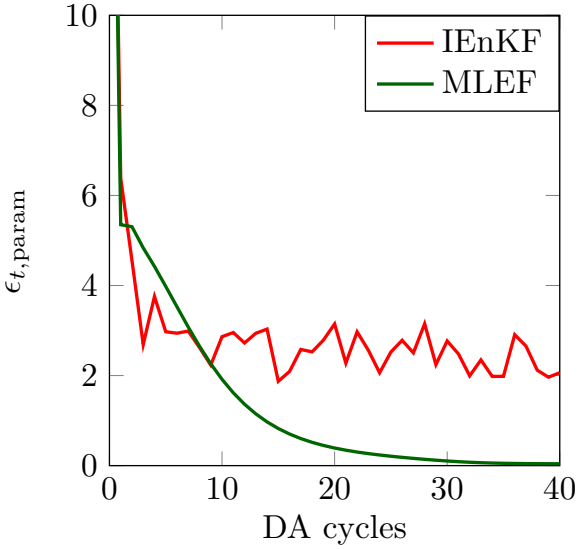
Figure 7.4: Comparisons of different measures between the IEnKF (red line) and MLEF (green line) methods.

7.1.3 Errors in Both Initial Condition and Multiple Parameters

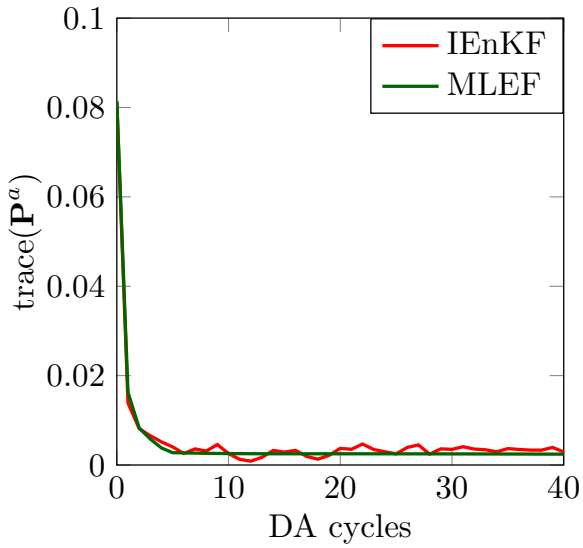
We investigate the performance of the MLEF and IEnKF methods with the same nonlinear observation operator for the situation in which the uncertainties exist in both the initial condition and the three model parameters (γ, μ, and c) simultaneously. For each model parameter, members are selected randomly by following a normal distribution. The normal distribution curves used here have the mean value of 3.0 and standard deviation of 1.0 for generating the initial guess of μ and c, and using the mean value of 4.0 and standard deviation of 2.0 for the initial guess of γ. The DA setups remain the same as those in the above cases. Since the magnitude of the errors in the model parameters and the initial condition are of a different order of magnitude, tracking the overall true error would not be particularly instructive. Instead, we plot the trajectories of the convergence history of the true error estimation for the model parameters and the solution variable separately in Figs. 7.5a and 7.5b. We clearly can see that some fluctuations occur in the results by the IEnKF method, while the results by the MLEF method show nice and smooth convergence history along the DA process. The error takes a much longer time to be adjusted or converged to a lower magnitude for this case. This is not surprising because the model parameters are not observed components in the given observation and can only get corrected through the covariance between the coupled uncertainties with the solution variable. Similar oscillation phenomena are observed in the results of the IEnKF method when tracking the trajectories of the trace value of only the solution variable part of $\mathbf{P}^a$ matrix along the entire DA process in Fig. 7.5c. Figure 7.5d compares the final DA predictions by the IEnKF and MLEF methods to the truth along with the free CFD runs. In addition, the trajectories of each model parameter by the two DA methods are plotted in Fig. 7.6. Clearly, MLEF can efficiently and effectively improve the estimate of the model parameters and the prediction.



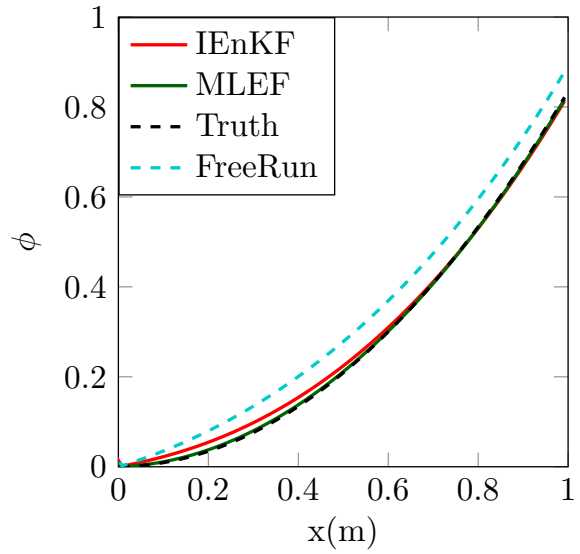
(a) Errors in the solution variable.



(b) Errors in the model parameters.



(c) Trace of $\mathbf{P}^a$ matrix.



(d) Predictions after 40 DA cycles.

Figure 7.5: Comparisons of various measures between the IEnKF (red line) and MLEF (green line) methods.

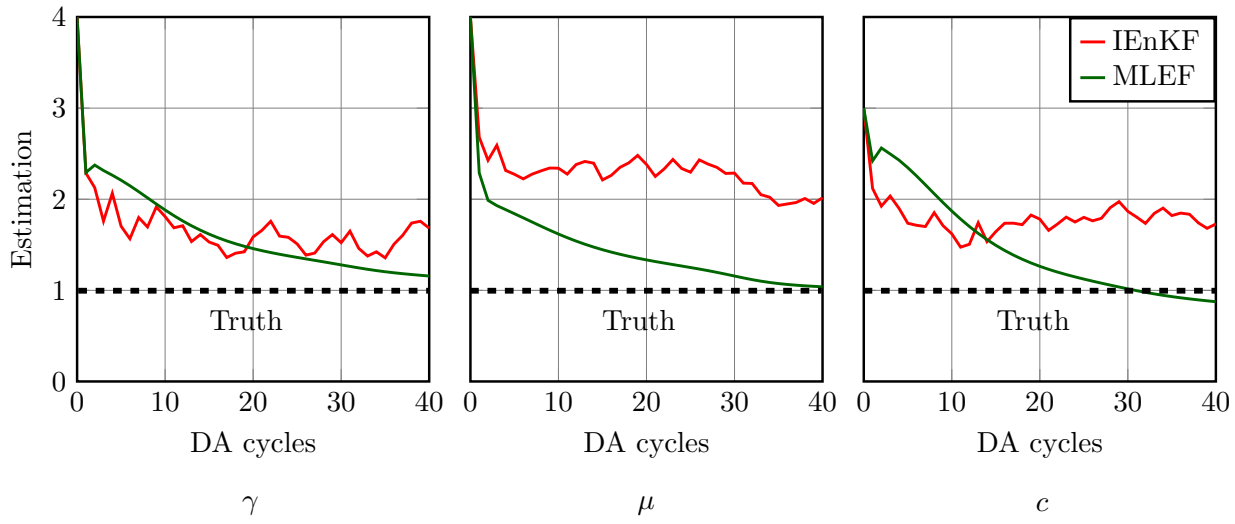


Figure 7.6: The estimation of each model parameter in terms of ensemble mean ($\bar{\gamma}$, $\bar{\mu}$, and $\bar{c}$) by IEnKF (red line) and the control state ($\hat{\gamma}$, $\hat{\mu}$, and $\hat{c}$) by MLEF (green line).

7.2 Lorenz-96 Model

7.2.1 Problem Configuration

We continue the investigation of the DA impact on the imperfect initial condition of L96 model. It is commonly used as a model problem in data assimilation. The model equation is given by,

$$\frac{dx_j}{dt} = (x_{j+1} - x_{j-2})\, x_{j-1} - x_j + F\,, \quad j = 1, 2, \ldots, N_x\,, \tag{7.7}$$

where N_x represents the total number of variables. In this study, we use $N_x = 20$. F is the source term for enforcing the system to be the chaotic regime, which has the value of 8.17. A strong positive or negative correlation is observed for each variable with respect to the adjacent variable components in the vector. The periodic boundary condition is applied in the system. The numerical solution is obtained by discretizing Eq. (7.7) in the temporal space using a fourth-order Runge-Kutta time marching scheme with a constant time step $\Delta t = 0.001$. The truth of this problem is adopted as the solution that is propagated from the initial condition of setting the first

model variable to be 0.1 and the rest of them to be zero. This test focuses on filter performance with a nonlinear quadratic observation operator.

7.2.2 Imperfect Initial Condition in Chaotic-mode Problem

A sensitivity study of the ensemble size is performed first using the IEnKF method to consider both the solution accuracy and computational efficiency. A linear observation operator is applied in the form of

$$\mathcal{H}(X_j) = X_j\,, \quad j = 1, 2, \ldots, N_{\text{obs}}\,. \tag{7.8}$$

The observations are synthesized from the perfect model for every component by adding noise with the error distribution $\mathrm{N}(0, 0.5)$. The initial members are generated by randomly perturbing the initial condition. The DA frequency is set to be every 100 time steps, and the time step is $\Delta t = 0.001$ s. Six ensembles, consisting of 4, 6, 8, 10, 12, and 14 members, are generated. The total computational costs and the cost increments on average are reported in Table 7.1 with respect to each ensemble size. The trajectories of the convergence history of the RMS error are plotted in Fig. 7.7. We find that an ensemble consisting of 12 members shows the best performance in the aspects of both the solution accuracy and the computational efficiency along the DA process. In the following test cases, we use 12 members in the ensemble.

Table 7.1: The computational cost and the cost increment of each ensemble.

Ensemble size	Total CPU cost (sec)	Increment (%)
4	0.563	–
6	0.702	24.64
8	0.885	26.03
10	0.997	12.60
12	1.107	11.04
14	1.218	10.08

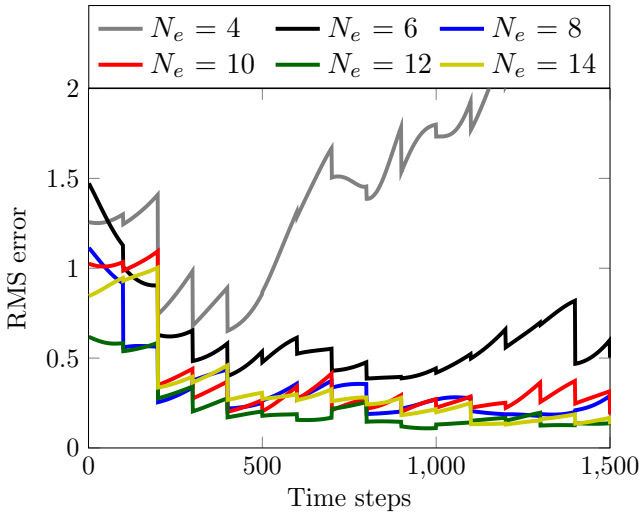


Figure 7.7: Comparison of the RMS errors of the six ensembles.

The DA performance is firstly assessed by comparing the solution of the model states between the predictions by the three DA methods and the truth. We use the same linear observation operator as defined in Eq. (7.8). The observations are still synthesized from the perfect model, but now they are at every other cell by adding noise with the error distribution $\mathrm{N}(0, 0.75)$. The DA frequency and the time step size remain the same as those in the above case. The initial members are generated

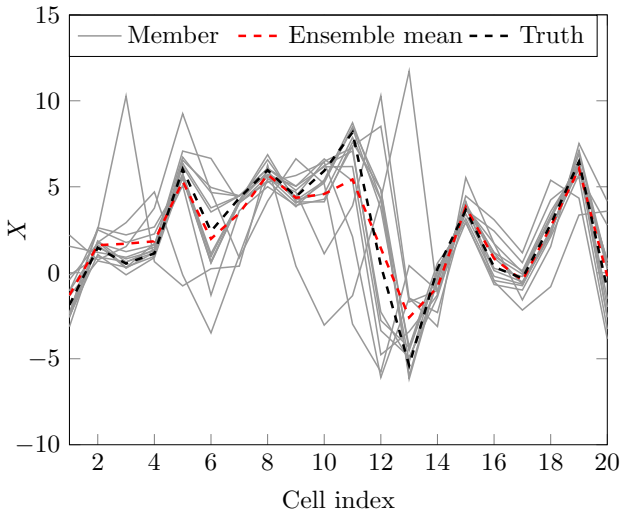
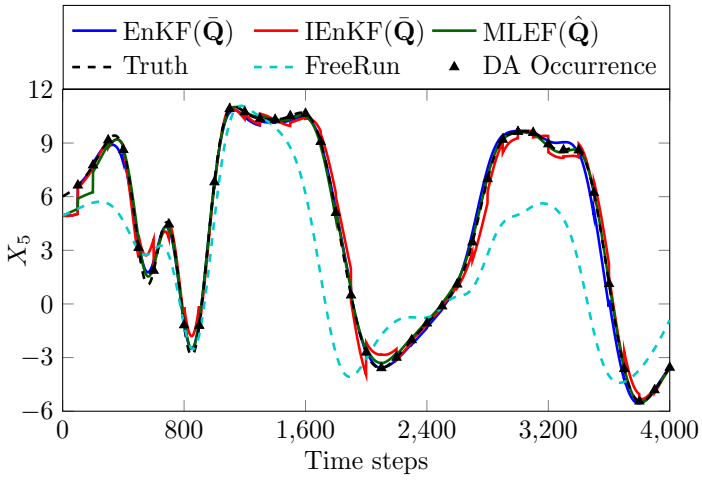
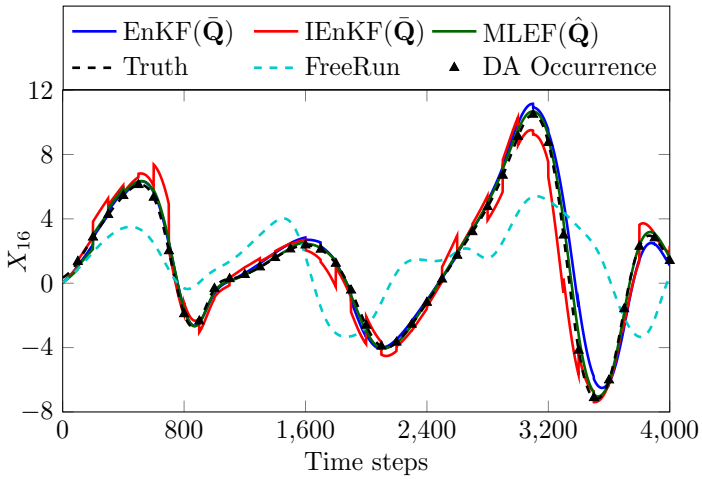


Figure 7.8: The members in the initial ensemble.

by the lagged forecast method with $t_\tau = 250$. Figure 7.8 shows the spread of the members, the ensemble mean, and the truth. The results in Fig. 7.9 present the trajectory comparisons of both the observed component (X_5) and unobserved component (X_{16}) predictions along the entire DA process. The small black triangles indicate when the observations are available and assimilated into the DA+CFD system. By comparing the overall difference between the truth, the free CFD runs, and



(a) The observed component.

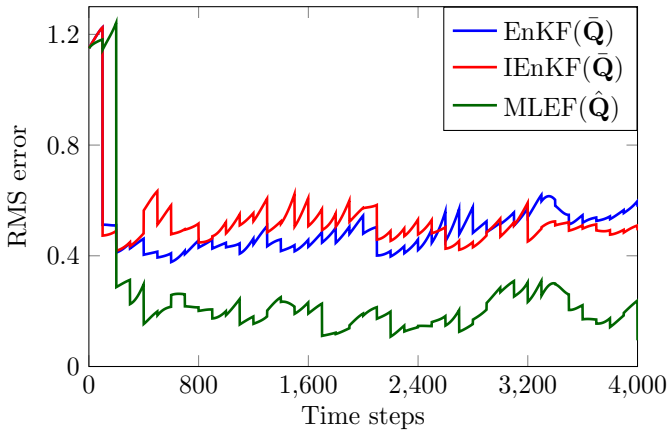


(b) The unobserved component.

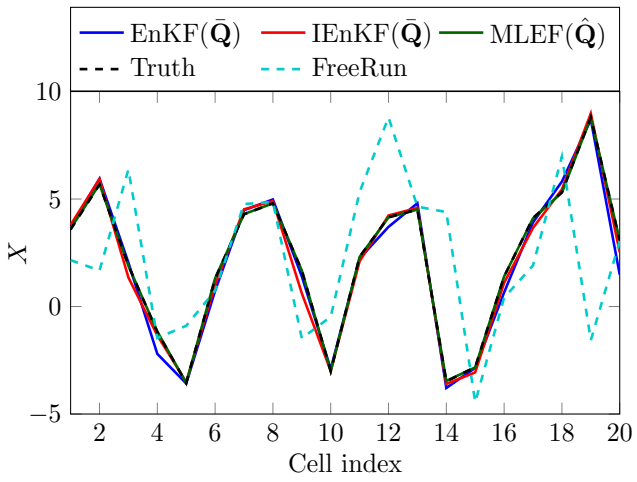
Figure 7.9: Trajectory comparisons on the observed and unobserved components among the EnKF (blue line), IEnKF (red line), and MLEF (green line) methods.

the predictions from the EnKF, IEnKF, and MLEF methods, all DA predictions are pulled toward the observation at the beginning of the DA process. However, only the MLEF prediction gets better converged toward the "truth" along the entire process. Partially, this may be due to the optimization at the analysis stage in the MLEF method.

We further assess the performance based on the measure of the RMS error. By tracking the trajectories of the convergence history of the RMS error over the DA cycles and comparing among the three different DA methods in Fig. 7.10a, we find that data assimilation is effective in the error reduction for the chaotic mode problem since all the RMS error values are reduced as more DA



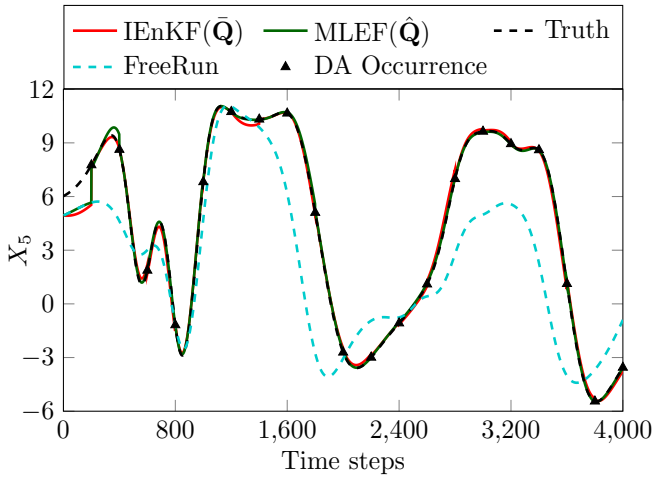
(a) RMS error estimation.



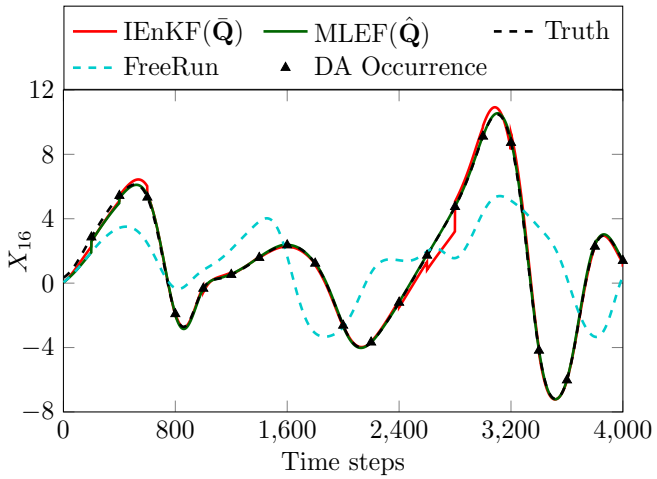
(b) Predictions after 40 DA cycles.

Figure 7.10: Comparisons of different measures among the EnKF (blue line), IEnKF (red line), and MLEF (green line) methods.

cycles are performed. Some fluctuations occur in the results by 3 DA methods. However, the result of the MLEF method converges to lower magnitude along the DA process. The final prediction by MLEF converges closer to the truth after 40 DA cycles in Fig. 7.10b, which shows that the MLEF has better convergence performance than the other two methods on addressing the chaotic dynamics.



(a) The observed component.



(b) The unobserved component.

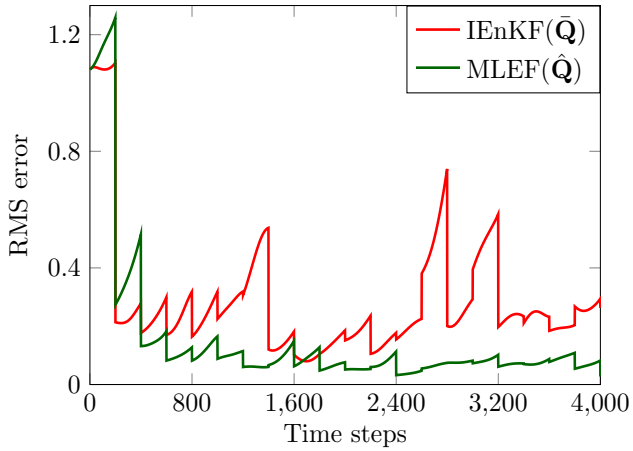
Figure 7.11: Trajectory comparisons on the observed and the unobserved components between the IEnKF (red line) and MLEF (green line) methods.

More interestingly, we investigate the performance of the MLEF and IEnKF methods for the nonlinear quadratic observation operator, where $\mathcal{H}(\cdot)$ is given by

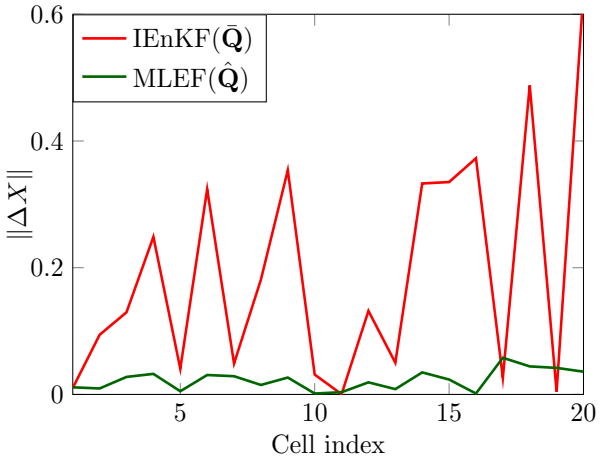
$$\mathcal{H}(X_j) = 2.5X_j^2 + 1.5X_j\,, \quad j = 1, 2, \ldots, N_{\text{obs}}\,. \tag{7.9}$$

Similarly, the observations are synthesized from the perfect model at every other component by adding noise with the error distribution $\mathrm{N}(0, 0.25)$. In this case, because of the chaotic model problem and the nonlinearity introduced from the given observations, a very small perturbation in the DA initial conditions can lead to a significant deviation when testing both the IEnKF and MLEF methods. Instead, the observation error is reduced in order for us to focus on understanding the uncertainty reduction between the IEnKF and MLEF methods. The DA frequency is set to be every 200 time steps, and the time step is still $\Delta t = 0.001$ s. The predictions of the observed component (X_5) and unobserved component (X_{16}) predictions are compared among the results of the IEnKF and MLEF, the truth, and the free CFD run in Fig. 7.11 along the entire DA process. The small black triangles still indicate when the observations are available and assimilated into the DA+CFD system. Clearly, both DA methods can pull the forecasts to the truth, but the prediction by the MLEF method converges closer to the truth. Nevertheless, as more DA cycles are performed, the solution is getting closer and closer toward the truth, not surprisingly, and apparently than the free CFD run.

The assimilation performance is further investigated by tracking the RMS error over the DA cycles and comparing the two DA methods. The trajectories of the convergence history are plotted in Fig. 7.12a. They are clearly consistent with what we observed in Figs. 7.4c and 7.10a, which show that data assimilation is effective in the error reduction. The absolute solution difference is calculated and presented in Fig. 7.12b. The final prediction of MLEF shows a smaller difference than that in the prediction of the IEnKF, which can indicate that the MLEF method has better filter performance on addressing the nonlinearity of the observation operator than the IEnKF method does.



(a) RMS error estimation.



(b) Solution difference ($\|\Delta X\|$) after 20 DA cycles.

Figure 7.12: Comparisons of different measures between the IEnKF (red line) and MLEF (green line) methods.

7.3 Conclusion

MLEF consistently produces more accurate and efficient solutions than the other two methods and provides more information on both states and their uncertainties during the verification process. This study provides enough evidence to support us in choosing the MLEF method as an appropriate DA algorithm when establishing the DA+CFD system for further investigations in current research.

Chapter 8

Verification and Validation

Before applying data assimilation to improve the predictions of practical turbulent reacting flows, the DA+CFD system established in the current work needs to be verified and validated. The CDR problem, as described in the previous chapter, verified the MLEF method. Here, a turbulent Couette flow is used to validate the framework. This flow problem is chosen because it is a canonical fluid dynamics problem and has literature data for the performance assessment of the framework. Moreover, the prediction of the turbulent viscosity model is of particular interest. The CFD predictions are compared between those with the use of data assimilation and those without the use of data assimilation. The assimilation performance is assessed based on the measures, such as the χ-squared diagnostics and the RMS errors.

8.1 RANS Modeling of Turbulent Couette Flow

The CFD modeling of the plane turbulent Couette flow is based on Reynolds-averaged Navier-Stokes equations. The Spalart-Allmaras model [81, 82] is employed. Since the turbulent eddy viscosity, $\rho\tilde{\nu}$, has to be solved in the coupled system for the turbulent Couette flow, the system of governing equations is slightly different from Eqs. (2.1) - (2.3) on a rectangular physical domain, given by

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}\cdot(\rho\vec{u}) = 0 \tag{8.1}$$

$$\frac{\partial}{\partial t}(\rho\vec{u}) + \vec{\nabla}\cdot\left(\rho\vec{u}\vec{u} + p\vec{\vec{I}}\right) = \vec{\nabla}\cdot\left(\vec{\vec{\tau}} + \vec{\vec{\tau}}_t\right) \tag{8.2}$$

$$\frac{\partial}{\partial t}(\rho e) + \vec{\nabla}\cdot\left[\rho\vec{u}\left(e + \frac{p}{\rho}\right)\right] = \vec{\nabla}\cdot\left[\left(\vec{\vec{\tau}} + \vec{\vec{\tau}}_t\right)\cdot\vec{u}\right] - \vec{\nabla}\cdot(\vec{q} + \vec{q}_t) \tag{8.3}$$

$$\begin{aligned}\frac{\partial \rho\tilde{\nu}}{\partial t} + \vec{\nabla}\cdot(\rho\vec{u}\tilde{\nu}) &= C_{b_1}(1 - f_{t2})\rho\tilde{S}\tilde{\nu} + \vec{\nabla}\cdot\left[\frac{\rho}{\sigma}(\nu + \tilde{\nu})\vec{\nabla}\tilde{\nu}\right] - \frac{1}{\sigma}(\nu + \tilde{\nu})\vec{\nabla}\rho\cdot\vec{\nabla}\tilde{\nu} \\ &+ \frac{C_{b_2}\rho}{\sigma}\vec{\nabla}\tilde{\nu}\cdot\vec{\nabla}\tilde{\nu} - \rho\left(C_{w_1}f_w - \frac{C_{b1}}{k^2}f_{t2}\right)\frac{\tilde{\nu}^2}{d^2},\end{aligned} \tag{8.4}$$

where the model closure coefficients are detailed described in the works by Allmaras et al. [81] and Spalart and Allmaras [82]. All the solution variables are time-averaged quantities in equations. In the equations, $\vec{\vec{\tau}}_t$ is the turbulent Reynolds stress tensors, and $\vec{q}_t$ is the turbulent heat flux vector. The other mathematics symbols are the same as those described in Chapter 2.

8.2 CFD and DA Configurations

The flow is channeled by two infinite long plates, as shown in Fig. 8.1, which are parallel in x coordinate and separated by a height of H = 0.0292 m in y. The computational domain is $0.0292\mathrm{m}$ $\times$ $0.0292\mathrm{m}$ with a grid of 32×32 cells. The top wall moves in x-direction with a constant speed of U_w = 104.18 $\mathrm{m/s}$. The Reynolds number is around Re_y = 10000 based on the moving wall velocity and the channel height. The fluid is air at $80\ ^{\circ}\mathrm{C}$, with a kinematic viscosity (ν) of $2.09\times10^{-5}\ \mathrm{m^2/s}$, a density of $1.177\ \mathrm{kg/m^3}$ and a thermal conductivity of $0.0243\ \mathrm{W/mK}$. Periodic BCs are set for the x-direction. For the y-direction, the top wall moves with U_w while the bottom wall is set to no-slip. The x-velocity of the flow field is initialized using the laminar Couette flow velocity profile (see Gao et al. [36]). The eddy turnover time (τ_e) is about $5.61\ \times\ 10^{-4}$ s based on the channel height and initial velocity.

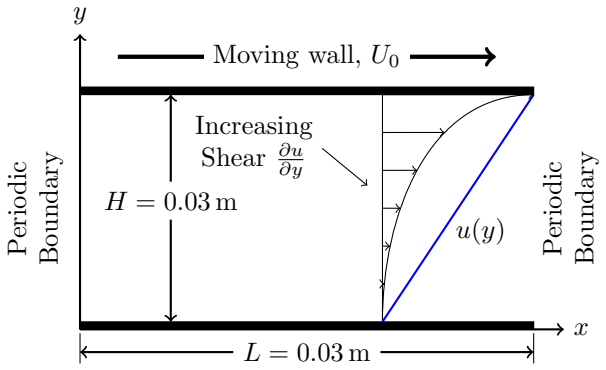


Figure 8.1: Diagram of the turbulent Couette flow case showing boundary conditions and initial velocity profile at the starting point.

For DA configuration, the lagged forecast method [16], as mentioned in the previous section, is implemented for creating the initial ensemble by running the forward model from -2$\mathrm{T_{DA}}$ to 2$\mathrm{T_{DA}}$,

where the data assimilation frequency ($\mathrm{T_{DA}}$) is set to be $2\tau_e$. The central time of the initial lagged forecast region is $10.5\tau_e$. Members ($N_e = 10$) are randomly selected along this lagged forecast region. The deterministic control state ($\hat{\mathbf{Q}}_0$) is the solution at the central time of the region. In the present study, uncertainties are assumed to exist in both the initial condition and the two model parameters (c_{b2} and c_{w3}) simultaneously. The reference values for c_{b2} and c_{w3} are 0.622 and 2.0, respectively. For each model parameter, members are selected randomly by following a normal distribution. The normal distribution curve used for generating the initial guess of c_{b2} has the mean value of 3.0 and standard deviation of 1.0, and using the mean value of 6.0 and standard deviation of 1.5 for the initial guess of c_{w3}. The control vector $\hat{\mathbf{Q}}$ is constructed by $c_{b2}, c_{w3}, \rho, \rho u, \rho v, \rho\epsilon$ and $\rho\tilde{\nu}$ in form of

$$\hat{\mathbf{Q}} = \left(c_{b2}, c_{w3}, \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_{N_\mathrm{c}} \end{pmatrix}^{\mathrm{T}}, \begin{pmatrix} (\rho u)_1 \\ \vdots \\ (\rho u)_{N_\mathrm{c}} \end{pmatrix}^{\mathrm{T}}, \begin{pmatrix} (\rho v)_1 \\ \vdots \\ (\rho v)_{N_\mathrm{c}} \end{pmatrix}^{\mathrm{T}}, \begin{pmatrix} (\rho\epsilon)_1 \\ \vdots \\ (\rho\epsilon)_{N_\mathrm{c}} \end{pmatrix}^{\mathrm{T}}, \begin{pmatrix} (\rho\tilde{\nu})_1 \\ \vdots \\ (\rho\tilde{\nu})_{N_\mathrm{c}} \end{pmatrix}^{\mathrm{T}} \right)^{\mathrm{T}}, \quad (8.5)$$

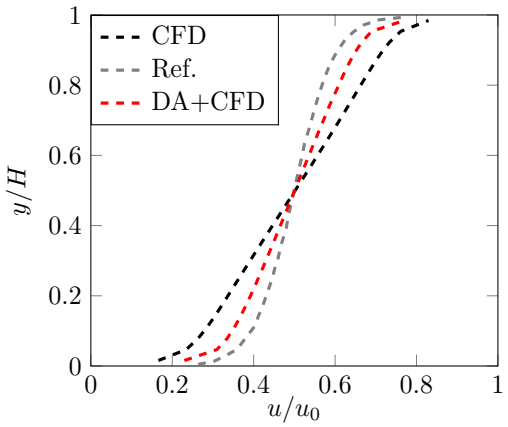
where N_c denotes the total number of the cell of the computational domain. Note that the internal energy, not the total energy variable, is selected as a control variable because of its "independent" relationship with other model variables in order to keep its error as Gaussian. In addition, from that point of view, using independent control variables makes the data assimilation more efficient and possibly more robust. A literature data of steady-state u profile is available in work by Spalart et al. [83]. The observations in the current work are synthesized from this profile at every other cell by adding noise with the error distribution N(0, 4). It is important to point out that in the case of decaying modes, such as those in a statistically steady-state system like the Couette flow, perturbed and non-perturbed simulations would eventually come to the same point, which implies minimal uncertainty. But, for the verification purpose, by assimilating the steady-state solution as the observation into the system, we focus on proving that the data assimilation can reduce the uncertainty faster than that by the dissipation mechanism in the forward model, which leads to the eventual steady state.

8.3 Performance Assessment

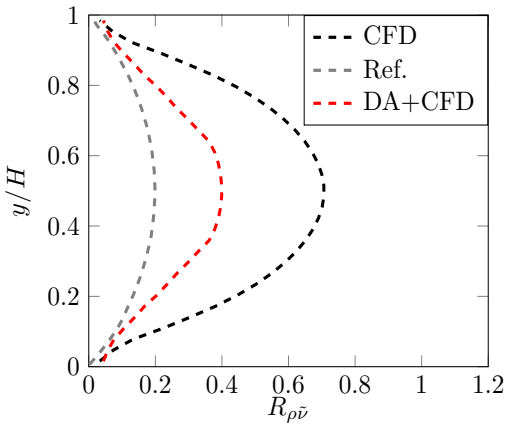
The profiles of axial velocity (u) and the eddy viscosity quantity ($R_{\rho\tilde{\nu}}$) are shown in Fig. 8.2 after the 1$^{\text{st}}$, 5$^{\text{th}}$ and 10$^{\text{th}}$ DA cycles. The variable $R_{\rho\tilde{\nu}}$ is calculated by

$$R_{\rho\tilde{\nu}} = \frac{\overline{u'v'}\rho\tilde{\nu}}{u_\tau H \frac{du}{dy}}\,, \tag{8.6}$$

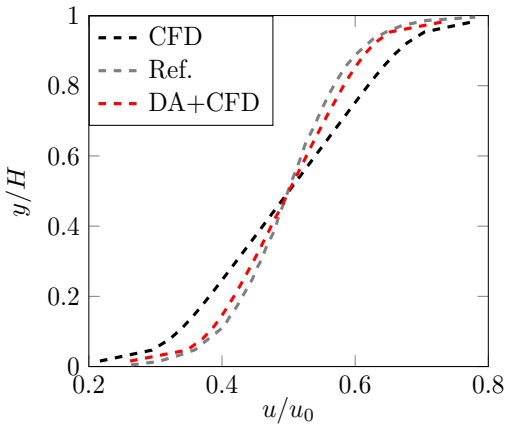
where $\overline{u'v'}$ is a component of the Reynolds stresses and u_τ is the friction velocity. These y-profiles are investigated and plotted at $x = L/2$. The reference solutions of u and $R_{\rho\tilde{\nu}}$ come from the work by Spalart et al. [83]. The free CFD predictions (i.e., the solutions without using data assimilation) are compared to the predictions with data assimilation along with the literature. As more data assimilation cycles are performed, the solution for each variable is apparently getting closer and closer toward the reference than the free CFD simulations. It is worth pointing out that $\rho\tilde{\nu}$ is treated as a non-observed variable in the system. However, a positive impact on its prediction by data assimilation is obtained along the DA process, which clearly indicates that the DA impact on the non-observed variable can be realized through the coupled dynamics and the solution-dominant covariance between the coupled uncertainties generated in the DA+CFD system. Table 8.1 shows that the optimization with data assimilation reduces the solution error faster than the free CFD run, implying that data assimilation effectively reduces uncertainty. As a reminder, χ^2 test is used to evaluate the correctness of the innovation covariance matrix, and the values are expected to be closer to 1.0 as possible.



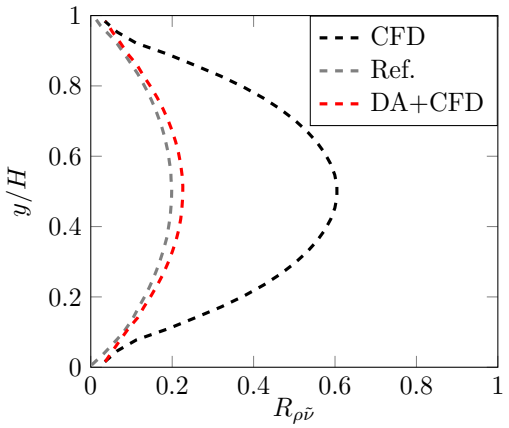
(a) u after 1^{st} DA cycle.



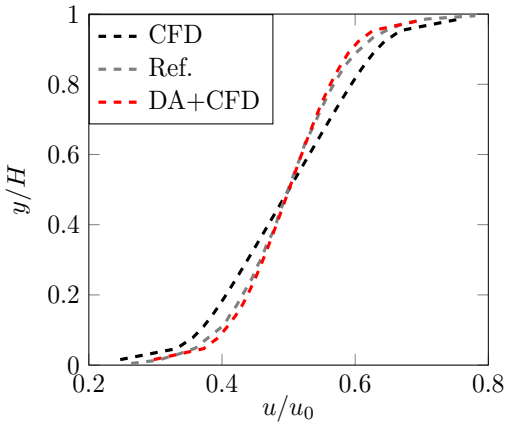
(b) $R_{\rho\tilde{\nu}}$ after 1^{st} DA cycle.



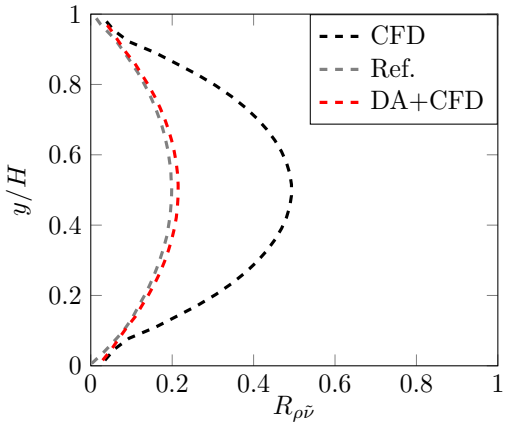
(c) u after 5^{th} DA cycle.



(d) $R_{\rho\tilde{\nu}}$ after 5^{th} DA cycle.



(e) u after 10^{th} DA cycle.



(f) $R_{\rho\tilde{\nu}}$ after 10^{th} DA cycle.

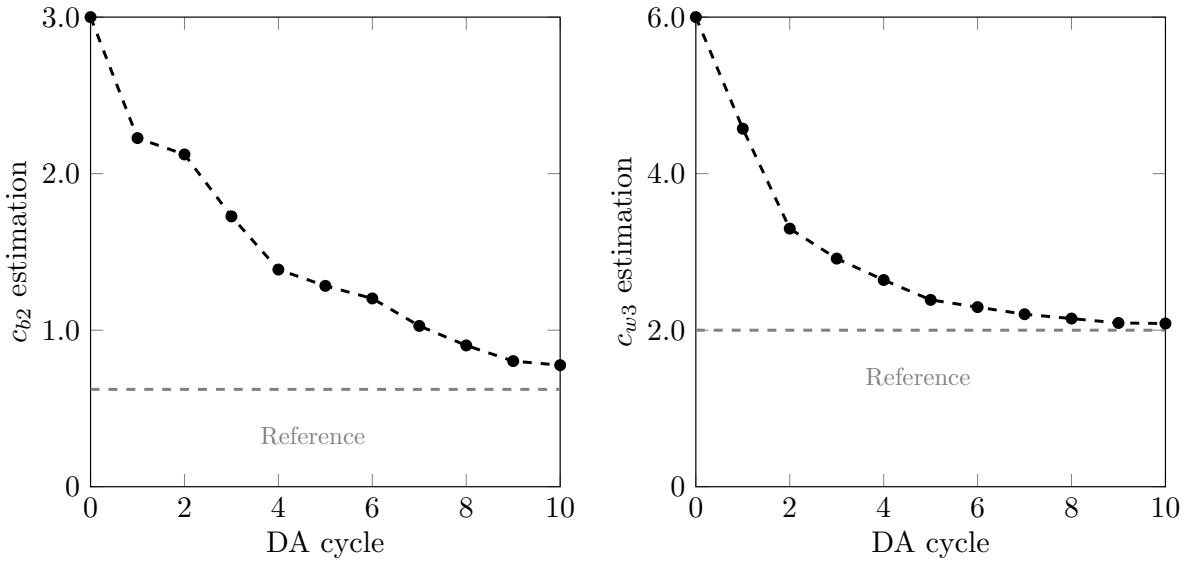
Figure 8.2: Comparisons for u and $R_{\rho\tilde{\nu}}$ profiles at $x = \frac{H}{2}$ after the 1^{st}, 5^{th} and 10^{th} DA cycles.

Table 8.1: χ^2 for the CFD and DA+CFD predictions.

DA cycle	χ^2 (CFD)	χ^2 (DA+CFD)
1	6.72951	1.27838
2	6.39608	1.19123
3	5.03123	1.20321
4	4.11283	1.03216
5	2.62701	1.01494
6	2.33488	1.11669
7	2.42313	0.99635
8	1.96743	0.96314
9	1.73231	1.08076
10	1.59313	1.10593

Furthermore, the trajectories of each model parameter are plotted in Fig. 8.3. MLEF effectively improves the estimate of the model parameters and thus further helps improve the prediction. To keep in mind, since the synthetic observation is not perfect, we do not expect that the predictions converge perfectly to the reference data.

The relative RMS errors of axial velocity from the free CFD and DA+CFD predictions are compared along the data assimilation process at the analysis stage, as illustrated in Fig. 8.4. Along each trajectory, the RMS error is normalized by its forecast RMS error for the first DA cycle. The DA+CFD result converges to a lower magnitude along the DA process. It confirms the effectiveness of the MLEF method in error reduction by using data to optimize the prediction. Especially, the data assimilation helps to reduce uncertainty more significantly than the dissipation mechanism in the model.



(a) c_{b2} **(b)** c_{w3}.

Figure 8.3: The estimation of each model parameter along 10 DA cycles.

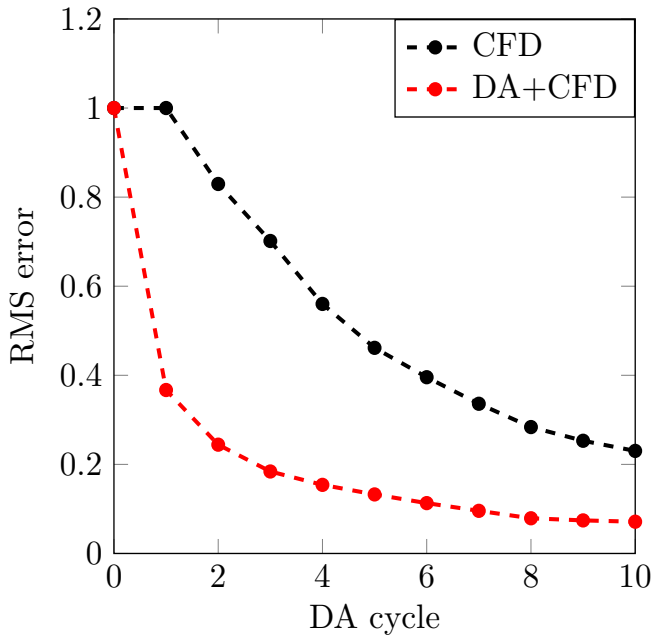


Figure 8.4: The relative RMS error of axial velocity comparisons between CFD (black line) and DA+CFD (blue line) along 10 DA cycles.

Chapter 9

Results and Discussion of the DA+CFD system

The verified and validated DA+CFD framework is now applied to improve the CFD predictability of the time-evolving shear-layer mixing with CH_4-air combustion, correct the boundary conditions of turbulent flow in a bluff-body combustor, and understand the uncertainties of the turbulent flows with and without C_3H_8-air combustion in the same combustor. These problems are selected since they mimic the physics process, such as the development of flow and flame of the shear layers, occurring in practical combustors, where some stabilization mechanisms are often embraced to anchor flames. For example, a bluff body is usually adopted as a flame holder in practical combustion devices. It is important to understand the mixing process and vortex dynamics in practical combustors for the design of the combustor and the overall combustion performance. In particular, data assimilation is applied to improve the prediction of such physics. As mentioned previously, the concept of using data to improve the predictability of CFD modeling of turbulence and combustion has been brought into the field of science and engineering only recently. However, its application to turbulent combustion is still at the fledgling stage due to the complexity of chemical reactions and the intricate interactions between chemistry and flow dynamics. This chapter will focus on investigating how data assimilation works for complex problems and what can be learned by assimilating data to correct the boundary condition and improve the CFD prediction of the flow and flame dynamics.

9.1 Numerical Tests and DA Configurations

9.1.1 Reacting Shear Layer Flow

As illustrated in Fig. 9.1, two parallel, gaseous fluid streams, the upper and the lower halves, flow oppositely along the axial direction in an infinitely long duct bounded by two walls in the y-direction. Specifically, the upper stream is CH_4 while the lower stream is O_2. The mean axial

velocity in the upper stream is $U_1 = 102.0882$ m/sec while the lower stream has a mean axial velocity $U_2 = -U_1$. The y- and z-velocity are initialized to be zero. A hyperbolic tangent profile is used for initializing the $U(y)$ and the mass fraction profile of the chemical species (O_2 and CH_4). In addition, the Busemann–Crocco relationship [84] is used for the initial temperature with $T_1 = 300$ K and $T_2 = 1800$ K. The slip-wall BC is specified for the walls, while periodic BCs are applied for other physical boundaries. The initial Reynolds number based on the vorticity thickness is around 640. The domain is $2\text{mm} \times 4\text{mm} \times 0.25\text{mm}$ with a computational grid of $64 \times 128 \times 8$ cells. Due to the velocity difference, the shearing force causes the two layers to mix in the center of the duct. When CH_4 and O_2 mix at a high temperature, the reaction occurs. The reaction is modeled by a 12-species, 38-steps CH_4-O_2 chemistry mechanism developed by Xu et al. [2].

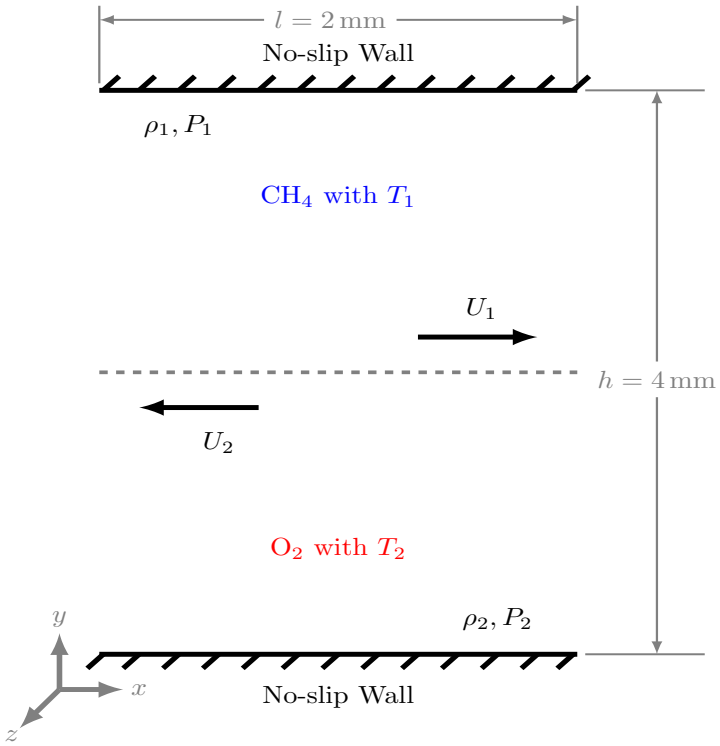


Figure 9.1: Diagram of the reacting shear layer flow.

All conservative variables are assumed to have uncertainties. With using Eq. (7.4), the observations are synthesized from the perfect model for all the species mass fractions, the x- and y-momentum, and the density at every other cell. For each species mass fraction, the observa-

tion error is set to be 5-8% of its true global maximum value. For the density and the x- and y-momentum, we set the observation errors to be 8-10% of individual global maximum true value, as $\sigma_\rho = 0.2\,, \sigma_{\rho u} = 5$, and $\sigma_{\rho v} = 2$. The average acceptance rate of the observations is around 63% in the first 3 DA cycles. Then, it increases to 80% after 5 DA cycles are performed. The ensemble consists of 12 members, created from the lagged-forecast method by setting the $\mathrm{T_{DA}}$ = 9000 time steps.

9.1.2 Non-reacting Flow in a Bluff-body Combustor

The bluff-body combustor developed by the US Air Force Research Lab (AFRL) [85], as illustrated by Fig. 9.2, has a 38.1 mm equilateral triangle flame holder centered in the y-direction and resembles the practical combusting devices in terms of forming three core regions: the anchored flame, recirculation zone, and shear-layer mixing. The computational domain is configured as the inflow and outflow for the x-boundaries, no-slip, adiabatic walls in y-direction, and periodicity in z-direction. The center xy-plane is $L_x \times L_y = 0.829\mathrm{m} \times 0.127\mathrm{m}$. For the inflow air condition at the left x-boundary, the velocity and temperature are specified as u =14.9 m/s and 310 K, respectively. The Reynolds number is 50,164 based on the trailing edge length of the bluff body and the free stream inlet velocity. The total pressure at the right x-boundary is set to be 100 KPa. No-slip and adiabatic conditions are applied to the solid surfaces of the bluff body.

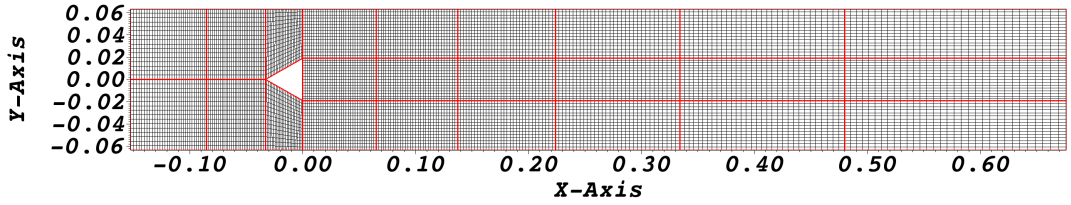


Figure 9.2: The configuration of the bluff-body combustor.

For data assimilation, the ensemble consists of 19 members and 1 control case. Although the details are not shown here, it is important to mention that the sensitivity study of the ensemble size was performed first for the consideration of both the solution accuracy and computational

efficiency. The control vector is $\hat{\mathbf{Q}} = (\rho,\ \rho u,\ \rho v,\ \rho\varepsilon)^{\mathrm{T}}$, with the size of N_s = 48,128. We use the same linear observation operator as described in Eq. (7.4). The observed variables are the density and the x- and y-momentum for two cases, as shown in Fig. 9.3. In case 1, there are 27,800 data-points (observations) being synthesized for the xy-plane at $x \times y$ = [0, 0.677] m $\times$ [-0.635, 0.635] m, whereas case 2 has 144 data-points synthesized for $y \in$ [-0.635, 0.635] m at x = 0.0005 m. Both cases represent the locations where typically experimental data are measured. The observation errors are set to be 5-8% of their global true maximum values, giving $\sigma_\rho = 0.1\,, \sigma_{\rho u} = 5$, and $\sigma_{\rho v} = 3$. The averaged acceptance rate of the observations is close to 88% in both cases along the DA process. Assimilation frequency is $\mathrm{T_{DA}}$ = 0.021 second.

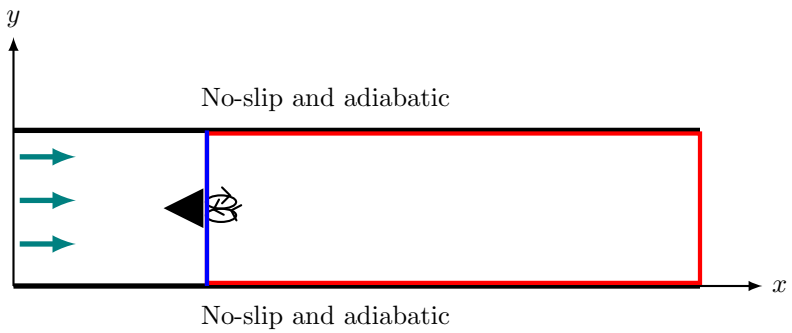


Figure 9.3: Diagram of two observation regions for BBC case.

9.1.3 Reacting Flow in a Bluff-body Combustor

Data assimilation has been performed for the premixed C_3H_8-air reacting flows at ϕ = 0.65 in the bluff-body combustor by focusing on 2D configurations first. Three benefits are achieved by using the 2D DA investigation. First, we can gain confidence in applying the DA+CFD system to complex reacting flows in a burner with realistic geometry and assimilation of experimental measurements. Establishing the proper workflow and assimilation criteria is essential before committing an enormous amount of CPU hours for the 3D DA investigation. The solution process of DA+CFD should be largely independent of spatial dimensions, and the resulting procedure from 2D is expected to be applicable to 3D. Secondly, we can assess the DA performance and parametric

sensitivity for fast turnaround times between investigations, such as the ensemble size, the assimilation frequency, and the experimental data usage, including the instantaneous or the averaged quantities (time-averaging, spatial-averaging, or space-time averaging, signals or derived quantities). Although not all information is presented herein, it is cataloged as guidance for the 3D study. Finally, scalability of specific operators and estimation of computational cost for 3D simulations are made, providing a basis for CPU time allocations for the 3D DA study in future work.

The computational domain as illustrated by Fig. 3.1 is reduced to a $xy-$plane with $z = 0$ for the 2D cases. The C_3H_8-air reactions are modeled by the chemical mechanism developed by Zettervall et al. [1] ("Z66"), including 25 species and 66 reactions. The boundary conditions for the premixed C_3H_8-air combustion are the same as described above. The mesh used for this study is particularly chosen with a similar resolution that is consistent with the literature references [86–89] for the purpose of comparisons. This "4mm-mesh", commonly referred to in these works, was created by taking into account a few important physical length scales to turbulent combustion, such as the boundary layer of the bluff body affecting the dynamics of shearing and recirculating. Cocks et al. [86] estimated this length scale to be approximately $1/\sqrt{\mathrm{Re}_D}$, requiring the spatial resolution of 5 mm; Re_D is 38,658 based on the trailing edge length of the bluff body and the inlet flow condition. In the present work, the base grid consists of 24 mapped blocks, with 16,384 computational cells in total. The cell distribution is either the 8×16 cells or 16×16 cells per block. In most regions of the main combustor channel, the cells are nearly isotropic with a resolution of 5.5 mm, but the cells are refined near the walls by stretching the mesh as shown in Figs. 3.1 and 3.2. The stretching ratios in x and y-direction are limited to be less than 10%, giving some fine cells with a resolution of 1 mm. Cell resolution in the plenum is much lower than the rest of the domain since it simply exists to enforce the boundary conditions properly.

Two DA tests are configured. The first case examines the consequence of an imperfect initial condition on the solution field. The second test investigates the situation in which the uncertainties exist simultaneously in both the initial condition and the inlet boundary condition, specifically, the axial velocity (u_{inlet}) at the leading edge of the slip wall region, as shown by the black arrowed-

lines in Fig. 3.1. In each case, the ensemble consists of 10 members and 1 control case. To define an initial state, including all the solution variables and their uncertainties, we select each member in the ensemble from the solution time region of $2.25\tau_f \leq t \leq 2.7\tau_f$ by using the lagged forecast method. The unit flow-through time, τ_f, is defined as the total amount of time that a particle travels through the duct part of bluff-body combustor along the streamline directory by $\tau_f = L_x/u_{\text{inlet}}$. For the first case, τ_f is calculated to be 0.0521 s. For the second test, τ_f is calculated based on the mean inlet velocity in the control case, which is about 0.0908 s. The DA frequency occurs every third of 1 flow-through time in the first case and every half flow-through time in the second case, respectively. To initialize the u_{inlet} condition in each member but avoid introducing unphysical noise into u_{inlet} profile, we add a small number of perturbations into the profile equation by,

$$u_{\text{inlet}}(y, t=0) = \begin{cases} 100\beta_1 y^2 + \beta_2\,, & \text{if } \beta_3 > 0 \\ \beta_2\,, & \text{otherwise} \end{cases}, \tag{9.1}$$

where we select β_1, β_2, and β_3 randomly by following a normal distribution as

$$\beta_1, \beta_2 \in N(9.5, 2)\,, \quad \beta_3 \in N(0, 1)\,. \tag{9.2}$$

The reference inlet velocity data is available in the work by Caswell et al. [90]. The initial members of the inlet velocity profiles are presented and compared to the reference data in Fig. 9.4.

The experimental measurement setup and area in the work by Caswell et al. [90], as illustrated in Fig. 9.5, are a 2D plane with dimension of $3.47D \times 2.91D$ as the viewing window, where $D = 38.1\text{mm}$. Note that there is a difference in the transversal width between the computational and experimental configurations as shown by the side view in Fig. 9.5b, and this difference needs to be taken into consideration when 3D investigations are considered in future work (e.g., normalizing the data with the corresponding inlet velocity). The data, including instantaneous u and ρc_{OH}, are assimilated for the first 42 DA cycles in the first test and the first 19 DA cycles in the second test. The time-averaged quantities are also assimilated, for the last 8 DA cycles in the first case and

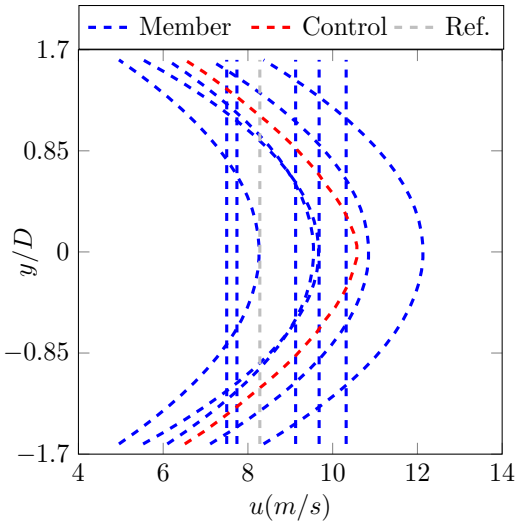
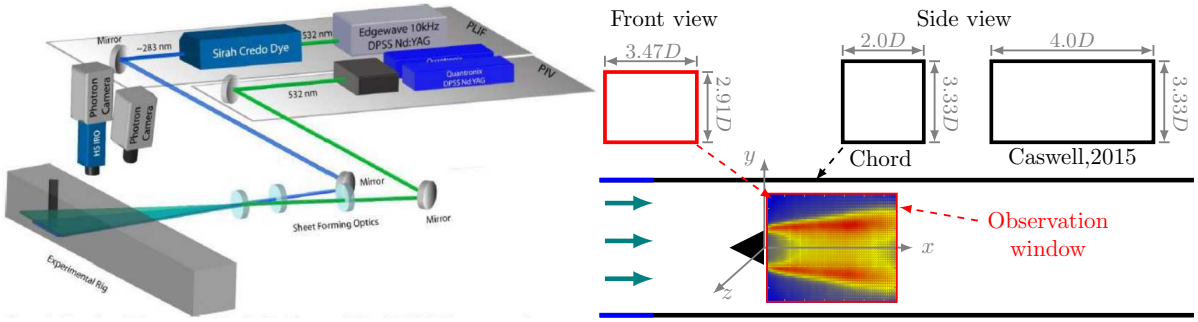


Figure 9.4: Initial inlet u profiles for the members.

the last 11 DA cycles in the second case, respectively, to demonstrate the data impact. The errors in the x-velocity and the ρc_{OH} are 5-10% of their global maximum values, giving $\sigma_u = 4.5$ and $\sigma_{\rho c_{\mathrm{OH}}} = 8.5 \times 10^{-5}$, respectively. The average acceptance rate of the observations is close to 75% along the DA process.



(a) Experimental measurements setup [90].

(b) Observation location (i.e., viewing window).

Figure 9.5: Schematic of the experimental measurements setup and the viewing window.

Before assimilation, the experimental measurements of ρc_{OH} are rescaled from the normalized OH-PLIF data for the equivalence ratio of $\phi = 0.65$. A 1D free flame simulation in Cantera version 2.5.1 [91] using Z66 kinetics and multi-component transport was performed at the specified equivalence ratio. The maximum value of the mass fraction of OH is 0.00063, which is used

as the scaling factor (α_{coeff}) during the pre-processing of experimental data. It is important to mention that the experimental data is for ϕ=0.6, and there are available scaling factors in multiple variables for rescaling normalized OH-PLIF for a range of equivalence ratios. In particular, for ϕ = 0.65, the factors for the mass fraction, mole fraction, and concentration for OH, and the total fluid density are 0.002941, 0.004869, 3.7124 $\times 10^{-5}$, and 0.2146, respectively, based on the 1D free flame simulation in Cantera using Z66 kinetics and multi-component transport performed. The maximum location of OH concentration through the flame was located, and the mass fraction, mole fraction, concentration, and density at that point are reported as the scaling factors in the table. Unfortunately, a caveat is that the point of maximum concentration does not always correspond to the point of maximum mass fraction or maximum mole fraction. However, the differences between these points are small, and to ensure consistency between density and mass and mole fractions, only the values at the point of maximum concentration are reported.

9.1.4 High Performance Computing

In the current work, the reacting shear layer flow and the non-reacting flow in a bluff-body combustor were carried out on Atlantis, an internal high-performance compute server managed by the Computational Fluid Dynamics and Propulsion group at Colorado State University. It consists of nine compute nodes and 189 TB of storage connected by a 40 Gbps InfiniBand network. In total, there are 200 cores on the compute nodes. For the turbulent reacting flow in a bluff-body combustor, the member simulations were carried out on Narwhal and Onyx managed by the Department of Defense (DoD) Navy Supercomputing Resource Center (DSRC) and the U.S. Engineer Research & Development Center (ERDC) DSRC, respectively. Technical details on the system architectures, programming environments, memory complement, and disk subsystems can be found in https://centers.hpc.mil/systems/unclassified.html.

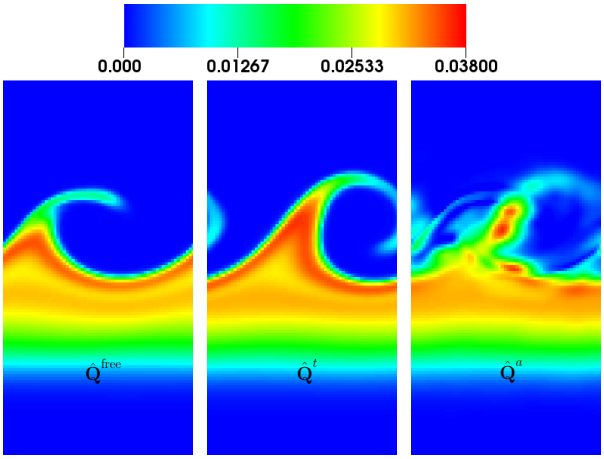
9.2 Results and Discussion

9.2.1 Reacting Shear Layer Flow

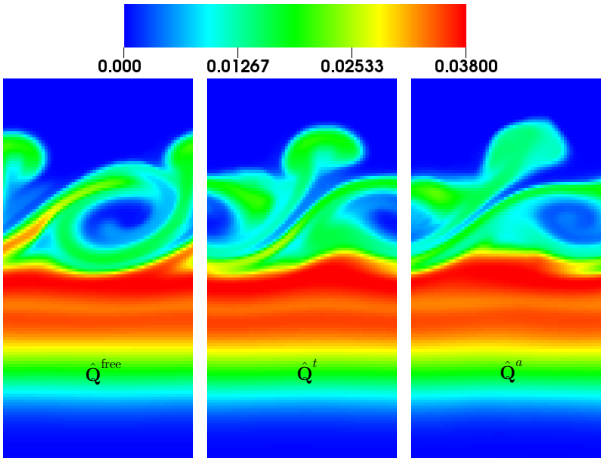
Figure 9.6 compares the DA+CFD predictions of $\rho c_{\mathrm{H_2O}}$ to the synthesized truth and the free CFD simulations for assimilation cycles 1, 5, and 10. This figure consists of 3 rows and 3 columns. Each row (from top to bottom) shows the results for a particular assimilation cycle. Columns from left to right correspond to the free CFD simulation (e.g., without using data, $\hat{\mathbf{Q}}^{\text{free}}$), the truth ($\hat{\mathbf{Q}}^{t}$), and the optimal analysis from data assimilation($\hat{\mathbf{Q}}^{a}$), respectively. The flame structure indicated by the $\rho c_{\mathrm{H_2O}}$ species distribution is well corrected, which is promising. It demonstrates that the optimization algorithm is robust because a combusting fluid dynamical system can be very sensitive to the species variations, and the optimization algorithm may diverge easily or have slow convergence issues. The analysis also reveals that the changes (e.g., the difference between $\hat{\mathbf{Q}}^{a}$ and $\hat{\mathbf{Q}}^{t}$) become less pronounced after cycle 5. This is the phenomenon we expect to see, and when this occurs, it is encouraging because a larger error reduction should be expected to be seen in earlier assimilation cycles than later ones, and this fact demonstrates that the data is effectively driving the system to the true trajectory. We have also observed that all solution variables behave consistently over the cycles. This is remarkable for a combusting system, positively indicating that the dynamical system is in a good balance during data assimilation. All indicates that the Chord-MLEF framework is effective, and the data quality is good.

It is worth pointing out that $\rho\epsilon$ is treated as a non-observed variable, and hence a positive impact on its predication by data assimilation is realized through the coupled dynamics and the solution-dominant covariance between the coupled uncertainties generated in the DA+CFD system. As stated above, all quantities behave consistently well. This parametric study has an important implication for realistic applications; that is, data is not always available for some variables, some locations, or some times. The forecast error covariance expresses the relationship between solution variables. It transmits information in the changes of observed variables to unobserved variables.

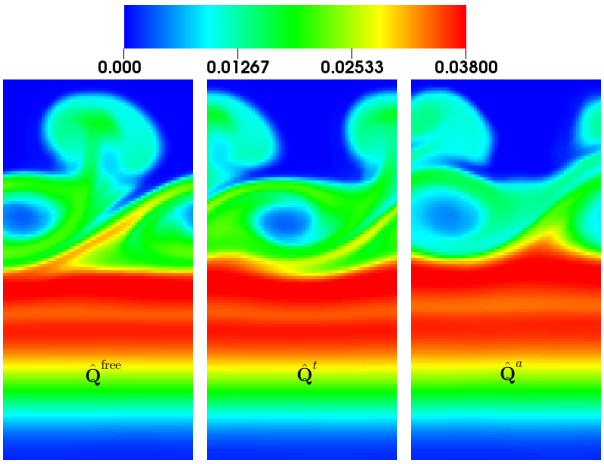
The performance is further assessed by measures of χ^2 and the error reduction of the total cost function. For each assimilation cycle, the error reduction of the total cost function can be evaluated



(a) At 1st DA cycle



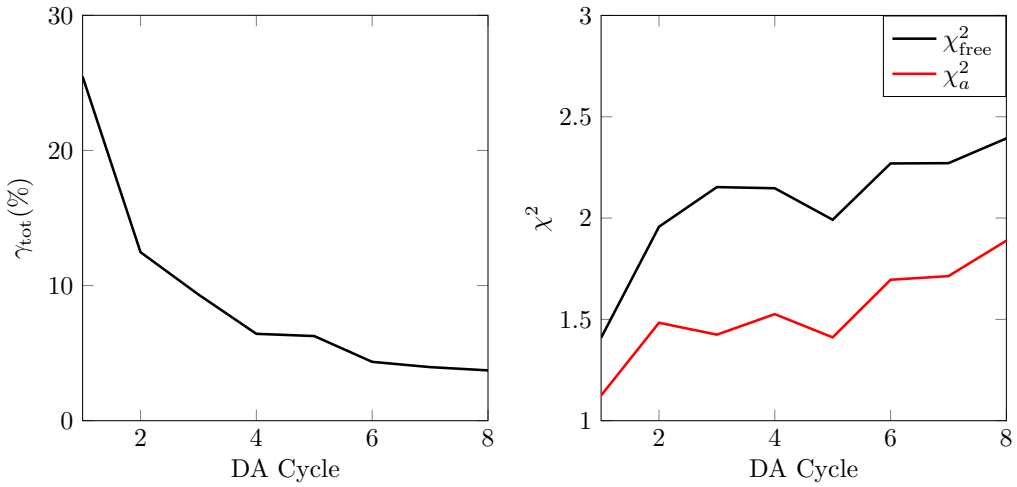
(b) At 5th DA cycle



(c) At 10th DA cycle

Figure 9.6: The comparisons for $\rho c_{\mathrm{H_2O}}$ between the free CFD ($\hat{\mathbf{Q}}^{\text{free}}$), the synthesized truth ($\hat{\mathbf{Q}}^{t}$), and the DA+CFD ($\hat{\mathbf{Q}}^{a}$) solutions at the midpoint of z-axis ($z = 0.125$ mm).

as $\gamma_{tot} = \left(\left| J(\hat{\mathbf{Q}}^f) - J(\hat{\mathbf{Q}}^a) \right| / J(\hat{\mathbf{Q}}^f) \right)$. Figure 9.7a shows the promising error reduction over the cycles, where χ_a^2 is for the data-assimilated CFD simulations and χ_{free}^2 is for the free CFD simulations. As the data assimilation predictions converge toward the truth by the optimization algorithm, the overall impact of data assimilation on the CFD simulations decreases, which is expected. By tracking the χ_a^2 along the assimilation process and comparing it to the free CFD run in Fig. 9.7b, it also shows that χ_a^2 is always smaller than χ_{free}^2 in each DA cycle. This demonstrates that data assimilation is effective in uncertainty reduction by examining γ_{tot}. As seen, the error reduction is about 25% in the first DA cycle, and the error reduction becomes smaller as more DA cycles are performed, as expected when data assimilation is performed properly. Similarly, RMS error can be considered for the quality assessment of data assimilation. Figure 9.8 shows that the analysis RMS error is always smaller than the forecast RMS error during the data assimilation process, indicating the positive impact of data assimilation on the dynamical system with uncertainties.



(a) The historical trajectory of γ_{tot}.

(b) The χ^2 comparison.

Figure 9.7: The trajectory of γ_{tot} and the χ^2 comparison between the data-assimilated CFD (χ_a^2) and the free CFD (χ_{free}^2) simulations over the assimilation cycles.

While the present results show a promising application of data assimilation to combustion, more exciting questions may arise when highly turbulent combusting flows are considered. How

often does data need to be assimilated since a numerical prediction of such a dynamical system may deviate from the "truth" rather quickly? In such a highly chaotic system with many modes, what would happen when all quantities do not behave consistently during data assimilation? A future follow-up study will be conducted to address these questions.

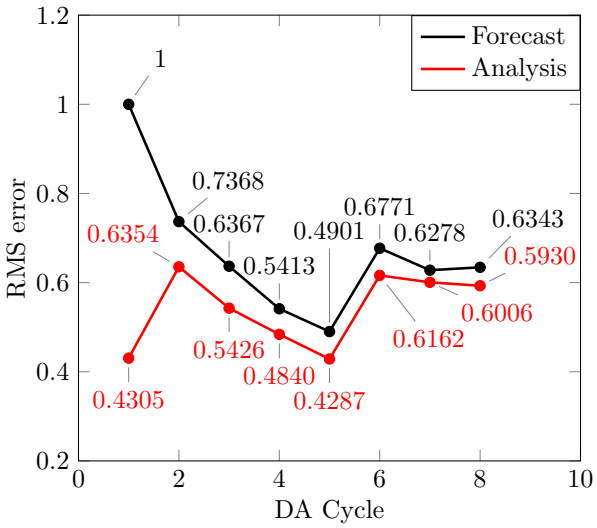


Figure 9.8: The relative RMS errors of the forecast and the analysis solutions.

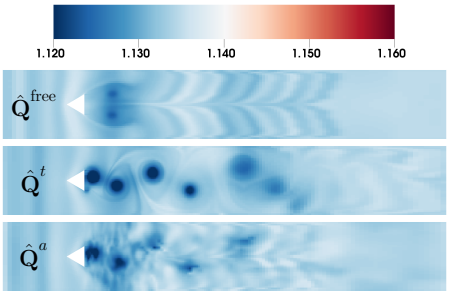
9.2.2 Non-reacting Flow in a Bluff-body Combustor

Figures 9.9-9.10 compare the optimal solution driven by assimilating the 2D-plane data to the free CFD simulation and the synthesized truth for up to 8 cycles. Each figure has 2 columns and 3 rows. The left column shows the 2D contours at the center xy-plane, while the right shows the y-profiles at x = 0.0005 m. Rows from top to bottom are for density, momentum, and energy, respectively. In each contour and line profile, the comparison is made between $\hat{\mathbf{Q}}^a$, $\hat{\mathbf{Q}}^{\text{free}}$, and $\hat{\mathbf{Q}}^t$. Figure 9.9 shows the flow structure is immediately pulled toward the truth after 1 data assimilation cycle and the y-profiles also show such behavior in detail. The solution fields for those quantities (ρ and ρu) who have observed data being assimilated are corrected considerably to the synthesized truth with 4 assimilation cycles, while the non-observed energy takes a couple of more cycles to be corrected, as demonstrated by Fig. 9.10. Despite the discrepancy in the amplitude for ρe between

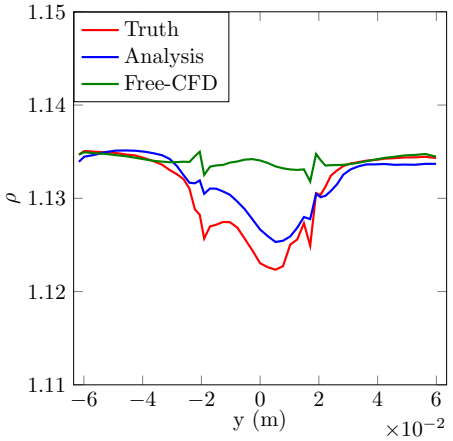
$\hat{\mathbf{Q}}^a$ and $\hat{\mathbf{Q}}^t$, the 1D profiles show the phase of the vortex shedding dynamics is properly corrected. The Strouhal number of the vortex shedding, $\mathrm{St} = fL_y/u$, for the data assimilation result, is approximately 1.6, which is the same as the true value.

Figures 9.11-9.14 compare the optimal solution driven by assimilating the 1D data to the free CFD simulation and the synthesized truth for up to 20 cycles. The layout of each figure is the same as above. In contrast to the above case of assimilating 2D-plane data, the flowfields take much longer to be adjusted or pulled to the truth for this case of assimilating 1D data. This is not surprising. The more quality data is assimilated, the faster the system is being driven toward the optimal state. As shown in Figs. 9.11-9.14 over the 20 cycles, the flow structure behind the bluff body is slowly responding to the data, and the local correction eventually influences the global flowfield as assimilation proceeds. Starting from the 12th cycle, visually, the flow structure is being pulled toward the truth, albeit taking more cycles. From this, factors, such as *where, what* and *when* data should be assimilated, are important to consider for assimilation efficiency and solution accuracy. Similarly, when flows become more and more turbulent, how likely is the "correction" would be washed out if not enough data is assimilated? A follow-up study will focus on highly turbulent flows with more complex geometries. Nevertheless, the present study demonstrates a new application of data assimilation for a moderate turbulent flow with the presence of geometry to improve the prediction of vortex dynamics. The result is promising.

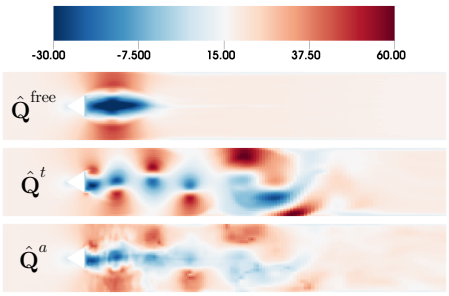
For turbulent flows, statistical analysis of flow fields (spatial or temporal averages) is very useful for engineering designs. For case 2, the time-averaged profiles are compared in Fig. 9.15 showing 3 columns and 4 rows. In each column, the y-profiles of $\hat{\mathbf{Q}}^a$, $\hat{\mathbf{Q}}^{\text{free}}$, and $\hat{\mathbf{Q}}^t$ are compared at x = 0.01 m for ρ, ρu and ρe. Rows are for showing the averaged amplitudes over 1-5, 6-10, 11-15, and 16-20 assimilation cycles, respectively. For all quantities, the averaged behavior is consistent. The phase error is large in the first set of the averaged profiles (i.e., averaged over 1-5 cycles), but significantly reduced for the rest. The averaged amplitude of $\hat{\mathbf{Q}}^a$ is much closer to $\hat{\mathbf{Q}}^t$ for ρu than others. The time-averaged profiles at different spatial locations may likely display somewhat



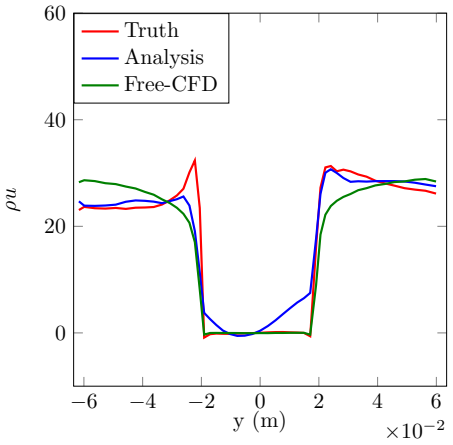
(a) 2D contour of ρ



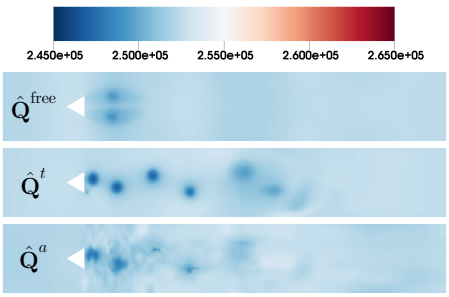
(b) 1D y-profile of ρ



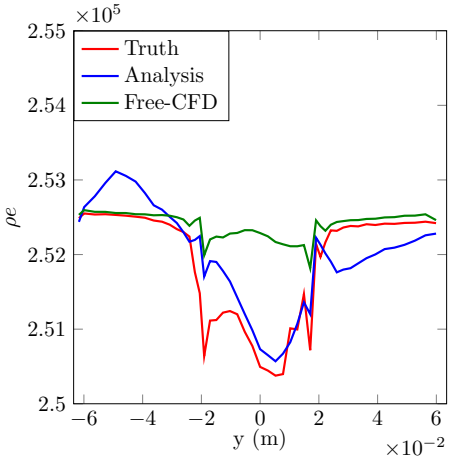
(c) 2D contour of ρu



(d) 1D y-profile of ρu

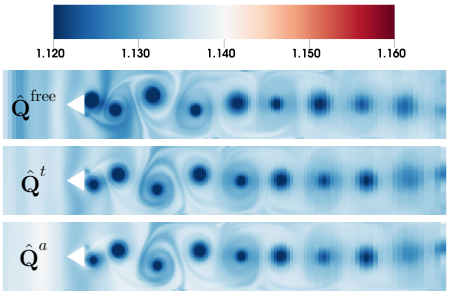


(e) 2D contour of ρe

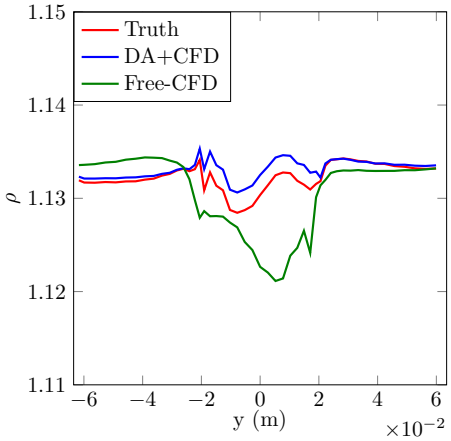


(f) 1D y-profile of ρe

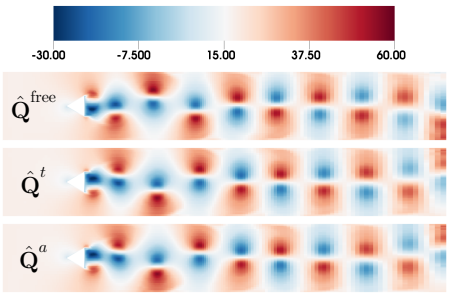
Figure 9.9: The forecast ($\hat{\mathbf{Q}}^{\text{free}}$), the synthesized truth ($\hat{\mathbf{Q}}^{t}$), and the analysis ($\hat{\mathbf{Q}}^{a}$) solutions at 1^{st} DA cycle when assimilating a 2D-plane data.



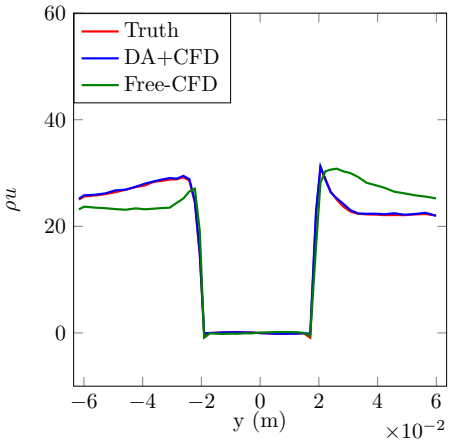
(a) 2D contour of ρ



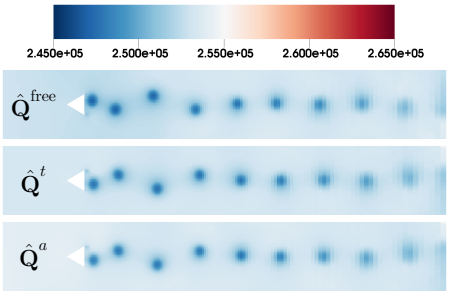
(b) 1D y-profile of ρ



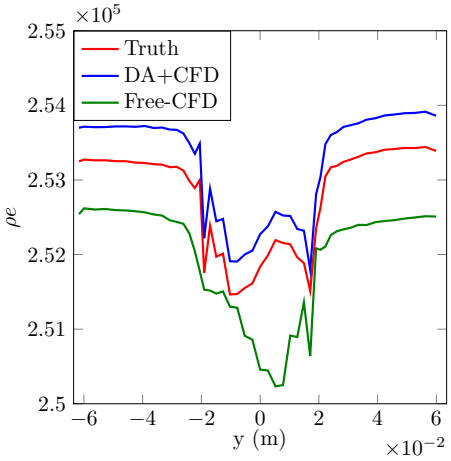
(c) 2D contour of ρu



(d) 1D y-profile of ρu

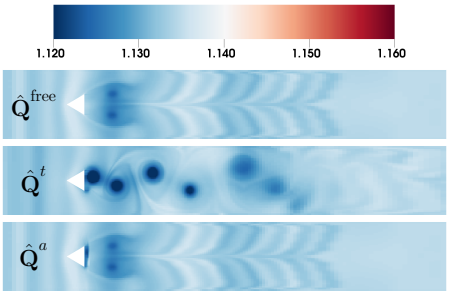


(e) 2D contour of ρe

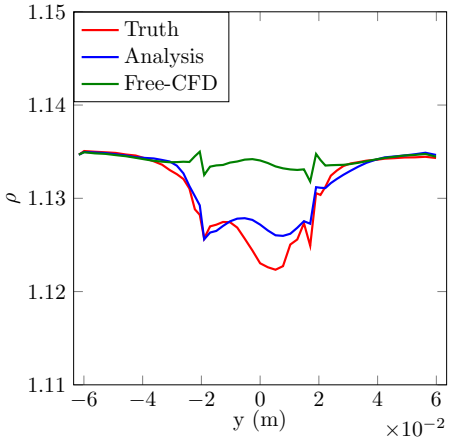


(f) 1D y-profile of ρe

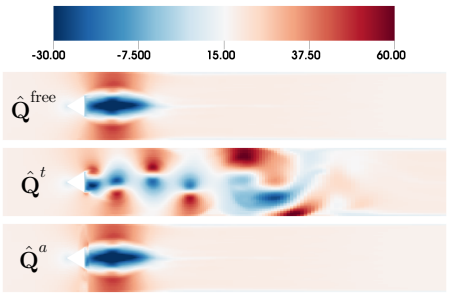
Figure 9.10: The free CFD ($\hat{\mathbf{Q}}^{\text{free}}$), the synthesized truth ($\hat{\mathbf{Q}}^{t}$), and the DA+CFD ($\hat{\mathbf{Q}}^{a}$) solutions at 8^{th} DA cycle when assimilating a 2D-plane data.



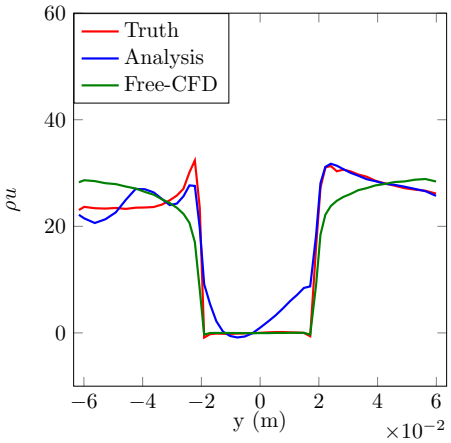
(a) 2D contour of ρ



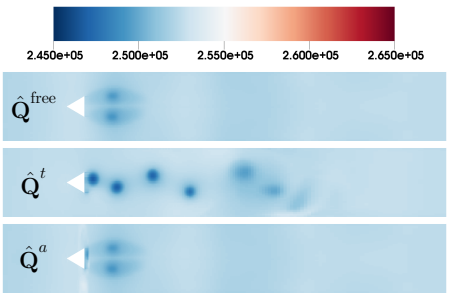
(b) 1D y-profile of ρ



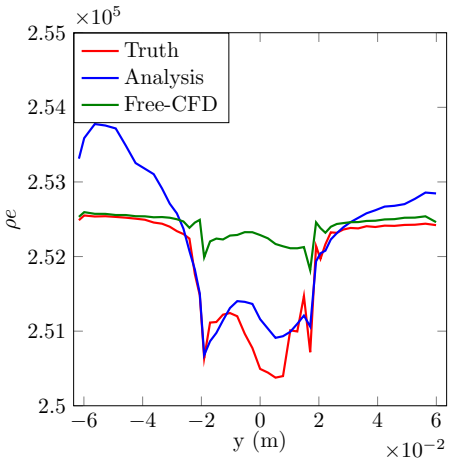
(c) 2D contour of ρu



(d) 1D y-profile of ρu

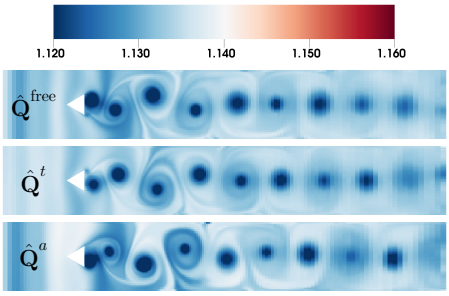


(e) 2D contour of ρe

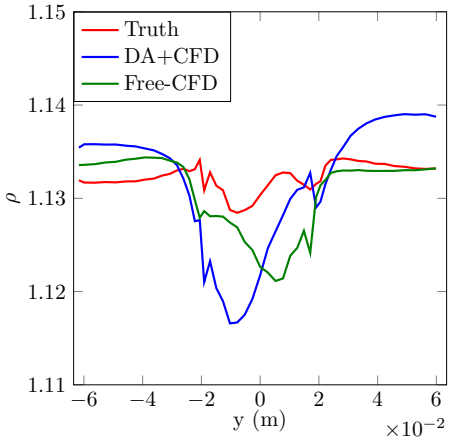


(f) 1D y-profile of ρe

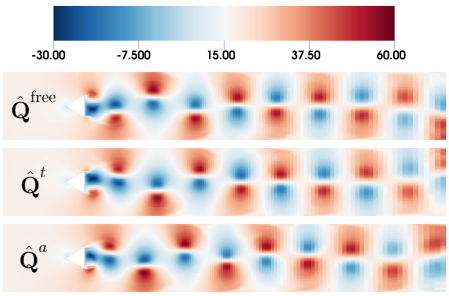
Figure 9.11: The free CFD ($\hat{\mathbf{Q}}^{\text{free}}$), the synthesized truth ($\hat{\mathbf{Q}}^{t}$), and the DA+CFD ($\hat{\mathbf{Q}}^{a}$) solutions at 1^{st} DA cycle when assimilating 1D data.



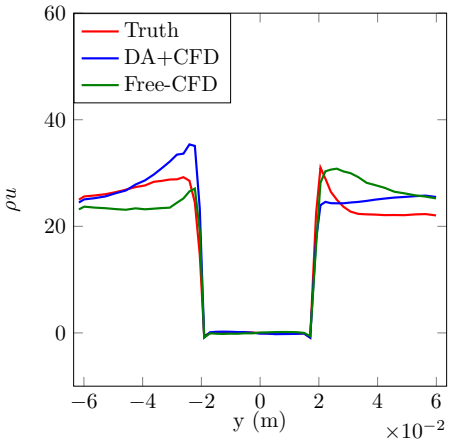
(a) 2D contour of ρ



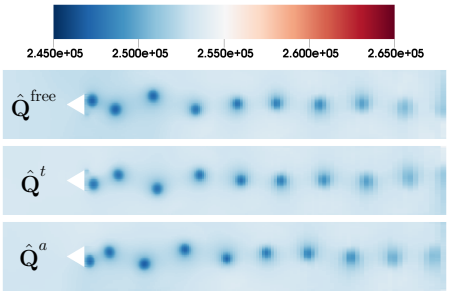
(b) 1D y-profile of ρ



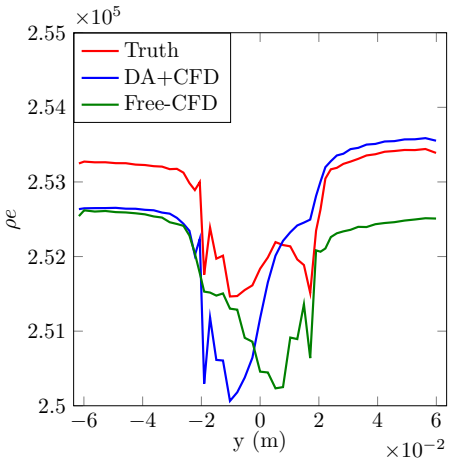
(c) 2D contour of ρu



(d) 1D y-profile of ρu

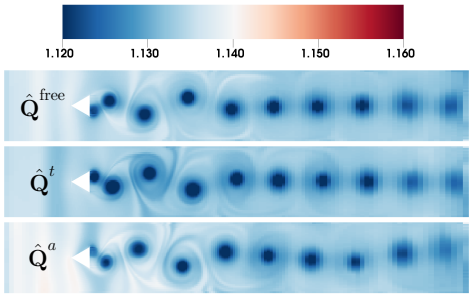


(e) 2D contour of ρe

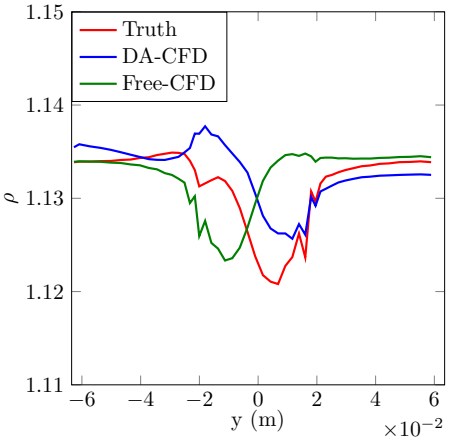


(f) 1D y-profile of ρe

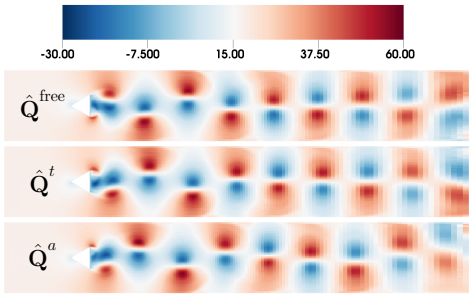
Figure 9.12: The free CFD ($\hat{\mathbf{Q}}^{\text{free}}$), the synthesized truth ($\hat{\mathbf{Q}}^{t}$), and the DA+CFD ($\hat{\mathbf{Q}}^{a}$) solutions at 8^{th} DA cycle when assimilating 1D data.



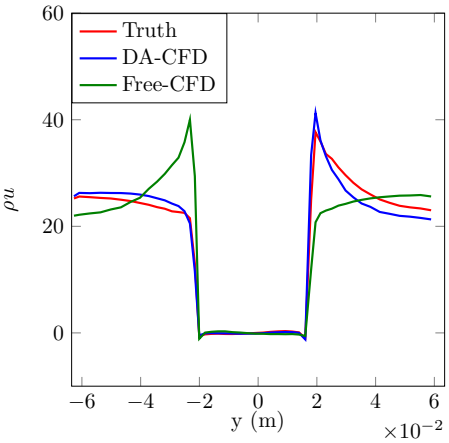
(a) 2D contour of ρ



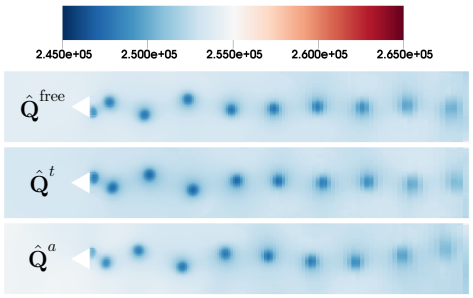
(b) 1D y-profile of ρ



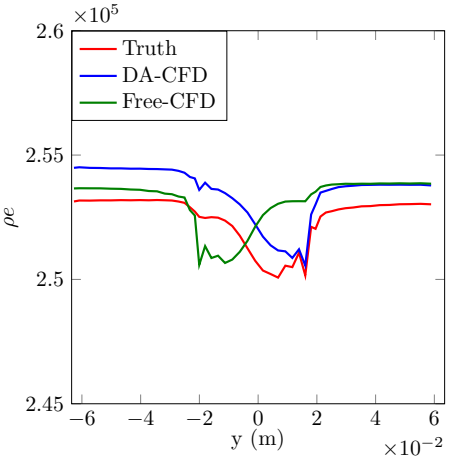
(c) 2D contour of ρu



(d) 1D y-profile of ρu

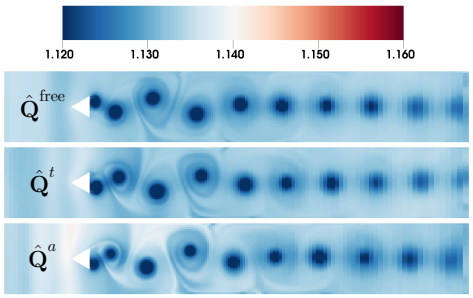


(e) 2D contour of ρe

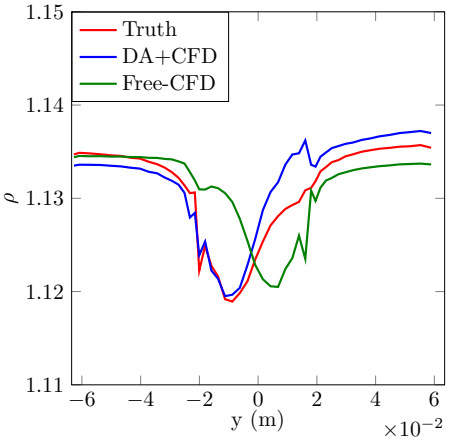


(f) 1D y-profile of ρe

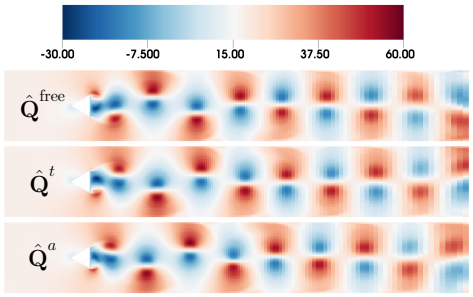
Figure 9.13: The free CFD ($\hat{\mathbf{Q}}^{\text{free}}$), the synthesized truth ($\hat{\mathbf{Q}}^{t}$), and the DA+CFD ($\hat{\mathbf{Q}}^{a}$) solutions at 12^{th} DA cycle when assimilating 1D data.



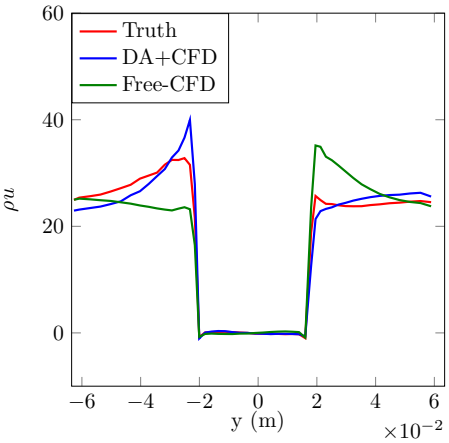
(a) 2D contour of ρ



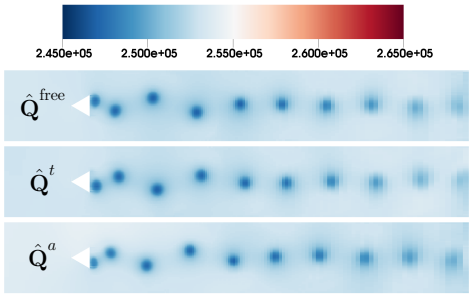
(b) 1D y-profile of ρ



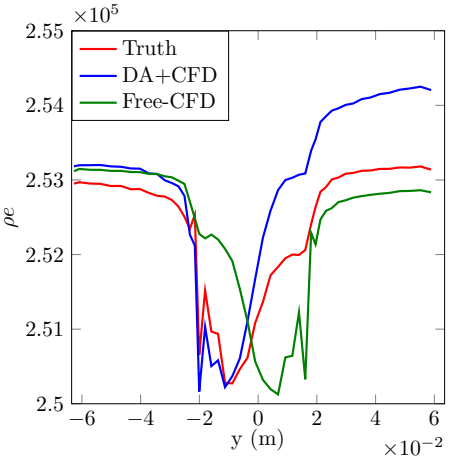
(c) 2D contour of ρu



(d) 1D y-profile of ρu

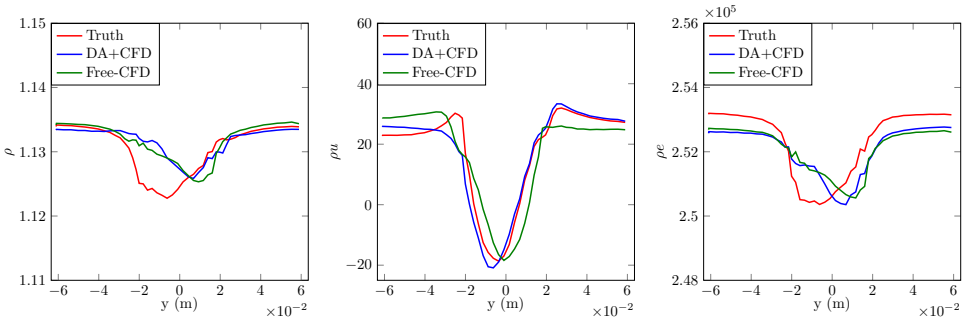


(e) 2D contour of ρe

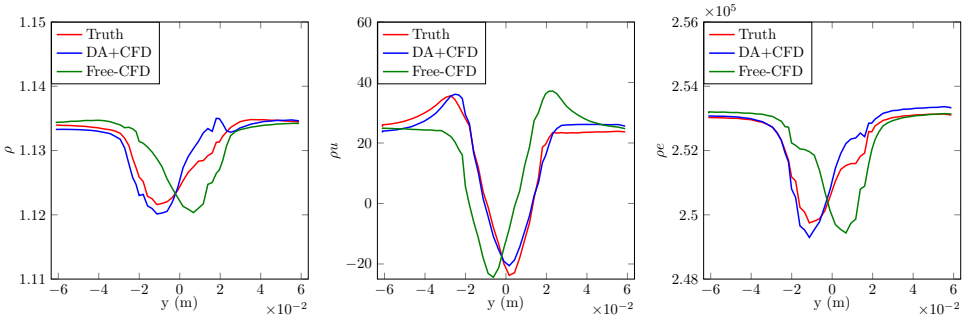


(f) 1D y-profile of ρe

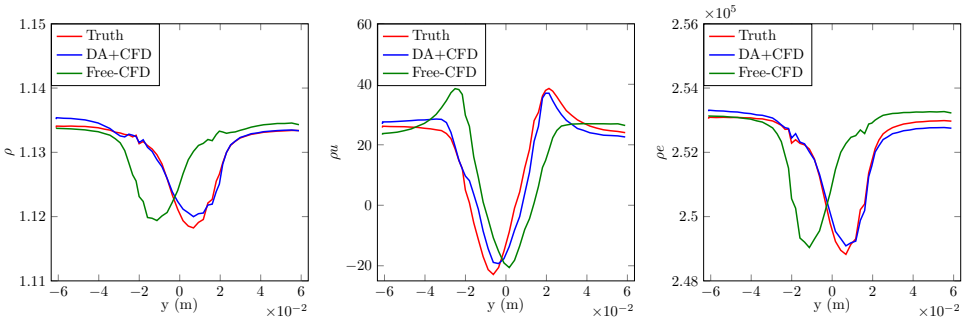
Figure 9.14: The free CFD ($\hat{\mathbf{Q}}^{\text{free}}$), the synthesized truth ($\hat{\mathbf{Q}}^{t}$), and the DA+CFD ($\hat{\mathbf{Q}}^{a}$) solutions at 20^{th} DA cycle when assimilating 1D data.



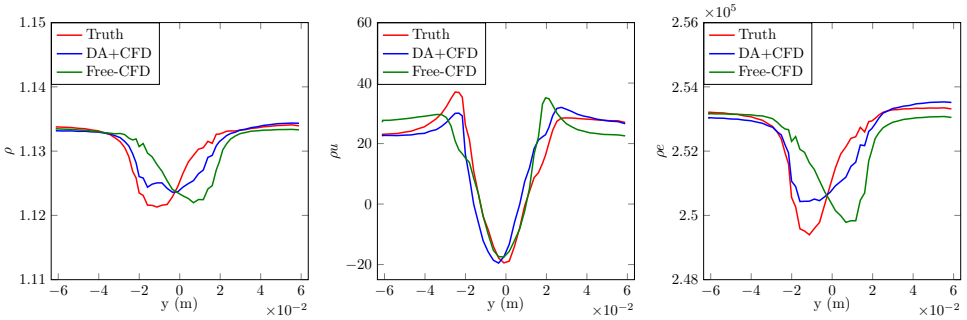
(a) Time-averaged ρ (left), ρu (mid), and ρe (right) over $5\mathrm{T}_{\mathrm{DA}}$.



(b) Time-averaged ρ (left), ρu (mid), and ρe (right) over (5 - 10)T_{DA}.



(c) Time-averaged ρ (left), ρu (mid), and ρe (right) over (10 - 15)T_{DA}.



(d) Time-averaged ρ (left), ρu (mid), and ρe (right) over (15 - 20)T_{DA}.

Figure 9.15: The y-profiles at x = 0.01 m for the time-averaged free CFD ($\hat{\mathbf{Q}}^{\mathrm{free}}$), the time-averaged synthesized truth ($\hat{\mathbf{Q}}^{t}$), and the time-averaged DA+CFD ($\hat{\mathbf{Q}}^{a}$) where 1D data is assimilated.

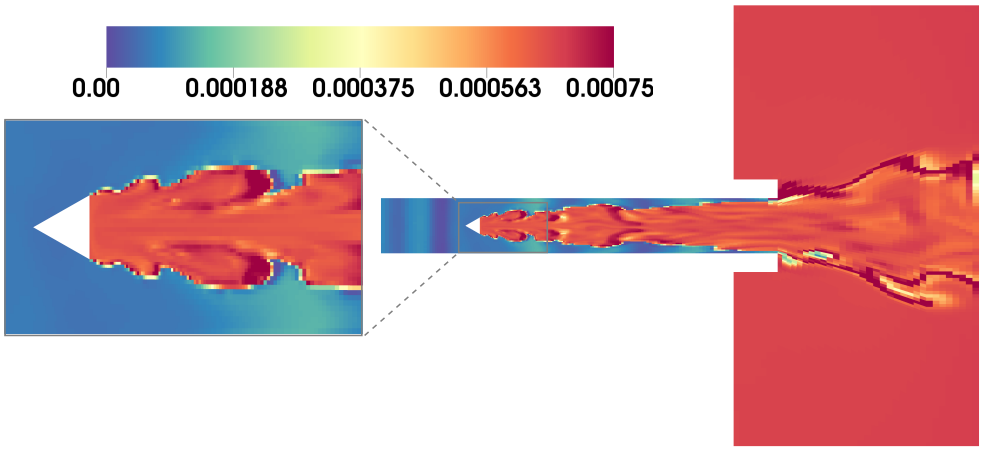
different behavior. Nevertheless, statistical analysis is important for assessing data assimilation performance and fluid dynamics for engineering purposes.

9.2.3 Reacting Flow in a Bluff-body Combustor

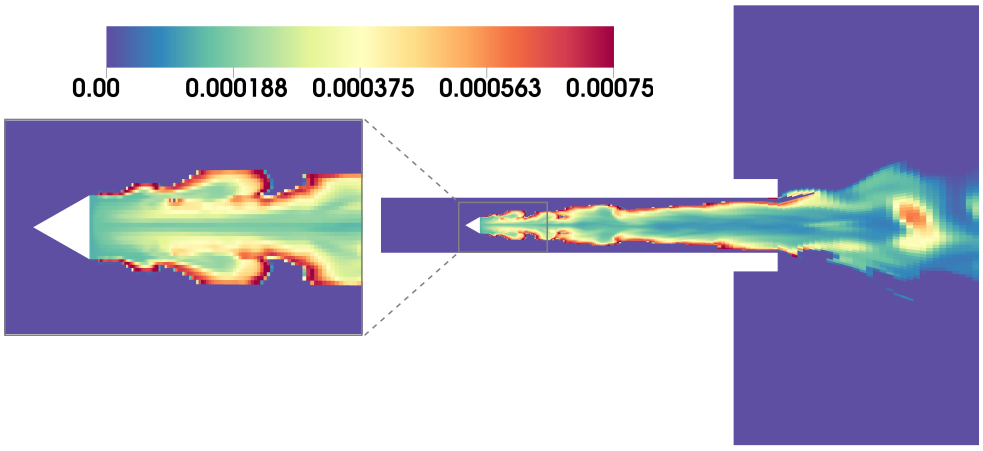
Imperfect Initial Condition

The instantaneous contours of the forecast states for ρe, ρc_{OH}, and $\rho c_{\mathrm{O_2}}$ at the 50th DA cycle are shown in Fig. 9.16 from the top to the bottom rows. Although many intermediate cycles are examined for all solution quantities, the final cycle is shown to demonstrate the resulting overall flame shapes and flow dynamics. In addition, the close-up of the flame immediately behind the bluff body is displayed for the viewing window. The overall flame shape is symmetric about the centerline (y/H = 0) parallel to the axial axis, agreeing with the experimental observation [90], so is clearly shown by the close-ups. Importantly, the symmetry maintains throughout the combustor, which was not the case in the 3D flame simulations as reported in literature [86–89, 92–95]. To see how assimilating experimental data changes the states, Figs. 9.17 and 9.18 show the difference between the forecast and the analysis states, revealing the spatial and temporal effects of DA on ρc_{OH} and ρe. These two solution variables are chosen to represent the observed and unobserved quantities, respectively. While spatial effects can be seen from individual contours, temporal effects are displayed by the sequence of contours in rows from top to bottom.

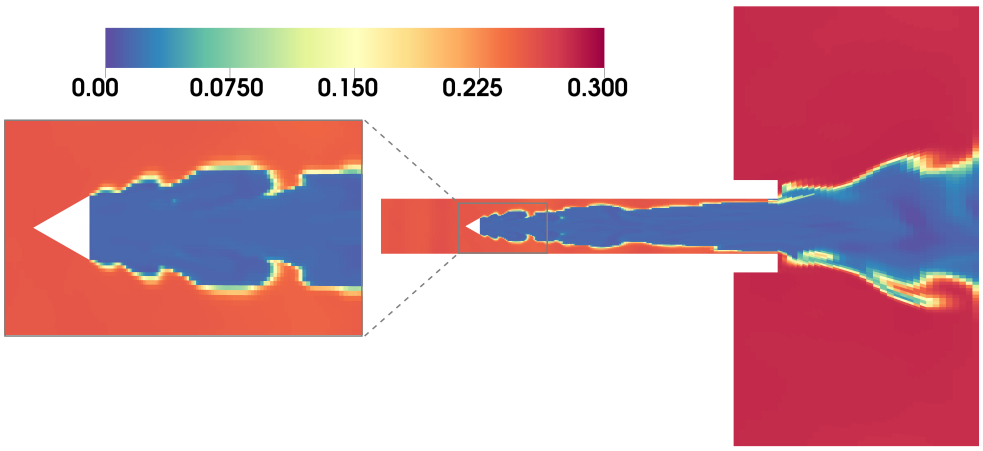
Figure 9.17 shows that assimilation changes the OH mass in the recirculation region near the trailing edge of the bluff body and the flame fronts at the 8^{th}, 20^{th}, 40^{th}, and 50^{th} DA cycles. But, the primary changes are found along the flame fronts after 40 DA cycles. While ρc_{OH} is used for assimilation, ρe is an unobserved quantity. Assimilating ρc_{OH} is expected to impact unobserved variables through nonlinear coupling dynamics. From Fig. 9.18, the footprints of changes in ρe follow the flame fronts closely. Large spatial gradients in the changes, $\vec{\nabla}\left((\rho e)^a - (\rho e)^f\right)$, are generally distributed along the flame fronts, the shear layer, and the outer edge of the recirculation zones, which are exactly the regions with complex fluid and flame structures. Similarly, large variations over time occur in the same regions. The flow and flame fields are physically and



(a) ρe

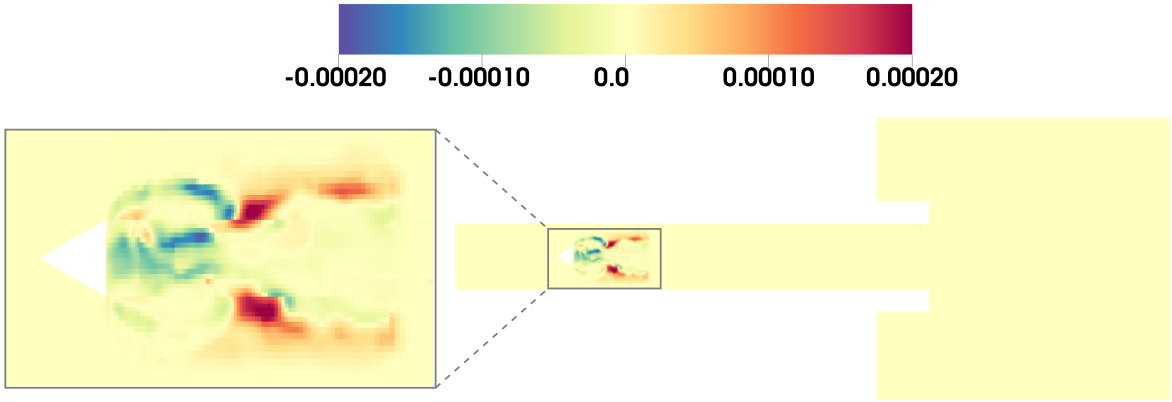


(b) ρc_{OH}

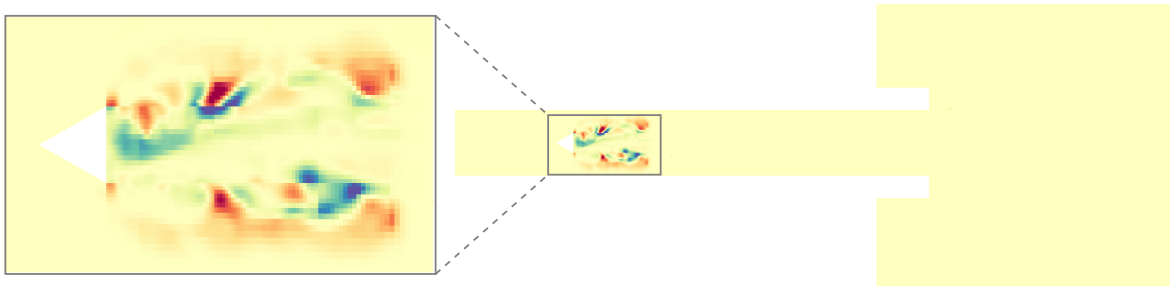


(c) $\rho c_{\mathrm{O_2}}$

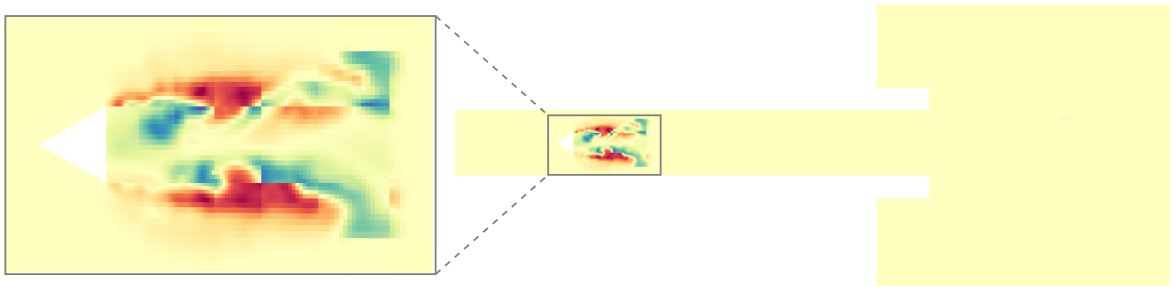
Figure 9.16: Instantaneous forecast state contours at the 50^{th} cycle. *top:* ρe, *middle:* ρc_{OH}, and *bottom:* $\rho c_{\mathrm{O_2}}$.



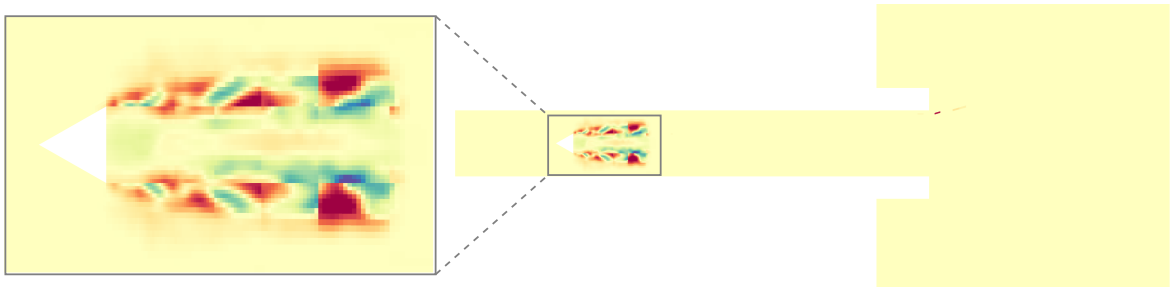
(a) 8^{th} cycle



(b) 20^{th} cycle

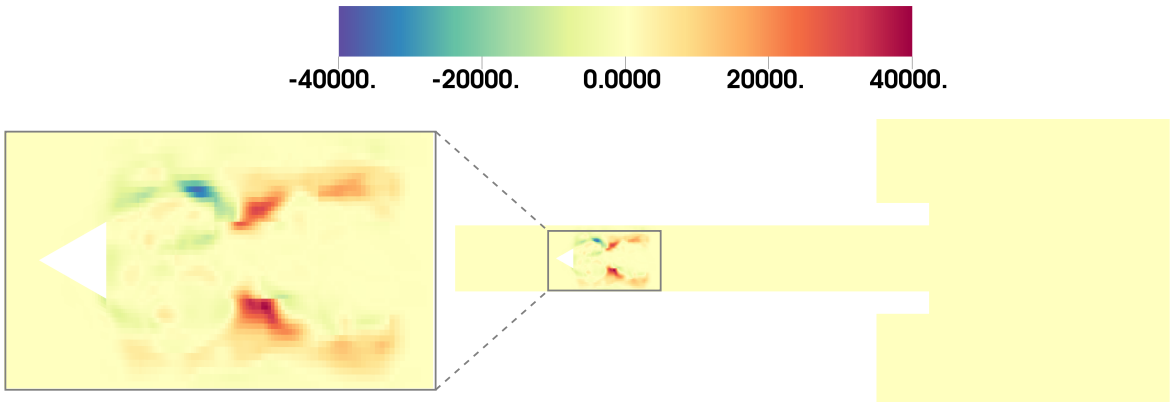


(c) 40^{th} cycle

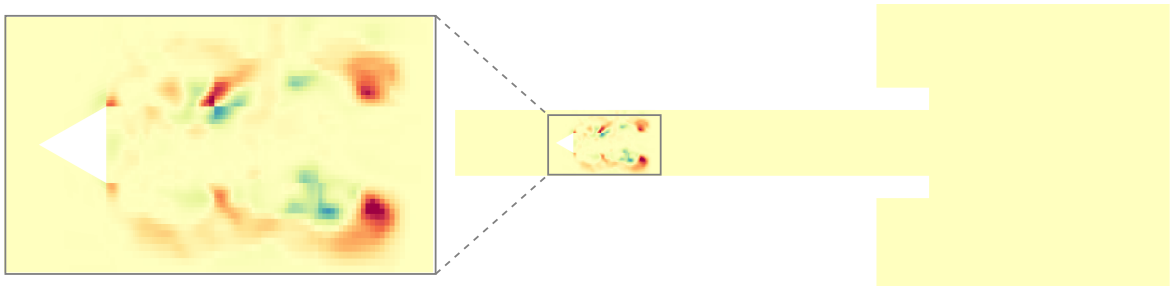


(d) 50^{th} cycle

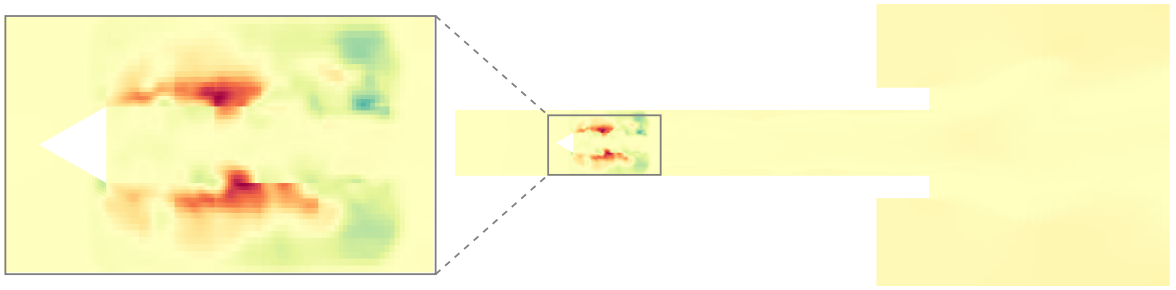
Figure 9.17: The model state difference ($\hat{\mathbf{Q}}^a - \hat{\mathbf{Q}}^f$) contours of ρc_{OH} at the 8^{th}, 20^{th}, 40^{th}, and 50^{th} DA cycles.



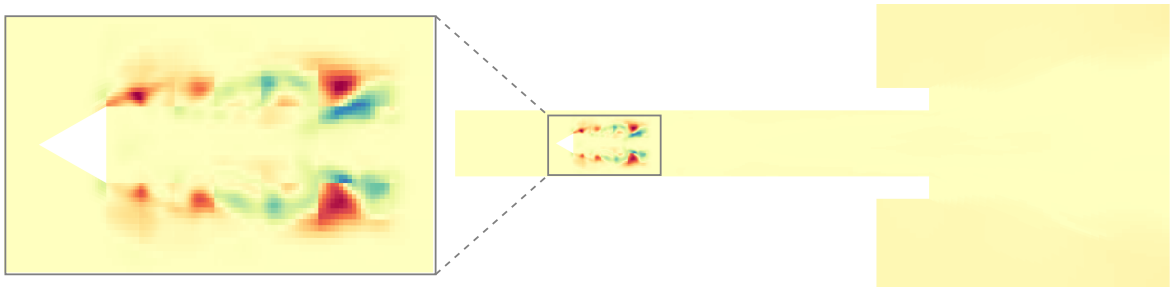
(a) 8^{th} cycle



(b) 20^{th} cycle



(c) 40^{th} cycle



(d) 50^{th} cycle

Figure 9.18: The model state difference ($\hat{\mathbf{Q}}^a - \hat{\mathbf{Q}}^f$) contours of ρe at the 8^{th}, 20^{th}, 40^{th}, and 50^{th} DA cycles.

numerically responding to DA, nonlinearly and dynamically, as desired. This consistent behavior is very encouraging in two perspectives: (i) model is responsive/sensitive to data in the regions where information transfer is high and interactions between data, model, and physics are compatible; (ii) data being assimilated is generally reliable. All solution variables are examined (not shown here) and demonstrate similar characteristics as that illustrated by ρc_{OH} and ρe. We also examined the eigensystem of $\mathbf{P}_a$, confirming that the magnitude and direction of information transfer correspond to dominant eigenvalues and eigenvectors. We assert that the DA+CFD system works properly.

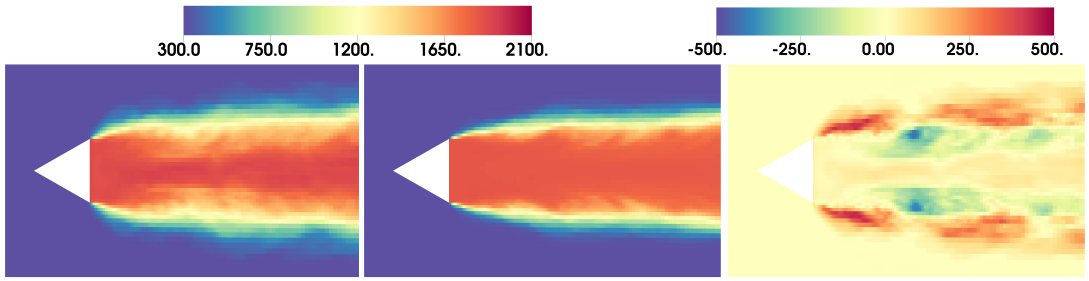
Figure 9.19 shows the time-averaged temperature contours in the viewing window, comparing the flame shapes obtained by the DA+CFD and the free CFD simulations and showing the difference between the two. In the matrix of subfigures, 5 rows$\times$3 columns, from top to bottom, each row shows the time-averaged contours over 10 DA cycles; that is 1-10, 11-20, 21-30, 31-40, and 41-50, respectively. From left to right, each column shows the flames obtained by the DA+CFD, the free run, and their difference, respectively. Clearly observed is that the DA+CFD flames are more wrinkled and slightly further expanded into the shear layers than the free CFD flames. The increase in the spatial gradient of temperature in the DA+CFD contrasts with the free CFD. The difference highlights the flame features, such as symmetry, thickened flame fronts, and wrinkled surfaces, all converging toward the proper flame shapes. Other quantities, such as velocities and species mass fractions, behave consistently and compatibly with the temperature. Indeed, this is not easy to achieve when the partial differential equations (PDEs) being solved include 29 solution variables with turbulence and combustion. Solutions can be easily destroyed if instability in one single variable (due to perturbations introduced during the DA process) can not be suppressed in this large, nonlinear, chaotic dynamical system. This demonstration of the stability and robustness of the DA+CFD system increases our confidence in applying the DA+CFD system to the 3D DA study in the follow-up work.

Upon assimilation, data introduces perturbations to the model. A desirable characteristic of DA is to create a dynamically-balanced analysis state. If a new analysis state contains unphysical wave modes, it can lead to spurious noise in prediction and eventually to erroneous prediction.

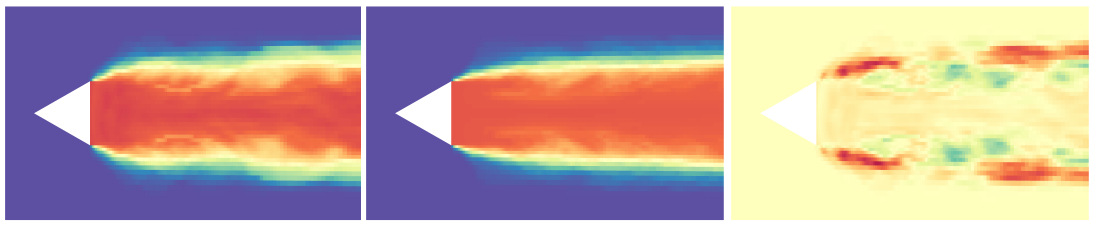
Therefore, it is important to assess the relative contribution of unphysical perturbations in the analysis and ensuing prediction. In our case, the unphysical perturbations are computed and found to be in the range of 0.2-3.5% of the reference conserved quantity on average after each DA cycle. MLEF appears to sufficiently suppress the erroneous modes introduced by assimilation, and with proper assimilation frequency, the unwanted waves are expected to be smoothed out along the propagation of the forward model.

Time-averaged flames (e.g., over the entire 50 DA cycles) are more useful to engineering. Figures 9.20 and 9.21 show the averaged solution contours for density, axial momentum, specific total energy, OH mass, CH_2O mass, and temperature. Each figure contains 2 columns of subfigures, with the left showing the DA+CFD solution and the right for the free CFD solution. In Fig. 9.20, the contours of ρ, ρu, and ρe are placed in the top, middle, and bottom rows, respectively. Figure 9.21 shows the ρc_{OH}, $\rho c_{\mathrm{CH_2O}}$, and temperature. Overall, the time-averaged flame shape is fluffed up slightly by DA. Moreover, the regions of recirculation behind the bluff body and the shear layers are more visually different between the DA+CFD and CFD solutions than others. Specifically, more fine structures appear in the DA+CFD solution while the free CFD contour is smooth. The averaged solution over 50 DA cycles is relatively smoother than the averaged solution over 10 DA cycles, which can be seen more clearly in the close-ups by Fig. 9.22. Remarkably, the DA+CFD flames remain symmetric. In the plots for $\rho c_{\mathrm{CH_2O}}$, two tiny spotty stripes along the outer edges of the recirculation zone appear behind the bluff body. This anomalous distribution turns out to be harmless to the solution stability. In fact, we have observed this phenomenon (spotty high-value $\rho c_{\mathrm{CH_2O}}$) in a 3D reference simulation where AMR is used without DA. They appear, disappear, and re-appear; sometimes short-lived, highly localized, this miniature explosion of radical CH_2O apparently has no damaging impact on the stability, which is good.

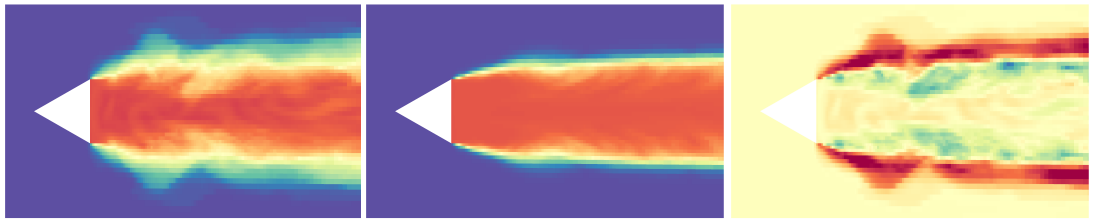
Furthermore, Figs. 9.23 and 9.24 compare the averaged y-profiles between the DA+CFD and free CFD predictions at $x/D = 0.0394$ and $x/D = 2.62$, respectively, for ρ, u, ρe, ρc_{OH}, ρc_{CH2O}, and temperature. Both locations have observed data for velocity and OH mass, which are assimilated to the model. At $x/D = 0.0394$, data appears to impact ρ, ρe, and temperature



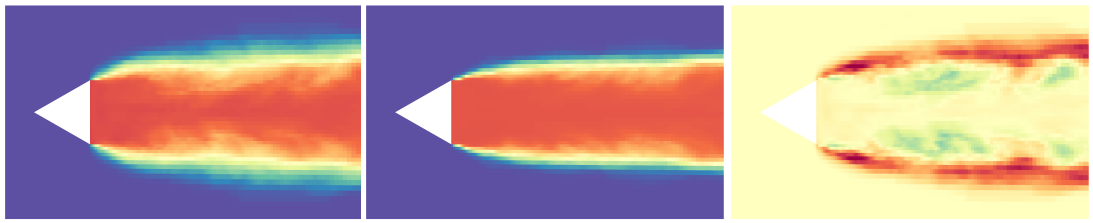
(a) Average over 1^{st} to 10^{th} DA cycles.



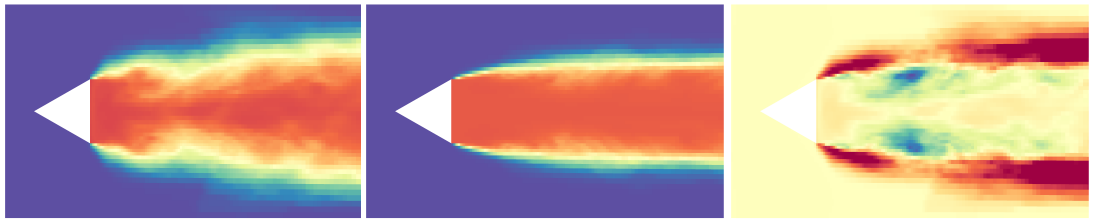
(b) Average over 11^{st} to 20^{th} DA cycles.



(c) Average over 21^{st} to 30^{th} DA cycles.

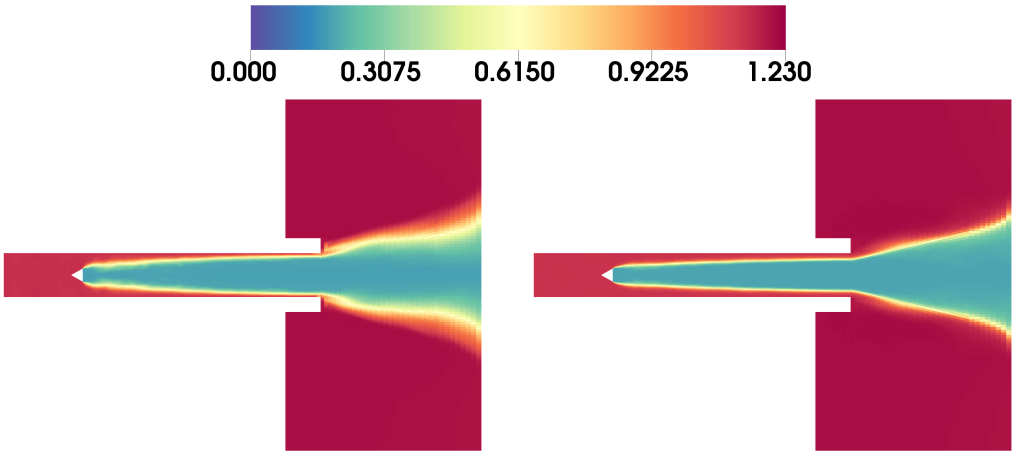


(d) Average over 31^{st} to 40^{th} DA cycles.

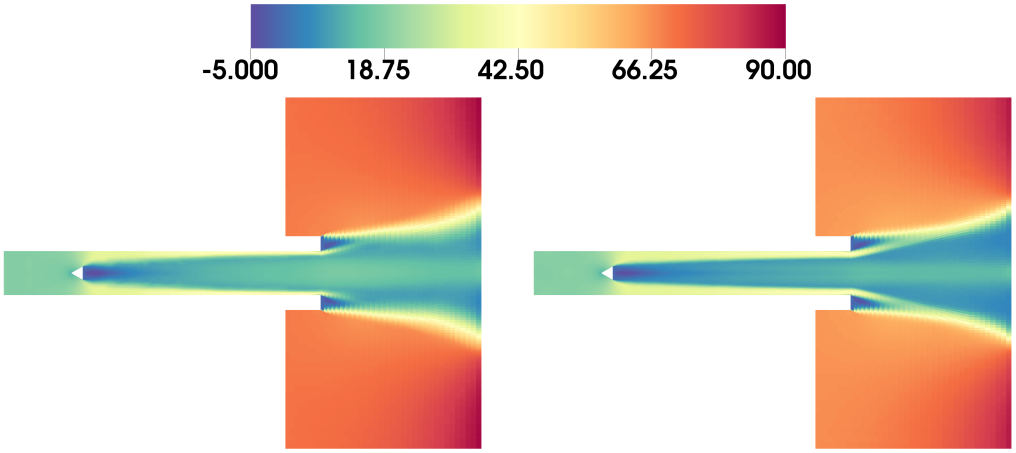


(e) Average over 41^{st} to 50^{th} DA cycles.

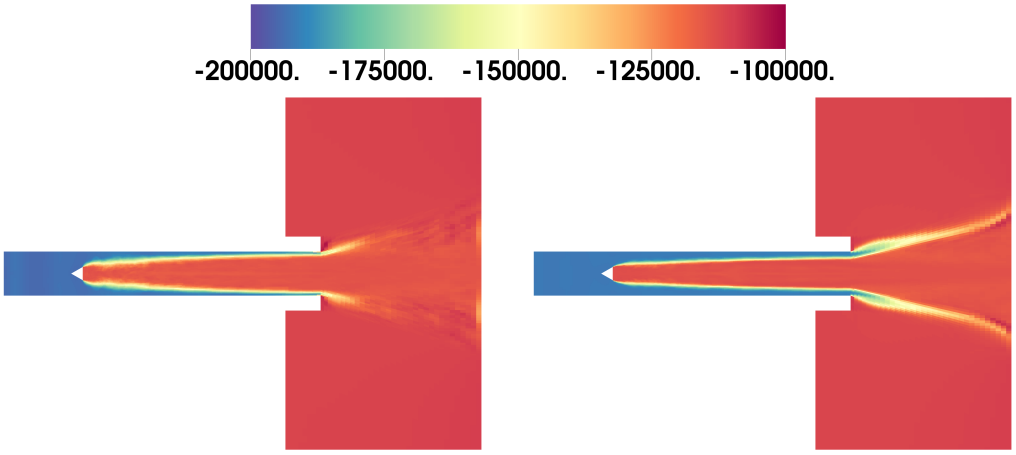
Figure 9.19: Time-averaged temperature contours comparisons for the viewing window. *left:* DA+CFD, *middle:* CFD, and *right:* Difference between DA+CFD and CFD.



(a) ρ.

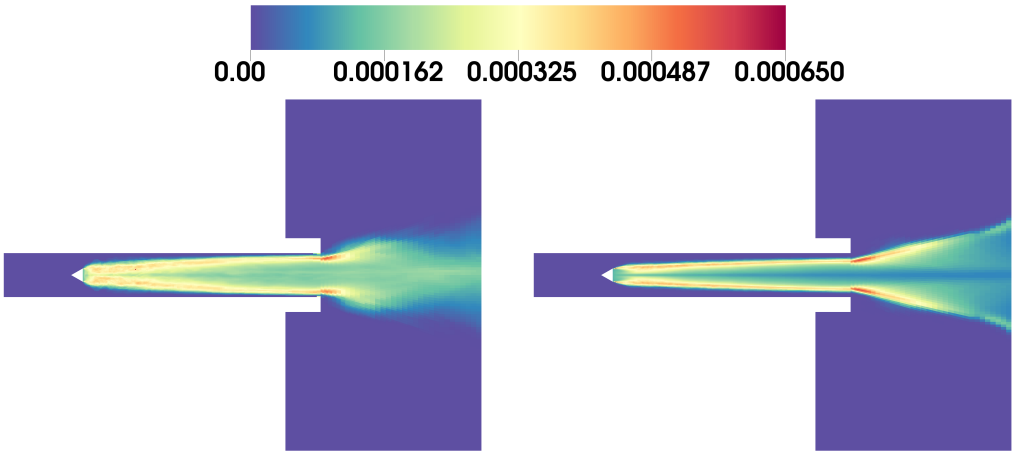


(b) ρu.

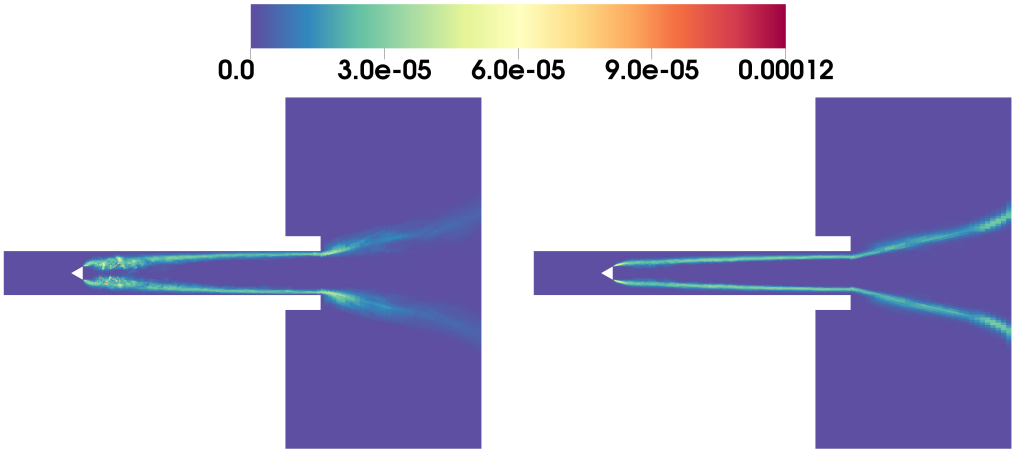


(c) ρe.

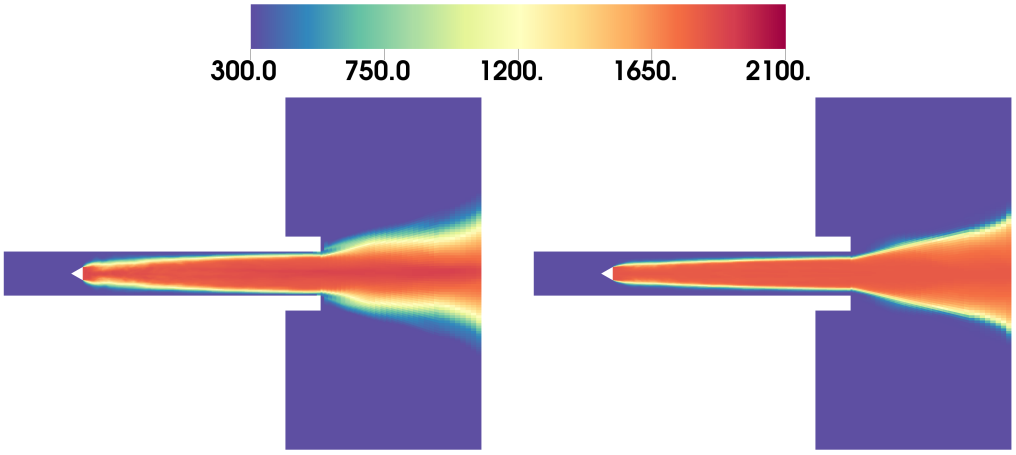
Figure 9.20: Time-averaged contours for ρ, ρu, and ρe over the time frame of 50 DA cycles. *left:* DA+CFD, *right:* CFD.



(a) ρc_{OH}.

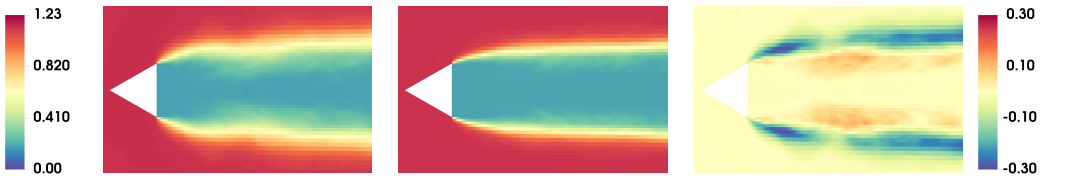


(b) $\rho c_{\mathrm{CH_2O}}$.

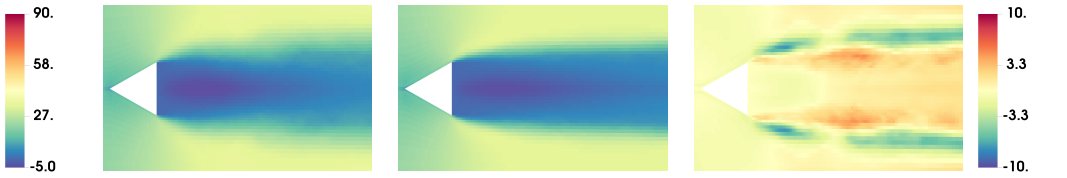


(c) Temperature.

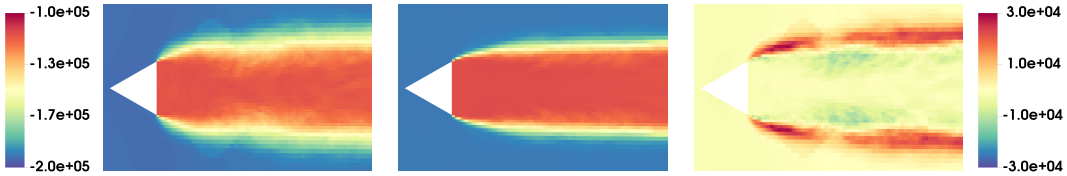
Figure 9.21: Time-averaged contours for ρc_{OH}, $\rho c_{\mathrm{CH_2O}}$ and temperature over the time frame of 50 DA cycles. *left:* DA+CFD, *right:* CFD.



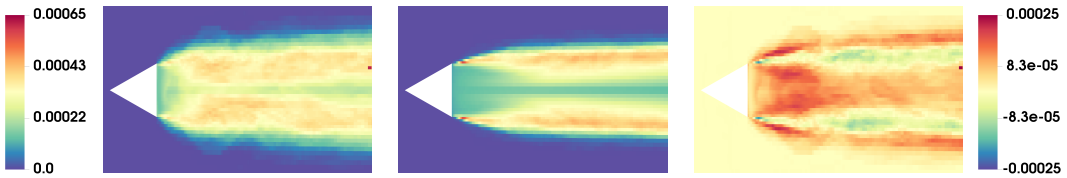
(a) ρ.



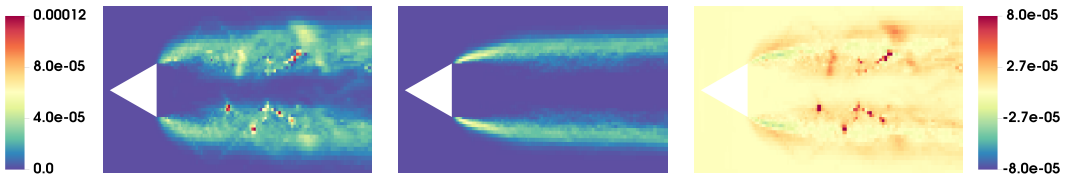
(b) ρu.



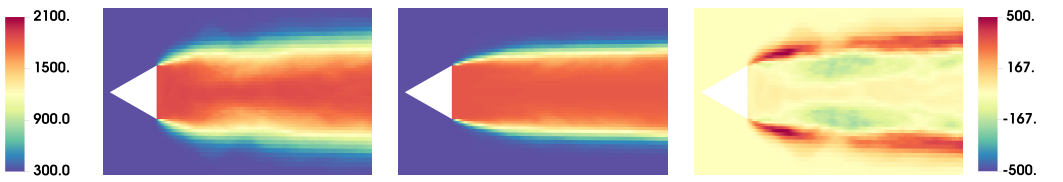
(c) ρe.



(d) ρc_{OH}.

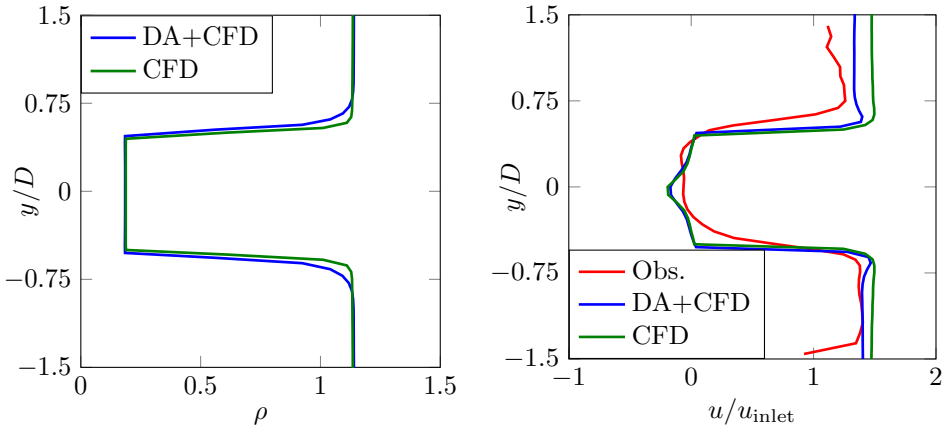


(e) $\rho c_{\mathrm{CH_2O}}$.



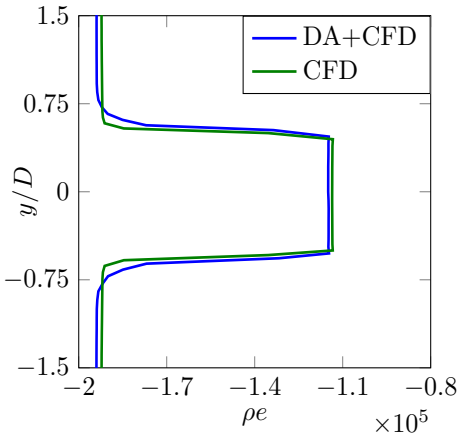
(f) Temperature.

Figure 9.22: Time-averaged contours for ρ, ρu, ρe, ρc_{OH}, $\rho c_{\mathrm{CH_2O}}$, and temperature in the viewing window over the time frame of 50 DA cycles. *left:* DA+CFD, *middle:* CFD, *right:* the difference.

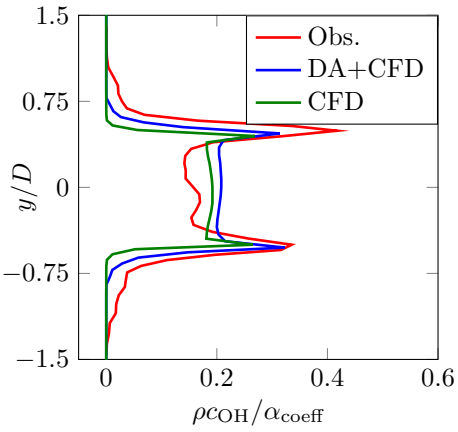


(a) ρ.

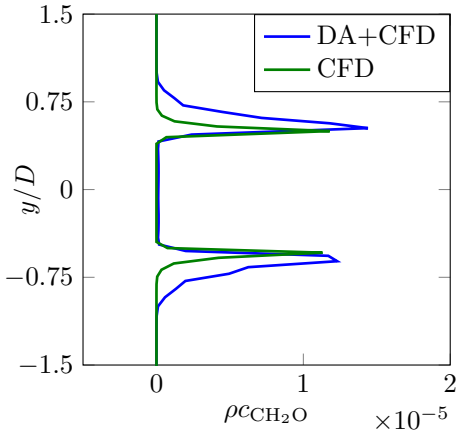
(b) u.



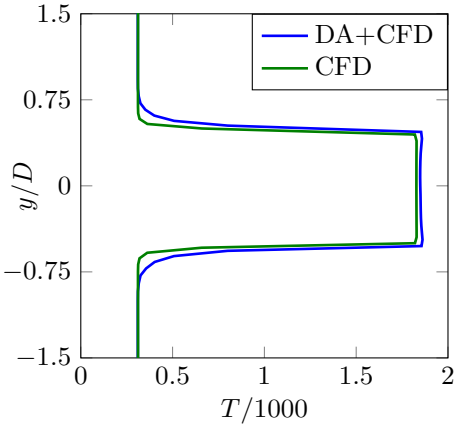
(c) ρe.



(d) ρc_{OH}.

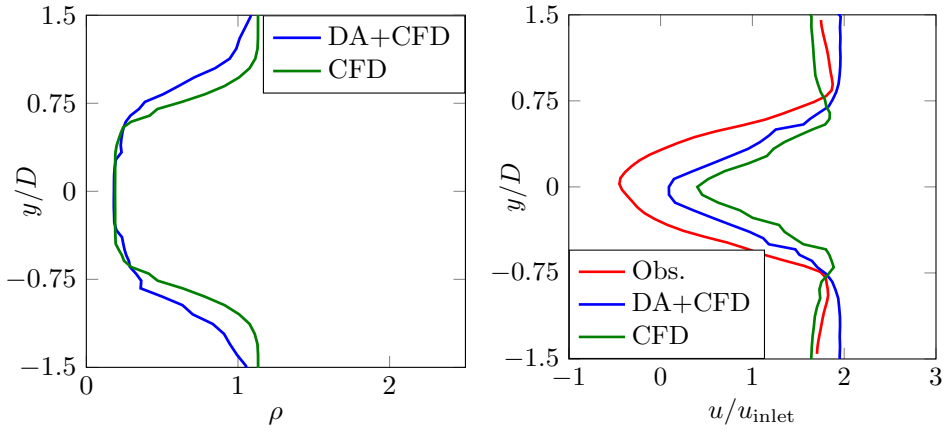


(e) $\rho c_{\mathrm{CH_2O}}$.



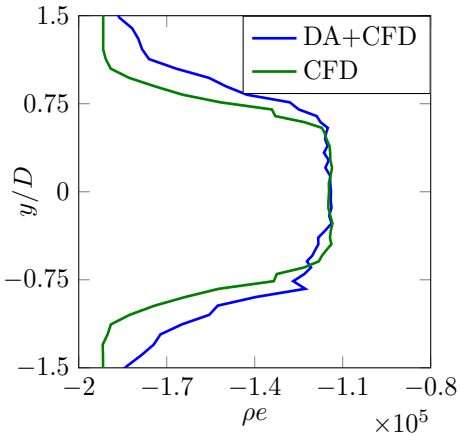
(f) Temperature.

Figure 9.23: Time-averaged 1D y-profiles comparisons at $x/H = 0.0394$ over the time frame of 50 DA cycles.

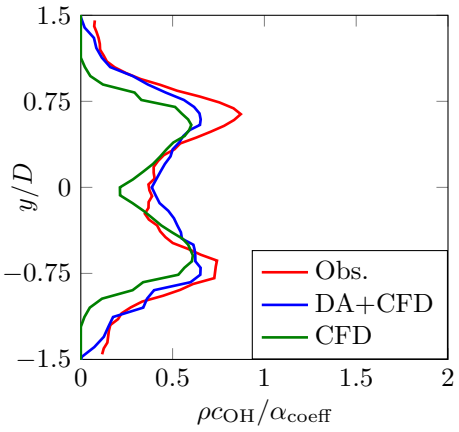


(a) ρ.

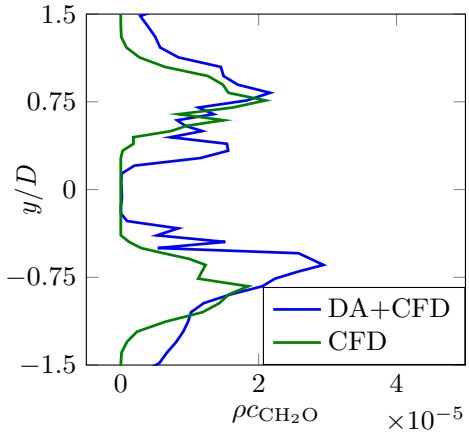
(b) u.



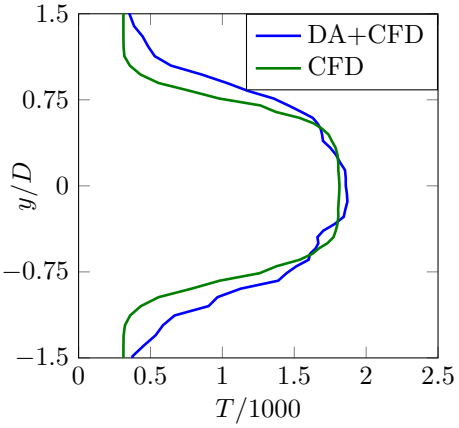
(c) ρe.



(d) ρc_{OH}.



(e) $\rho c_{\mathrm{CH_2O}}$.



(f) Temperature.

Figure 9.24: Time-averaged 1D y-profiles comparisons at $x/H = 2.62$ over the time frame of 50 DA cycles.

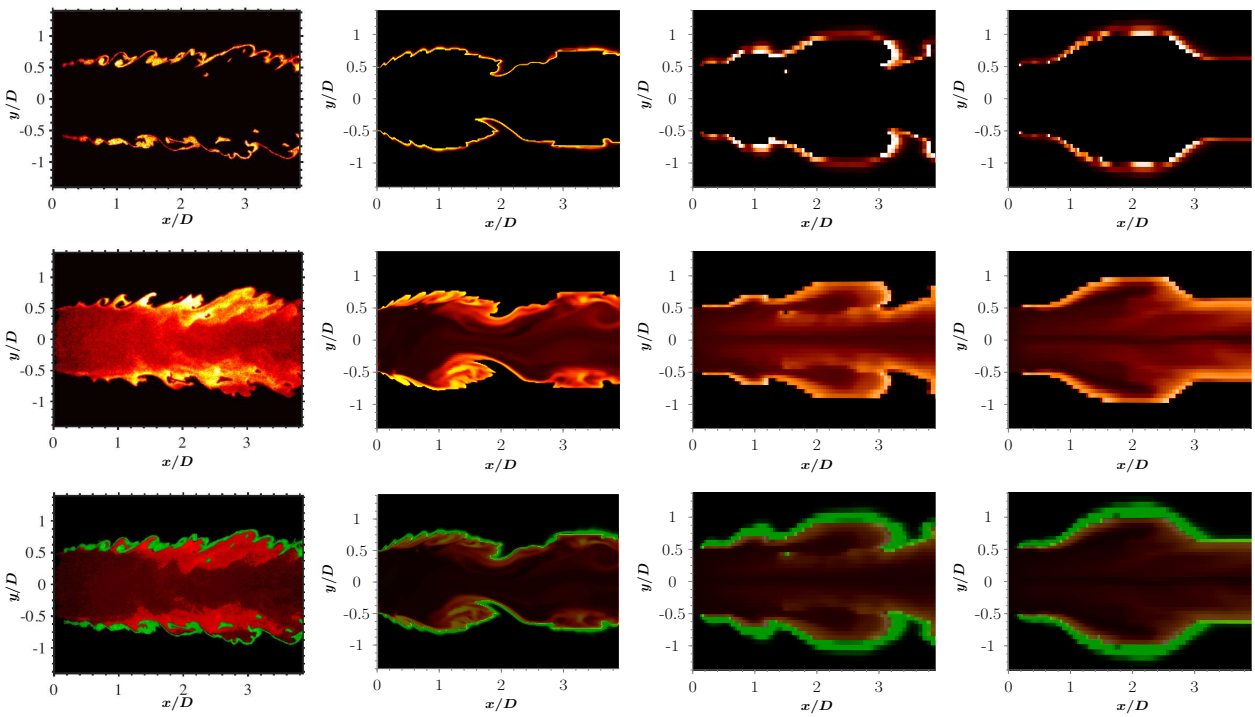


Figure 9.25: Instantaneous ρc_{CH2O} and ρc_{OH} comparisons between the experimental measurements and the simulations. *from left to right:* Column 1: experiment, Column 2: 3D AMR simulation, Column 3: 2D DA+CFD with a coarse grid, and Column 4: 2D free-CFD with the same coarse grid. In each column, the rows from top to bottom show the normalized ρc_{CH2O}, the normalized ρc_{OH}, and the overlap of ρc_{CH2O} (green) and ρc_{OH} (red), respectively.

insignificantly, but the positive impact on the observed quantities (axial velocity and OH mass) is apparent. From the profiles further downstream, positive DA impacts on model predictions for all quantities are noticeable. The profiles indicate flame fronts with more wrinkles, consistent with the observation contour. The solution structures manifest an effective DA process, driving model prediction toward data. However, DA performance saturates when the computational model contains structural errors or gravely under-resolves critical flow and flame dynamics. Figure 9.25 compares the instantaneous ρc_{CH2O} and ρc_{OH} between the measurements and the predictions. Rows 1 to 3 from top to bottom show the normalized ρc_{CH2O}, the normalized ρc_{OH}, and the overlap of ρc_{CH2O} (green) and ρc_{OH} (red), respectively. Columns from left to right are the experiment, the 3D AMR simulation, the 2D DA+CFD with a coarse grid, and the 2D free CFD with the same coarse grid. The experimental flame thickness is similar to the prediction by the 3D AMR grid but thinner than

that by the 2D coarse grid. Although much less meaningful in this comparison using instantaneous quantities, an important point to make is that DA has demonstrated its powerful impact on the 2D coarse solution, pulling the flame dynamics toward the reference AMR solution, and one can expect further improvement in the DA+CFD flames by performing a 3D DA study. Nevertheless, an even more important point to make from this investigation is that DA can identify the deficiency when the computational model is insufficient in capturing the fundamental structure of the dynamics. For example, model errors may come from the physics models and their parameters employed by the governing PDEs (e.g., the SGS model and the approximation of turbulent burning speed), the reaction rate parameters, the thermodynamic parameters, the transport properties, and the computational configurations, such as the initial/boundary conditions and the accuracy of the computational geometry as compared to the real one.

Along the DA process, values of the cost function are evaluated, which measure the DA performance, i.e., uncertainty reduction rates. The uncertainty error is expected to be reduced after observations are assimilated at each DA cycle, and the value of the cost function after the process of the minimization decreases as well. For the present investigation, 50 DA cycles are performed, and the trajectories of the values of the cost function are compared between the forecast and the analysis. Figure 9.26 shows that at each assimilation cycle, the value of the analysis stage is always smaller than that of the forecast stage, demonstrating that DA is performed properly and the optimization algorithm is robust.

The performance is further assessed by measures of the error reduction of the total cost function and the L_2-norm of the ensemble skill. For each DA cycle, the error reduction of the total cost function can be calculated in form of $\gamma_{\text{tot}} = (J(\hat{\mathbf{Q}}^f) - J(\hat{\mathbf{Q}}^a))/(J(\hat{\mathbf{Q}}^f)) \times 100\%$, where the superscripts f and a represent the forecast state ($\hat{\mathbf{Q}}^f$) and the analysis state ($\hat{\mathbf{Q}}^a$), respectively. The formulation for computing the L_2-norm of the ensemble skill can be expressed as

$$\|S\|_{L_2} = \frac{1}{N_{\text{obs}}} \left\| \mathbf{O} - \mathcal{H}(\hat{\mathbf{Q}}) \right\| , \tag{9.3}$$

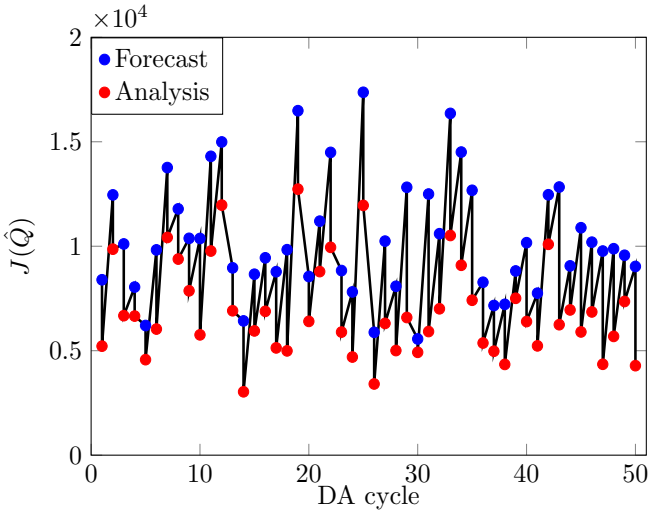


Figure 9.26: The trajectory of $J(\hat{\mathbf{Q}})$ along 50 DA cycles.

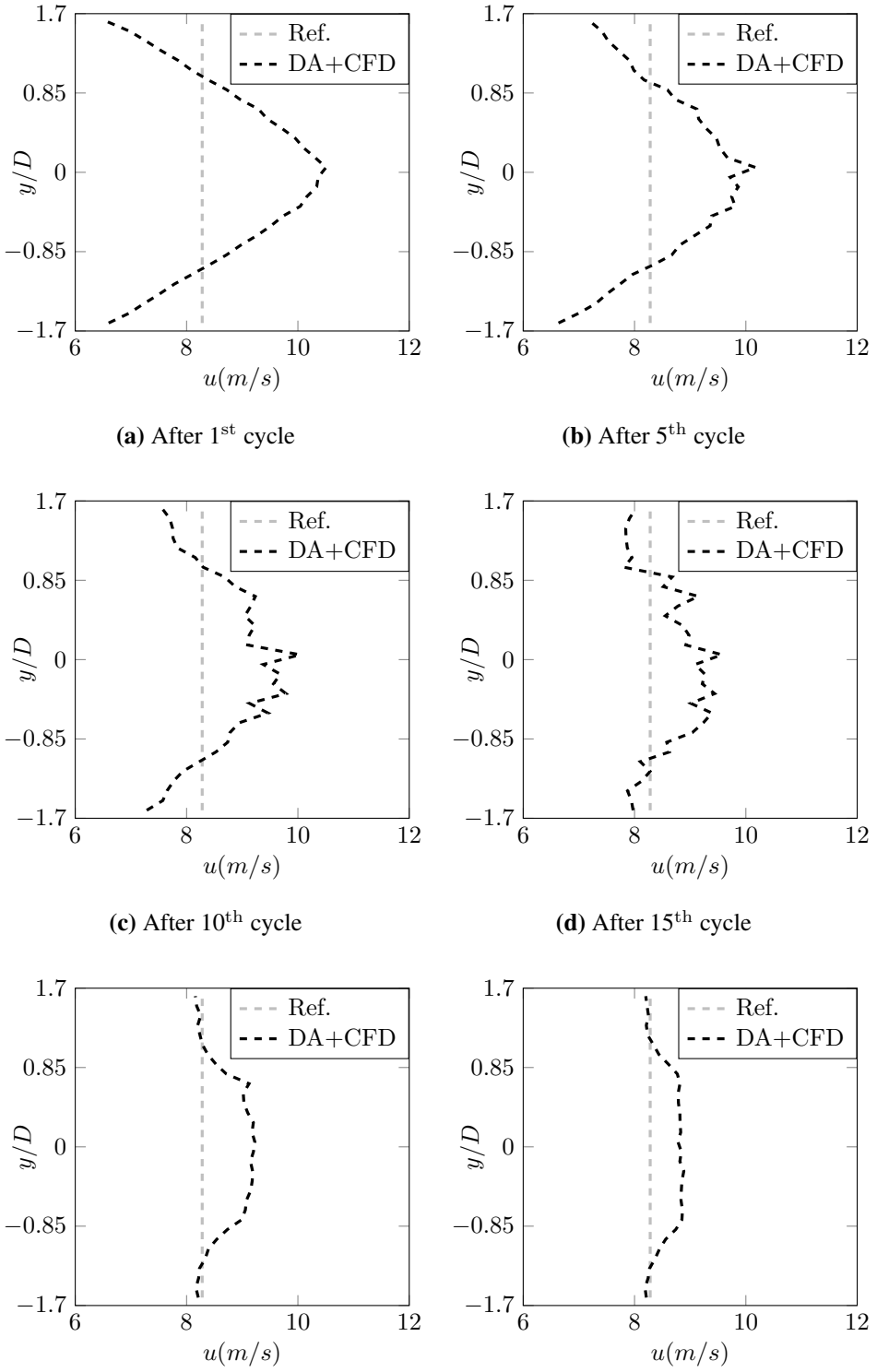
where N_{obs} is the total number of accepted observations by the quality control method. Skill evaluates the forecast or the analysis quality, measuring the overall level of agreement between the forecast/analysis and the observations. While the skill based on L_1 measures the absolute error of the deterministic vector, the L_2-based skill is used here to measure the square error of the deterministic vector. Table 9.1 shows the evaluations of $J(\hat{\mathbf{Q}})$, γ_{tot}, $\|S\|^f_{L_2}$ and $\|S\|^a_{L_2}$ along 50 DA cycles. As seen, for each observed quantity, the value of skill fluctuates along the DA process, in fact reflecting the chaotic nature of the dynamical system. However, the magnitudes of the L_2-based skill for u and ρc_{OH} at the analysis stage are reduced in comparison to those at the forecast stage and closer to their observations' errors at each DA cycle. In addition, the γ_{tot} values are between 0% and 100%. Larger values of γ_{tot} indicate a larger reduction of the cost function. Such properties signal promising results. Noticeably from Fig. 9.26 and Table 9.1, the last 8 DA cycles have consistently lower values for $J(\hat{\mathbf{Q}}^a)$. This may be an indication that assimilation of averaged data results in a smaller reduction in uncertainty than the use of instantaneous data, or that likely that the DA+CFD solution was indeed close to statistically steady since the reduction level does not fluctuate as much as that in the early DA cycles.

Table 9.1: The diagnosis results of $J(\hat{\mathbf{Q}}^f)$, $J(\hat{\mathbf{Q}}^a)$, $\|S\|^f_{L_2}$ and $\|S\|^a_{L_2}$ along 50 DA cycles.

DA cycle	$J(\hat{\mathbf{Q}}^f)$	$J(\hat{\mathbf{Q}}^a)$	γ_{tot} (%)	$\|S(u)\|^f_{L_2}$	$\|S(u)\|^a_{L_2}$	$\|S(\rho c_{\text{OH}})\|^f_{L_2}$ ($\times 10^{-5}$)	$\|S(\rho c_{\text{OH}})\|^a_{L_2}$ ($\times 10^{-5}$)
1	8394.5	3176.5	37.84	11.78	9.02	7.64	5.55
2	12457	9859.8	20.85	15.93	14.15	14.51	13.05
3	10109	6676.1	33.96	13.98	11.24	13.37	11.62
4	8046.3	6657.5	17.26	13.77	12.11	11.42	11.35
5	6205.9	4571.4	26.34	14.70	11.21	7.56	8.24
6	9824.9	6036.4	38.56	15.06	11.22	14.21	11.94
7	13764	10420	24.29	19.49	15.37	12.37	12.94
8	11787	9386	20.37	14.58	12.74	16.88	15.78
9	10378	7867.1	24.19	14.42	12.49	14.24	12.99
10	10371	5759	44.47	13.96	9.934	16.11	12.71
11	14303	9770.3	31.69	20.08	16.23	12.32	11.01
12	14991	11968	20.17	20.47	16.78	13.54	14.46
13	8964.6	6912	22.90	13.18	11.93	12.01	10.24
14	6425.1	3033.9	52.78	15.08	9.17	7.87	8.34
15	8660.1	5947.3	31.33	16.04	10.65	12.47	13.04
16	9445.1	6884	27.12	14.32	12.11	12.17	10.93
17	8779.9	5134.8	41.52	12.85	9.67	14.58	11.79
18	9833.5	4991.9	49.24	13.55	10.01	14.12	9.94
19	16487	12732	22.78	20.72	17.54	13.46	13.32
20	8550.5	6403.6	25.11	12.95	11.01	13.29	12.58
21	11200	8786.6	21.55	16.19	13.52	10.85	11.13
22	14491	9947.9	31.35	19.10	15.24	12.36	11.77
23	8835.3	5891.4	33.32	15.11	11.12	11.06	11.99
24	7813.3	4697.1	39.88	13.15	9.74	11.96	10.23
25	17368	11955	31.17	21.98	16.76	12.72	13.32
26	5880.5	3408.2	42.04	11.14	8.83	10.97	7.47
27	10244	6307	38.43	14.78	10.30	14.05	12.22
28	8083	5007.4	38.05	12.93	10.11	11.91	9.67
29	12821	6589.1	48.61	19.31	13.24	11.01	10.12
30	5561.5	4923	11.48	15.00	11.41	11.93	12.34
31	12498	5917.7	52.65	8.54	5.23	12.45	11.53
32	10599	7006.9	33.89	10.29	5.91	10.71	10.47
33	16357	10515	35.72	11.93	7.70	10.27	9.40
34	14507	9089.3	37.35	11.53	7.15	10.72	9.90
35	12677	7415.5	41.50	11.38	7.25	10.82	9.84
36	8276.8	5367.5	35.15	6.90	5.84	13.15	10.54
37	7162.7	4974.4	30.55	6.91	5.76	12.46	11.31
38	7218.7	4344.1	39.82	6.71	5.19	12.51	10.60
39	8816.1	7508.3	14.83	8.58	8.48	12.64	10.58
40	10169	6394.9	37.11	13.01	9.35	17.11	13.33
41	7753.2	5231.9	32.52	8.54	6.88	11.86	10.28
42	12463	10092	19.02	13.59	10.75	17.43	14.49
43	12833	6244.8	51.34	11.94	6.9	17.59	15.14
44	9059.5	6951.9	23.26	9.91	8.19	14.80	12.01
45	10885	5898.4	45.81	12.32	8.05	15.31	13.85
46	10194	6856	32.75	12.28	7.80	16.82	12.79
47	9769.5	4351	55.46	11.38	6.56	12.94	9.84
48	9880.5	5684.7	42.47	8.36	6.58	12.63	9.03
49	9567.1	7355.8	23.11	9.01	7.82	14.22	11.89
50	9031.9	4284.2	52.56	8.98	6.59	13.62	9.70

Errors in both Initial and Boundary Conditions

The performance of the MLEF method is further investigated for the situation in which the uncertainties exist in both the initial and inlet boundary conditions. Like the DA process as described in Chapter 5, at the analysis stage of each DA cycle, the predictions of the solution variables and u_{inlet} are updated simultaneously by the DA+CFD system with assimilating the observation data. Even though u_{inlet} is a non-observed variable in this study, the correlation relationship between the observed and unobserved variables is evaluated in the forecast error covariance matrix in the DA+CFD system, which is further used to transmit information in the changes of observed variables to unobserved variables during the DA correction process. After that, the updated solution variables and u_{inlet} are used to reinitialize the solution field and the boundary condition for the next forecast step in each member and the control case. In Fig. 9.27, the DA+CFD predictions of u_{inlet} profiles from the control case are plotted after the 1^{st}, 5^{th}, 10^{th}, 15^{th}, 20^{th}, and 30^{th} DA cycles. Clearly observed is that both the shape and estimation of u_{inlet} profiles are apparently getting closer and closer toward the reference data as more data assimilation cycles are performed. This phenomenon is encouraging because it demonstrates that the observation data effectively correct the erroneous boundary condition to a profile converging toward the reference along the DA process. However, by comparing the final DA+CFD profile and the reference profile, the DA+CFD prediction does not perfectly recover the reference data. The results are still acceptable since they show the DA impact on improving the boundary condition estimation over 30 DA cycles. Furthermore, tracking the difference between the two profiles along the DA process shows that the inlet velocity boundary condition is more responsive/sensitive to data in the regions near the top and bottom walls than in the center region. This is insightful in the sense that the DA+CFD system essentially tells us that some model characteristics at the inlet boundary may be most sensitive to perturbations in the assimilation process. Perhaps, measurements may be uninformative, or the information transfer process may be insufficient near the center region. A future follow-up study will be conducted to investigate and address the factors, such as what and when data should be assimilated into the system, to improve the center region.



(a) After 1st cycle

(b) After 5th cycle

(c) After 10th cycle

(d) After 15th cycle

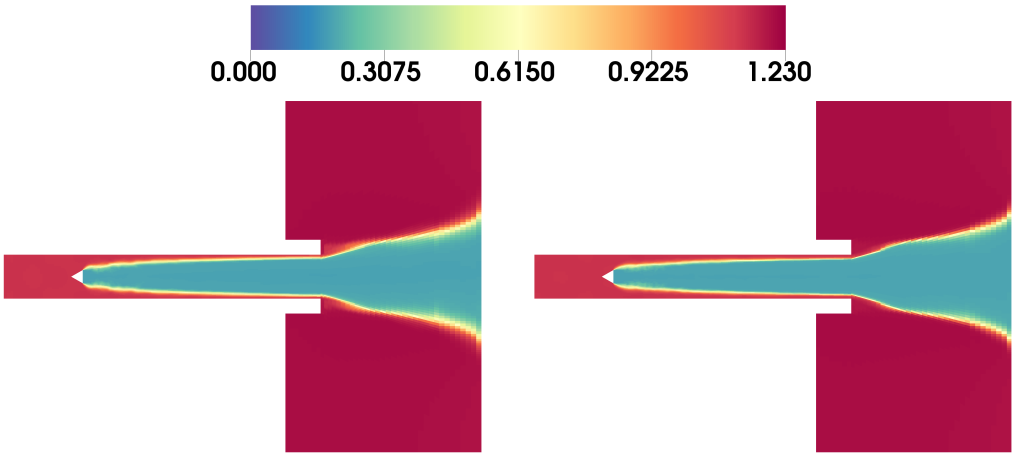
(e) After 20th cycle

(f) After 30th cycle

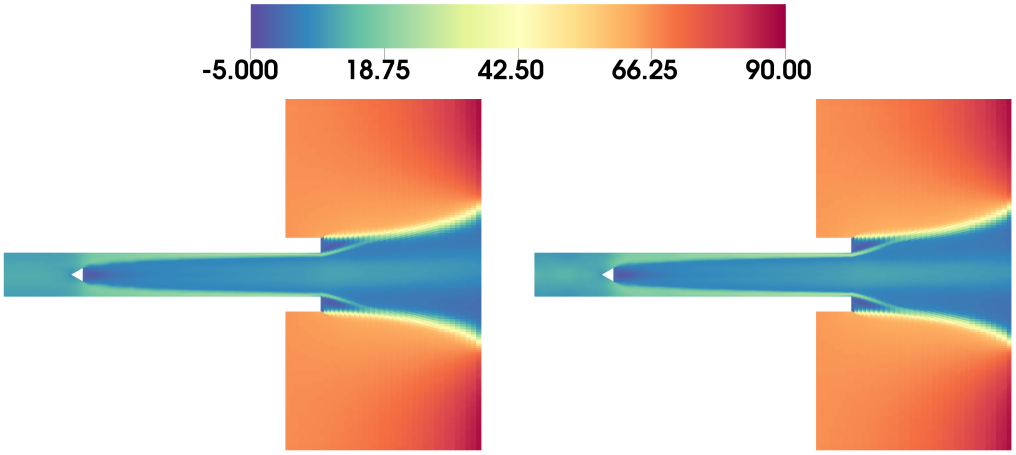
Figure 9.27: The analysis u_{inlet} profiles after the 1st, 5th, 10th, 15th, 20th, and 30th DA cycles.

The time-averaged solution contours, including density, axial momentum, specific total energy, OH mass, CH_2O mass, and temperature, over the entire 30 DA cycles are presented in Figs. 9.28 and 9.29. And, Fig. 9.30 shows the time-averaged solution contours in the viewing window, comparing the DA+CFD and the free CFD predictions and showing the difference between the two. The layout of each figure is the same as the previous DA case. By comparing the temperature contours in Fig. 9.29, we can see that the overall time-averaged flame shape in the DA+CFD prediction appears slightly bigger than the one in the free CFD prediction. All the other quantities, including density, axial momentum, specific total energy, and species mass fractions, behave consistently with the temperature. As desired, the DA+CFD flames still remain symmetric. In addition, the difference between the DA+CFD and CFD solutions can be seen more clearly in the regions of recirculation behind the bluff body and the shear layers. More fine structures are observed in the DA+CFD solution in Fig. 9.30. The DA+CFD flames are more wrinkled behind the bluff body and then fluffed up slightly in the shear layers, while the free CFD flames are just smooth in these regions.

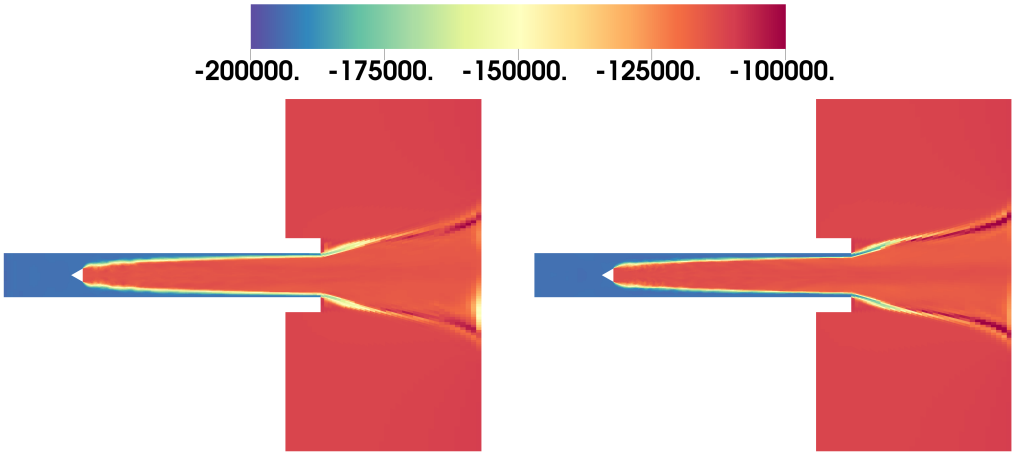
As shown in Figs. 9.31 and 9.32, the performance is further assessed by comparing the averaged y-profiles of the quantities, such as ρ, u, ρe, ρc_{OH}, ρc_{CH2O}, and temperature, between the DA+CFD and free CFD predictions at $x/D = 0.0394$ and $x/D = 2.62$, respectively. As a reminder, both locations have observed data for velocity and OH mass, which are assimilated to the model. At $x/D = 0.0394$, the positive impact by the data is only observed in the region of flame fronts. At $x/D = 2.62$, data appears to impact model predictions for all quantities significantly in the regions of both flame fronts and near the top and bottom walls. However, the DA impacts on those unobserved quantities in the center region are small, consistent with the impact seen in the inlet velocity profile estimation. Overall, the solution structures in the DA+CFD prediction indicate that the Chord-MLEF framework is effective and demonstrate that DA is performed properly.



(a) ρ.

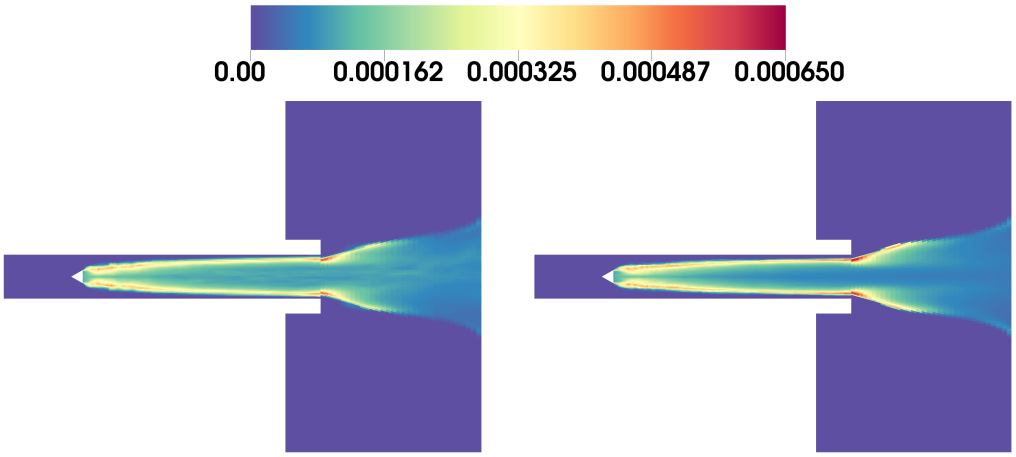


(b) ρu.

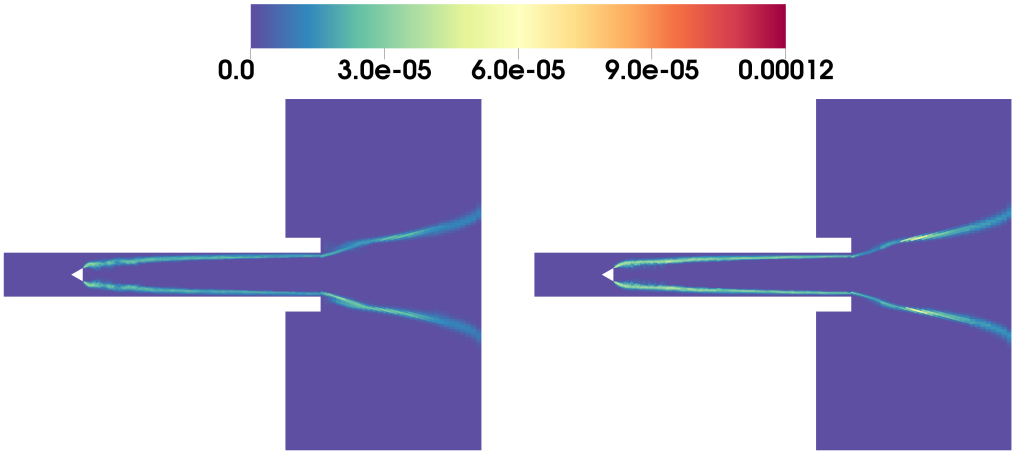


(c) ρe.

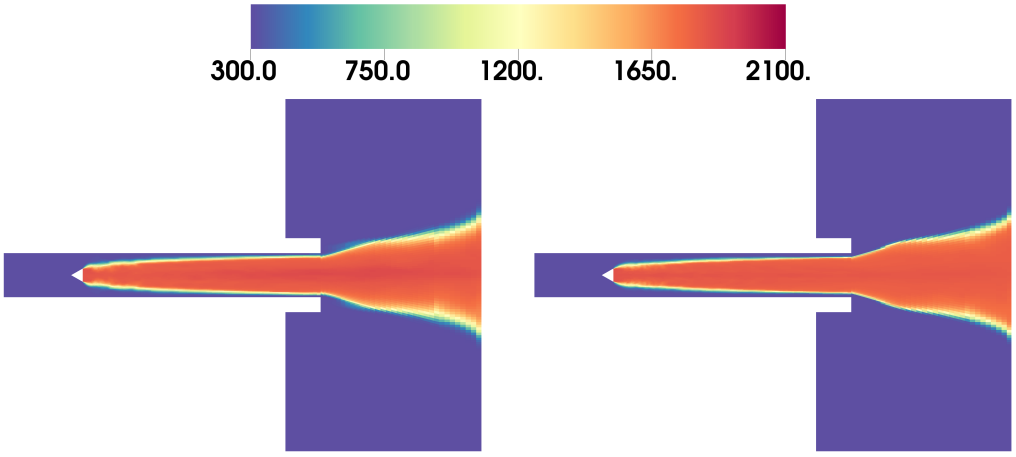
Figure 9.28: Time-averaged contours for ρ, ρu, and ρe over the time frame of 30 DA cycles. *left:* DA+CFD, *right:* CFD.



(a) ρc_{OH}.

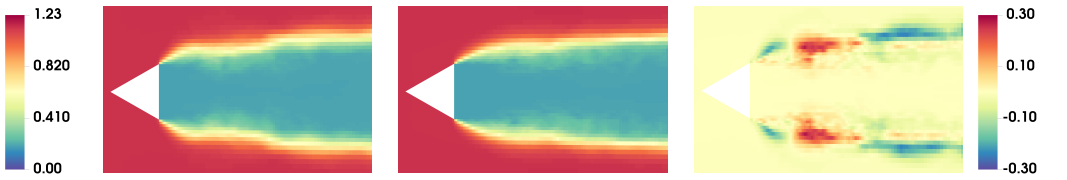


(b) $\rho c_{\mathrm{CH_2O}}$.

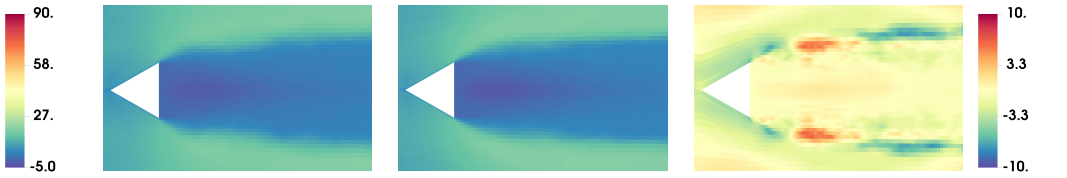


(c) Temperature.

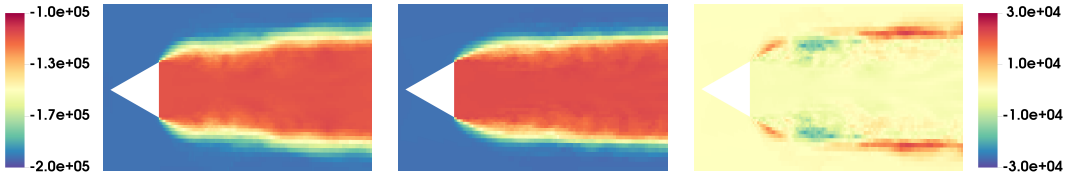
Figure 9.29: Time-averaged contours for ρc_{OH}, $\rho c_{\mathrm{CH_2O}}$ and temperature over the time frame of 30 DA cycles. *left:* DA+CFD, *right:* CFD.



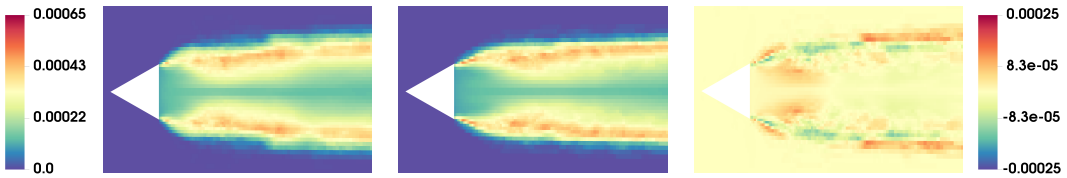
(a) ρ.



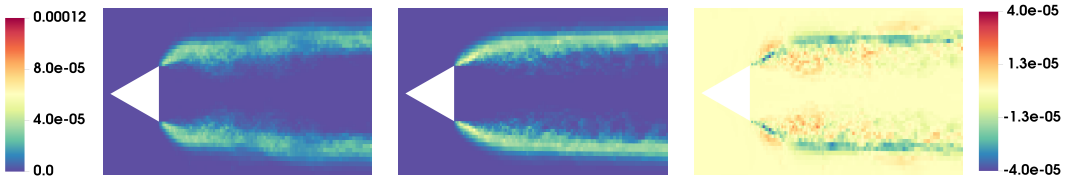
(b) ρu.



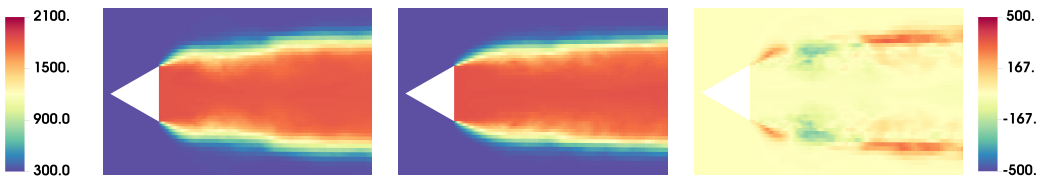
(c) ρe.



(d) ρc_{OH}.

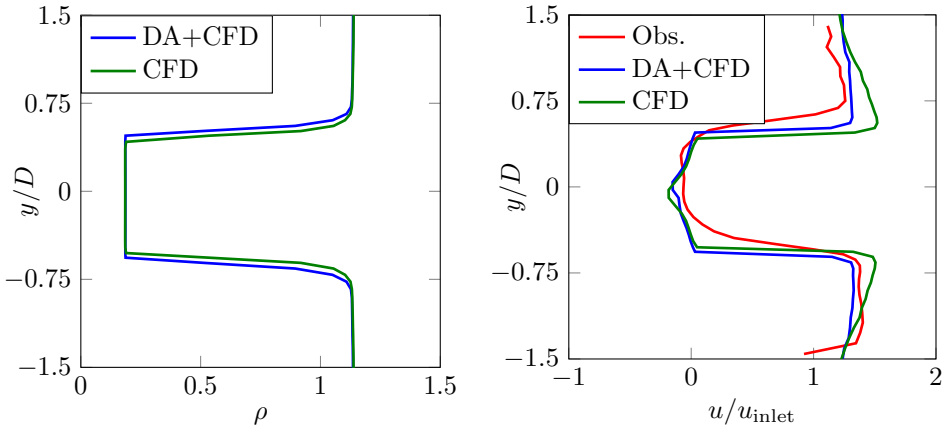


(e) $\rho c_{\mathrm{CH_2O}}$.



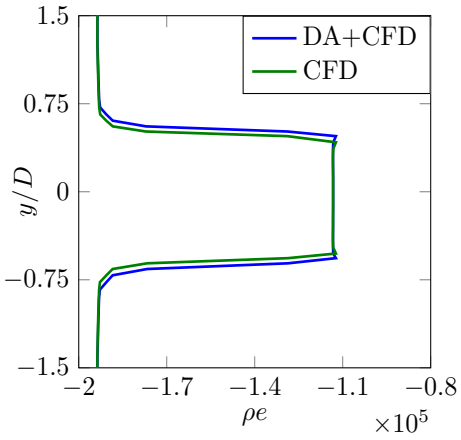
(f) Temperature.

Figure 9.30: Time-averaged contours for ρ, ρu, ρe, ρc_{OH}, $\rho c_{\mathrm{CH_2O}}$, and temperature in the viewing window over the time frame of 30 DA cycles. *left:* DA+CFD, *middle:* CFD, *right:* the difference.

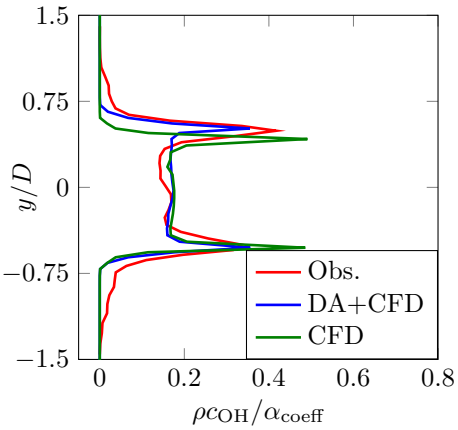


(a) ρ.

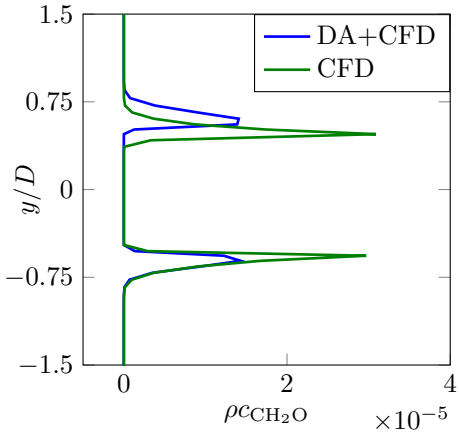
(b) u.



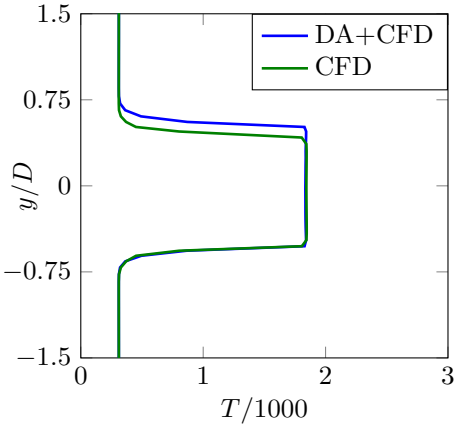
(c) ρe.



(d) ρc_{OH}.

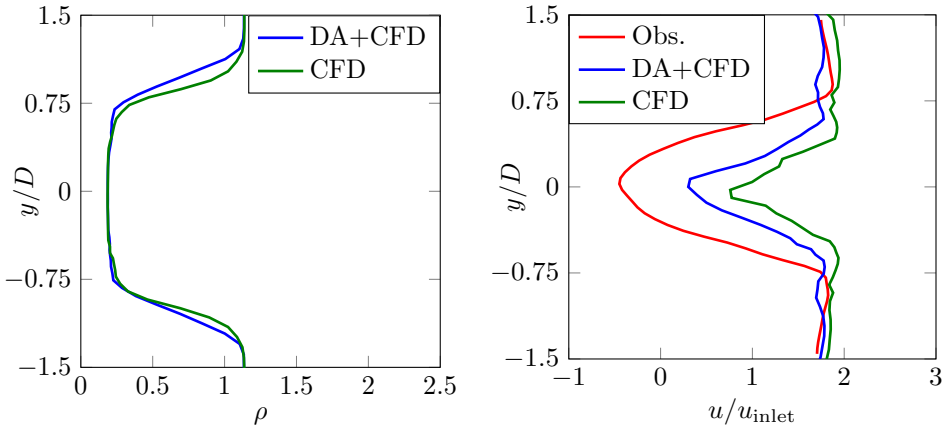


(e) $\rho c_{\mathrm{CH_2O}}$.



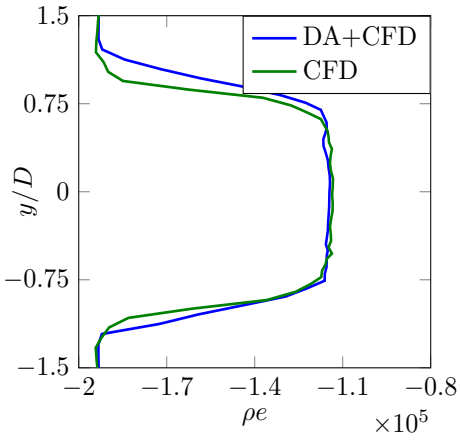
(f) Temperature.

Figure 9.31: Time-averaged 1D y-profiles comparisons at $x/H = 0.0394$ over the time frame of 30 DA cycles.

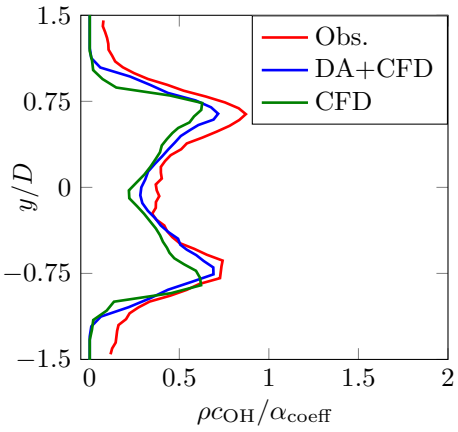


(a) ρ.

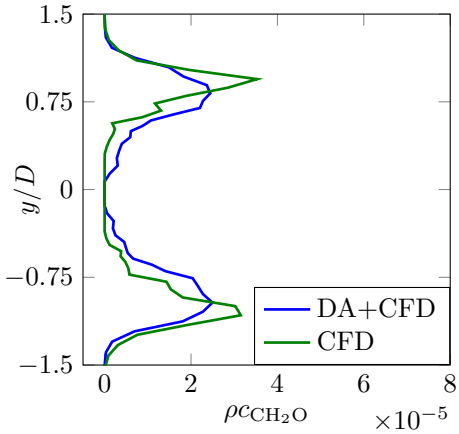
(b) u.



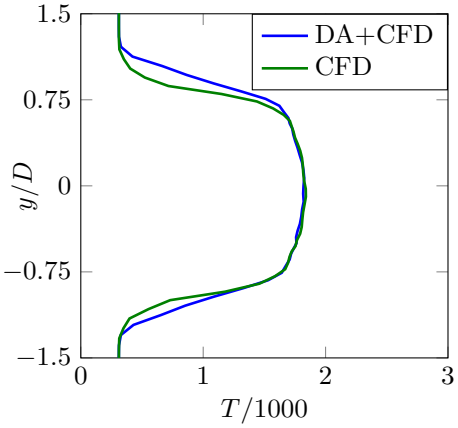
(c) ρe.



(d) ρc_{OH}.



(e) $\rho c_{\mathrm{CH_2O}}$.



(f) Temperature.

Figure 9.32: Time-averaged 1D y-profiles comparisons at $x/H = 2.62$ over the time frame of 30 DA cycles.

Chapter 10

Conclusions and Future Work

10.1 Conclusions

This dissertation details the development, verification, validation, and application of a multi-algorithm system that combines the CFD model and data through Bayesian data assimilation to increase the predictability of CFD modeling of turbulent and combusting engineering fluid flow problems. Specifically, we develop and apply a Bayesian computational framework by assimilating data into Chord using MLEF to improve the prediction of turbulent combustion flow in complex geometry and obtain a better understanding of the uncertainty of the dynamical system. Chord uses AMR, ARK_4, and MMB techniques altogether to solve practical problems with computational efficiency, enabling the ensemble calculation with reasonable CPU hours. MLEF is chosen as the main DA method, supported by a performance study using the CDR problem and L96 model for three ensemble-based DA methods, including MLEF, EnKF, and iterative EnKF. The assessment of DA performance demonstrates that the MLEF method is superior to the EnKF method in both the computational efficiency and the solution accuracy in the case of the linear observation operator and the IEnKF method in the case of the nonlinear observation operator. Especially in the cases of the nonlinear observation operator, the MLEF demonstrates consistent efficiency and effectiveness in improving the estimates of the model parameters and predictions. The turbulent Couette flow has demonstrated the validity of the DA+CFD system. The optimal states found by the DA+CFD system are significantly driven toward the true states and fast in the sense of the computational time, rendering the system a promising tool. By examining the vortex shedding frequency and the amplitude of the oscillations for the non-reacting flow in a bluff-body combustor, the optimal states are driven to the true states faster when assimilating more data from a 2D-plane than from a 1D-line. Notably, the vortex shedding frequency and the large-scale flow dynamics behind the bluff body are quickly corrected by assimilating more data. In addition, for those quantities without

observations being assimilated, the predictions are also improved through the coupling influence of solution variables and error covariances. This is deemed as a powerful impact of data assimilation on prediction because data is not always available for all variables, and deriving insights from available information on the non-observed quantities is an important aspect of data assimilation.

Moreover, the quality of the DA+CFD system is tested by the 2D turbulent reacting flow in a bluff-body combustor, which provides a reliable process of assimilating measurements from physical experiments. Since the measurements are often different from the computation data in type and resolution, the experimental data may not match the computation data type without post-processing, such as temperature or species PLIF images, and often, they are instantaneous. An observation operator must be developed to assimilate data properly into the model, and its accuracy impacts the assimilation performance and efficiency. We have developed proper nonlinear observation operators, applied observation-quality-control method, and measured performance through the present work. The results are promising, and what particularly merits a mention of the solution behavior is the consistent characteristics of model prediction by the DA+CFD system. This has two layers of meaning. First, the MLEF method demonstrates efficacy in damping any potential discontinuities that can arise during the DA process, particularly for a turbulent combustion system where radical species and thermodynamic states are sensitive to changes, even small ones. Secondly, the transfer of information from experimental data to the model appears to be smooth, indicating that the model characteristics in the spatial region, where data are available, and assimilation is performed, are stable and responsive to data in driving the dynamics. Furthermore, the investigation provides us a fundamental understanding of the scalability of the solution process developed from the low dimensions to the high dimensions, providing a quantitative basis for guiding the 3D DA studies on BBC and other DA application investigations. Similar positive DA impacts are observed in the results of the 3D DA studies on BBC. More fine structures appear in the DA+CFD solution immediately behind the flame holder, while the flame structures in the free CFD solution are much smoother in that region. The instantaneous DA+CFD flame is impacted by the observations to behave more wrinkled than the free CFD flame in the shear layer, which shows

a closer feature to the experimental flames. Furthermore, the DA+CFD system shows that some model characteristics may be most sensitive to perturbations in the assimilation process. From the inlet velocity profile comparison at the end of the DA process, as shown in Fig. 9.27, we remark that measurements from that spatial region could be less informative, and with more similar investigations using the DA+CFD system, recommendations on experimental measurements may be made, such as what quantities need or need not be measured, what regions contain sensitive dynamics, and how to better improve the information transmission from data to model for correcting the boundary condition.

10.2 Original Contributions

This dissertation makes novel contributions toward the development of the DA application for improving the CFD predictability of the turbulent combustion flows with realistic geometry, a high-order forward model, and experimental data. Some important original features of the research are as follows:

1. The development of a new alternative approach to improve the predictability of CFD modeling of turbulent combustion flows by reducing the uncertainty via data assimilation.

2. The derivation of new insight on factors, such as where, what, and when data should be assimilated, providing valuable guidance to data measurements and physical experimental design.

3. The demonstration of data assimilation as a potential powerful approach to improve CFD modeling of turbulent combustion in engineering applications and reduce the uncertainties with data.

10.3 Future Work

Through this research, we have investigated the data assimilation impact on the predictability of flow and flame dynamics structures. The future work will focus on

1. Extending the present DA+CFD system to present performance studies on the turbulent combustion applications of high Reynolds numbers.

2. Understanding the uncertainty in model parameters for developing and assessing data-driven models.

3. Demonstrating a new approach to address non-Gaussianity by integrating the MLEF with an implicit particle filtering method.

Bibliography

[1] N. Zettervall, K. Nordin-Bates, E. J. K. Nilsson, and C. Fureby. Large eddy simulation of a premixed bluff body stabilized flame using global and skeletal reaction mechanisms. *Combust. Flame*, 179:1–22, 2017.

[2] R. Xu, et al. A physics-based approach to modeling real-fuel combustion chemistry – II Reaction kinetic models of jet and rocket fuels. *Combust. Flame*, 193:520–537, July 2018.

[3] E. S. Titi. On approximate Inertial Manifolds to the Navier-Stokes equations. *J. Math. Anal. Appl.*, 149:540–557, 1990.

[4] C. Carton de Wiart and S. M. Murman. Extending the dynamic slip boundary condition to variational multi-scale methods. Technical report, Center for Turbulence Research, 2016. Proceedings of the Summer Program.

[5] P. Spalart and C. Rumsey. Effective inflow conditions for turbulence models in aerodynamic calculations. *AIAA J.*, 45(10), 2007.

[6] H. Kato and S. Obayashi. Approach for uncertainty of turbulence modeling based on data assimilation technique. *Comput. Fluids*, 85:2–7, 2012.

[7] A. J. Reagan, Y. Dubief, P. S. Dodds, and C. M. Danforth. Predicting flow reversals in a computational fluid dynamics simulated thermosyphon using data assimilation. *PLoS ONE*, 11(2), 2016.

[8] R. Diez. *Machine learning for RANS turbulence modelling of variable property flows*. PhD thesis, Delft University of Technology, 2018.

[9] K. Duraisamy, G. Iaccarino, and H. Xiao. Turbulence modeling in the age of data. *Annual Review of Fluid Mechanics*, 51:357–377, 2019.

[10] O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Journal of the Meteorological Society*, 113:1311–1328, 1987.

[11] O. Talagrand. Assimilation of observations-an introduction. *Journal of the Meteorological Society of Japan*, 75:191–209, 1997.

[12] E. Kalnay. *Atmospheric modeling, data assimilation and predictabilty*. Cambridge University Press, Cambridge, 1st edition, 2003.

[13] G. Evensen. *Data assimilation: the ensemble Kalman filter*. Springer, Heidelberg, 2007.

[14] P. J. VanLeeuwen. Particle filtering in geophysical system. *Mon. Wea. Rev.*, 2137:4089–4114, 2009.

[15] X. Wang. Application of the WRF hybrid ETKF–3DVAR data assimilation system for hurricane track forecasts. *Am. Meteorol. Soc.*, 26:868–884, December 2011.

[16] K. Suzuki and M. Zupanski. Uncertainty in solid precipitation and snow depth prediction for Siberia using the Noah and Noah-MP land surface models. *Frontiers of Earth Science*, 12(4):672–682, 2018. https://doi.org/10.1007/s11707-018-0691-2.

[17] J. Weinmeister, N. Xie, X. Gao, A. K. Prasad, and S. Roy. Combining a reduced polynomial chaos expansion approach with universal Kriging for uncertainty quantification. In *8th AIAA Theoretical Fluid Mechanics Conference*, number AIAA 2017-3481. AIAA Aviation Forum, 2017. https://doi.org/10.2514/6.2017-3481.

[18] J. Weinmeister, N. Xie, X. Gao, A. K. Prasad, and S. Roy. Analysis of a polynomial chaos-Kriging metamodel for uncertainty quantification in aerospace applications. In *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, number AIAA 2018-0911. AIAA SciTech Forum, 2018. https://doi.org/10.2514/6.2018-0911.

[19] J. Weinmeister, X. Gao, and S. Roy. Analysis of a polynomial chaos-Kriging metamodel for uncertainty quantification in aerodynamics. *AIAA Journal*, 57, 2019. https://doi.org/10.2514/1.J057527.

[20] X. Gao, Y. Wang, N. Overton, M. Zupanski, and X. Tu. Data-assimilated computational fluid dynamics modeling of convection-diffusion-reaction problems. *J. Comput. Sci*, 21:38–59, 2017.

[21] J. Ling and J. Templeton. Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty. *Phys. Fluids*, 27:085103, 2015.

[22] B. Tracey, K. Duraisamy, and J. Alonso. A machine learning strategy to assist turbulance model development. Technical report, 2015. 8th AIAA Aerospace Sciences Meeting.

[23] N. Kutz. Deep learning in fluid dynamics. *J. Fluid Mech.*, 814:1–4, 2017.

[24] T. Oliver and R. Moser. Bayesian uncertainty quantification applied to rans turbulence models. *Journal of PhysicsL Conference Series*, 318:042032, 2011.

[25] H Cheung, T. Oliver, E. Prudencio, S. Prudhomme, and R. Moser. Bayeisian uncertainty analysis with applications to turbulence modeling. *Reliab. Eng. Syst. Safety*, 96(9):1137–1149, 2011.

[26] W. Edeling, P. Cinnella, and R. Dwight. Predictive RANS simulations via Bayesian modelscenario averaging. *J. Comput. Phys.*, 275:65–91, 2014.

[27] C. H. Colburn, J. B. Cessna, and T. R. Bewley. State estimation in wall-bounded flow systems part 3. The ensemble Kalmam filter. *J. Fluid Mech.*, 682:289–303, 2011.

[28] T. Suzuki. Reduced-order Kalman-filtered hybrid simulation combining particle tracking velocimetry and direct numerical simulation. *J. Fluid Mech*, 709:249–288, 2012.

[29] A. Gronskis, D. Heitz, and E. Mémin. Inflow and initial conditions for direct numerical simulation based on adjoint data assimilation. *J. Comput. Phys.*, 242:480–497, 2013.

[30] H. Kato, A. Yoshizawa, G. Ueno, and S. Obayashi. A data assimilation methodology for reconstructing turbulent flows around aircraft. *J. Comput. Phys.*, 283:559–581, 2015.

[31] V. Mons, J.-C. Chassaing, T. Gomez, and P. Sagaut. Reconstruction of unsteady viscous flows using data assimilation schemes. *J. Comput. Phys.*, 316:255–280, 2016.

[32] J. Labahn, H. Wu, B. Coriton, J. Frnak, and M. Ihme. Data assimilation using high-speed meansurement and LES to examine local extinction events in turbulent flames. *Proc. Combust. Inst.*, 37:2259–2266, 2019.

[33] H. Xiao, J.-L. Wu, J.-X. Wang, R. Sun, and C. J. Roy. Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach. *J. Comput. Phys.*, 324:115–136, 2016.

[34] C. He, Y. Liu, and L. Gan. A data assimilation model for turbulent flows using continuous adjoint formulation. *Phys. Fluids*, 30:105108, 2018.

[35] M. Zupanski, I. M. Navon, and D Zupanski. The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *Q. J. R. Meteorol. Soc.*, 113(D20110), 2008.

[36] X. Gao, S. M. J. Guzik, and P. Colella. A fourth-order boundary treatment for viscous fluxes on cartesian grid finite-volume methods. AIAA 2014-1277, 52nd AIAA Aerospace Sciences Meeting, 2014.

[37] S. M. Guzik, X. Gao, L. D. Owen, P. McCorquodale, and P. Colella. A freestream-preserving fourth-order finite-volume method in mapped coordinates with adaptive-mesh refinement. *Comput. Fluids*, 123:202–217, 2015.

[38] S. M. Guzik, X. Gao, and C. Olschanowsky. A high-performance finite-volume algorithm for solving partial differential equations governing compressible viscous flows on structured grids. *Comput. Math Appl.*, 72:2098–2118, 2016.

[39] X. Gao, L. D. Owen, and S. M. J. Guzik. A parallel adaptive numerical method with generalized curvilinear coordinate transformation for compressible Navier-Stokes equations. *Int. J. Numer. Meth. Fluids*, 82:664–688, 2016.

[40] L. D. Owen, S. M. Guzik, and X. Gao. A high-order adaptive algorithm for multispecies gaseous flows on mapped domains. *Comput. Fluids*, 170:249–260, 2018.

[41] L. D. Owen, X. Gao, and S. M. Guzik. Techniques for improving monotonicity in a fourth-order finite-volume algorithm solving shocks and detonations. *J. Comput. Phys.*, 415, 2020.

[42] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A*, 38A:97–110, March 1986.

[43] R. N. Bannister. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.*, 143:607–633, 2017.

[44] G. Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn*, 53:343–367, 2003.

[45] P. L. Houtekamer and H. L. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Am. Meteorol. Soc.*, 129:123–137, 2001.

[46] C. Bishop, J. Etherton, and S. J. Majmudar. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspect. *Mon. Wea. Rev.*, 129:420–430, 2001.

[47] J. S. Whitaker and T. M. Hamill. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, 130:1913–1924, 2002.

[48] J. L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, 129:2884–2903, 2001.

[49] M. Zupanski. Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.*, 133:1710–1726, 2005.

[50] I. M. Navon and S. J. Wright. Conjugate gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.*, 115:1479–1502, 1987.

[51] D. F. Parrish and J. C. Derber. The national meteorological center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, 120:1747–1763, 1992.

[52] M. Zupanski. Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.*, 121:2396–2408, 1993.

[53] Z. Li and I. M. Navon. Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. R. Meteorol. Soc.*, 127:661–683, 2001.

[54] F. Rabier, H. Jarvinen, E. Klinker, and A. J. Simmons. The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, 126:1143–1170, 2007.

[55] X. Wang, D. M. Barker, C. Snyder, and T. M. Hamill. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observation system simulation experiment. *Mon. Wea. Rev.*, 136:5116–5131, 2008.

[56] D. T. Kleist and K. Ide. An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3d-hybrid results. *Mon. Wea. Rev.*, 143:433–451, 2015.

[57] B. M. Bell. The iterated Kalman smoother as a Gauss-Newton method. *SIAM J. Optim.*, 4:626–636, 1994.

[58] R. J. Lorentzen and G. Naevdal. An iterative ensemble Kalman filter. *IEEE Transactions on Automatic Control*, 56(8):1990–1995, 2011.

[59] S. Yang, E. Kalnay, and B. Hunt. Handling nonlinearity in an Ensemble Kalman Filter: experiments with the three-variable Lorenz model. *Mon. Wea. Rev.*, 140:2628–2646, 2012.

[60] P. Sakov, D. S. Oliver, and L. Bertino. An iterative enkf for strongly nonlinear systems. *Mon. Wea. Rev.*, 140:1988–2004, 2012.

[61] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[62] P. Sakov, J.-M. Haussaire, and M. Bocquet. An iterative ensemble Kalman filter in presence of additive model error. *Q. J. R. Meteorol Soc*, 144:1297–1309, 2018.

[63] X. Gao and C. P. T. Groth. A parallel adaptive mesh refinement algorithm for predicting turbulent non-premixed combusting flows. *Int. J. Comut. Fluid Dyn.*, 20(5):349–357, 2006.

[64] X. Gao. *A parallel solution-adaptive method for turbulent non-premixed combusting flows*. PhD thesis, University of Toronto, 2008.

[65] X. Gao and C. P. T. Groth. A parallel solution-adaptive method for three-dimensional turbulent non-premixed combusting flows. *J. Comput. Phys.*, 229:3250–3275, 2010.

[66] X. Gao, S. Northrup, and C.P.T. Groth. Parallel solution-adaptive method for two-dimensional non-premixed combusting flows. *Progress in Computational Fluid Dynamics, An International Journal*, 11(2):76–95, 2011.

[67] L. D. Owen, S. M. Guzik, and X. Gao. A fourth-order finite-volume algorithm for compressible flow with chemical reactions on mapped grids. (AIAA 2017-4498), 2017.

[68] J. D. Anderson. *Hypersonic and High Temperature Gas Dynamics*. McGraw-Hill, New York, 1989.

[69] S. Arrhenius. Über die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch säuren. *Z. Physik. Chem.*, 4:226, 1889.

[70] N. Overton-Katz, X. Gao, and S. M. Guzik. Mapped multi-block grids of arbitrary geometry for high-order finite-volume methods with adaptive mesh refinement. In *Comput. Math Appl.*, 2021. Manuscript in progress for submission.

[71] Z. J. Wang, K. Fidkowski, R. Abgrall, F. Bassi, D. Caraeni, A. Cary, H. Deconinck, R. Hartmann, K. Hillewaert, H. T. Huynh, N. Kroll, G. May, P.-O. Persson, B. van Leer, and M. Visbal. High-order CFD methods: current status and perspective. *Int. J. Numer. Methods Fluids*, 72(8):811–845, 2013.

[72] L. D. Owen. *A Fourth-Order Solution-Adaptive Finite-Volume Algorithm for Compressible Reacting Flows on Mapped Domains*. Ph.D. dissertation, Colorado State University, 2019.

[73] C. Kennedy and M. Carpenter. Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Applied Numerical Mathematics*, 44:139–181, 2003.

[74] J. C. Christopher. *Time integration for complex fluid dynamics*. Ph.D. dissertation, Colorado State University, 2021.

[75] J. C. Christopher, S. M. Guzik, and X. Gao. High-order implicit-explicit additive Runge-Kutta schemes for numerical combustion with adaptive mesh refinement. *Int. J. Numer. Methods Fluids*, 2022. https://doi.org/10.1002/fld.5084.

[76] G. Bugers, P. J. VanLeeuwen, and G Evensen. Analysis scheme in the ensemble kalman filter. *Mon. Wea. Rev.*, 126:1719–1724, 1998.

[77] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

[78] A. C. Lorenz. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, 112:1177–1194, 1986.

[79] B. Wu, X. Zhao, B. Roychowdhury, B. M. Cetegen, C. Xu, and T. Lu. A numerical investigation of the flame structure and blowoff characteristics of a bluff-body stabilized turbulent premixed flame. *Combust. Flame*, 202:376–393, 2019.

[80] S. J. Fletcher. *Data Assimilation for the Geosciences: From Theory to Application*. Elsevier Inc., 2017.

[81] S. R. Allmaras, F. T. Johnson, and P. R. Spalart. Modification and clarifications for the implementation of the spalart-allmaras turbulence model. *Seventh International Conference on Computational Fluid Dynamics*, ICCFD7-1902, July 2012.

[82] P. R. Spalart and S. R. Allmaras. A one-equation turbulence model for aerodynamic flows. *Recherche Aerospatiale*, 1:5–21, 1994.

[83] P. R. Spalart, A. Garbaruk, and M. Strelets. Rans solutions in couette flow with streamwise vortices. *International Journal of Heat and Fluid Flow*, 49:128–134, 2014.

[84] S. A. Ragab and J. L. Wu. Linear instabilities in two-dimensional compressible mixing layers. *Phys. Fluids A*, 1(6):957 – 966, 1989.

[85] A. Comer, M. Ihme, C. Li, S. Menon, J. Oefeline, B. Rankin, V. Sankaran, and V. Sankaran. Model validation for propulsion workshop. Accessed online, January 2018. https://community.apan.org/wg/afrlcg/mvpws/p/proceedings.

[86] P. A. T. Cocks, M. Soteriou, and V. Sankaran. Impact of numerics on the predictive capabilities of reacting flow les. *Combust. Flame*, 162:3394–3411, 2015.

[87] V. Sankaran and T. P. Gallagher. Grid convergence in LES of bluff body stabilized flames. *2017 AIAA SciTech Forum*, pages 10.2514/6.2017–1791, 2017.

[88] T. P. Gallagher and V. Sankaran. Affordable explicitly filtered large-eddy simulation for reacting flows. *AIAA JOURNAL*, 57(2):10.2514/1.J057213, 2019.

[89] V. Sankaran and T. P. Gallagher. A consistent reactive LES based on explicit filtering. *2019 AIAA SciTech Forum*, pages 10.2514/6.2019–0451, 2019.

[90] A. W. Caswell, B. A. Rankin, B. C. Huelskamp, A. C. Lynch, V. Belovich, and J. R. Gord. Spatiotemporal characterization of bluff-body stabilized turbulent premixed flames using simultaneous high-repetition-rate OH PLIF and PIV measurements. *53rd AIAA Aerospace Sciences Meeting*, pages 2015–0424, 2015.

[91] David G. Goodwin, Raymond L. Speth, Harry K. Moffat, and Bryan W. Weber. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. https://www.cantera.org, 2021. Version 2.5.1.

[92] C. Fureby. A comparative study of large eddy simulation (LES) combustion models applied to the volvo validation rig. *AIAA SciTech Forum*, pages 10.2514/6.2017–1575, 2017.

[93] B. Rochette, F. Collin-Bastiani, L. Gicquel, O. Vermorel, D. Veynante, and T. Poinsot. Influence of chemical schemes, numerical method and dynamic turbulent combustion modeling on les of premixed turbulent flames. *Combust. Flame*, 191:417–430, 2018.

[94] C. Fureby. A large eddy simulation (LES) study of the VOLVO and AFRL bluff body combustors at different operating conditions. *2019 AIAA SciTech Forum*, pages 10.2514/6.2019–0453, 2019.

[95] C. Fureby. Large eddy simulation of turbulent reacting flows with conjugate heat transfer and radiative heat transfer. *Proc. Combust. Inst.*, 38(2):3021–3029, 2021.

Appendix A

Reaction Mechanisms

Table A.1: Propane-air "Z66" mechanism [1]

Num.	Reaction	A_r	β_r	$E_{a,r}$ [cal mol^{-1}]
1	C3H8 => CH3 + C2H5	1.7E+16	0	54840
2	C3H8 + H => 3 CH3	1E+09	0	3000
3	C2H6 + H => C2H5 + H2	537	3.5	5200
4	CH3 + CH3 => C2H6	1E12	0	5000
5	C2H6 => CH3 + CH3	2.239E19	-1	88310
6	C2H5 + M => C2H4 + H + M	2E+15	0	39000
7	C2H4 + H + M => C2H5 + M	4.17E+10	0	11030
8	C2H5 + H => CH3 + CH3	3.16E+13	0	0
9	C2H4 + O => HCO + CH3	3.31E+12	0	1130
10	HCO + CH3 => C2H4 + O	1.58E+11	0	31180
11	C2H4 + OH => C2H3 + H2O	4.79E+12	0	1230
12	C2H3 + H2O => C2H4 + OH	1.2E+12	0	14000
13	C2H4 + CH3 => C2H3 + CH4	1E+13	0	13000
14	C2H3 + CH4 => C2H4 + CH3	3.02E+13	0	12580
15	C2H2 + H + M => C2H3 + M	1.23E+11	1	10360
16	C2H3 + H => C2H2 + H2	2E+13	0	2500
17	C2H + H + M => C2H2 + M	1.1E+09	1	770
18	C2H2 + H => C2H + H2	2E+14	0	19000
19	C2H2 + OH => C2H + H2O	6.03E+12	0	7000
20	C2H + H2O => C2H2 + OH	5.37E+12	0	16360
21	C2H2 + O => C2H + OH	3.24E+15	0.6	17000
22	C2H + OH => C2H2 + O	2.95E+14	0.6	910
23	C2H + O2 => HCO + CO	1E+13	0	7000
24	HCO + CO => C2H + O2	8.51E+12	0	138400
25	H + O2 => OH + O	2.65E+14	0	16800
26	OH + O => H + O2	1.5E+13	0	690
27	O + H2 => OH + H	1.80E+10	1	8826
28	OH + H => O + H2	8E+09	1	6760
29	H2 + OH => H2O + H	1.17E+09	1.3	3626
30	H2O + H => H2 + OH	6.0E+09	1.3	18200
31	OH + OH => O + H2O	6.0E+08	1.3	0
32	O + H2O => OH + OH	5.0E+09	1.3	15000
33	H + O2 + M => HO2 + M	1.1E+18	-0.8	0
	CH4:6.5/CO:0.75/CO2:1.5/H2:1/H2O:6.5/N2:0.4/O2:0.4			

Num.	Reaction	A_r	β_r	$E_{a,r}$ [cal mol^{-1}]
34	H + HO2 => OH + OH	1.5E+14	0	1004
35	H + HO2 => H2 + O2	2.5E+13	0	700
36	OH + HO2 => H2O + O2	2.0E+13	0	1000
37	CO + OH => CO2 + H	1.51E+07	1.3	-758
38	CO2 + H => CO + OH	1.57E+09	1.3	18200
39	CH4 (+ M) => CH3 + H (+ M)	6.3E+14	0	104000
	C3H8:3/CH4:6.5/CO:0.75/CO2:1.5/H2:1/H2O:6.5/N2:0.4/O2:0.4			
	Low/ 1.0E+17, 0, 86000			
40	CH3 + H (+ M) => CH4 (+ M)	5.20E+12	0	-1310
	C3H8:3/CH4:6.5/CO:0.75/CO2:1.5/H2:1/H2O:6.5/N2:0.4/O2:0.4			
	Low/ 8.25E+14, 0, -19310			
41	CH4 + H => CH3 + H2	2.20E+04	3	8750
42	CH3 + H2 => CH4 + H	9.57E+02	3	8750
43	CH4 + OH => CH3 + H2O	1.6E+06	2.1	2460
44	CH3 + H2O => CH4 + OH	3.02E+05	2.1	17422
45	CH3 + O => CH2O + H	6.80E+13	0	0
46	CH2O + H => HCO + H2	9.0E+13	0	3991
47	CH2O + OH => HCO + H2O	6.0E+13	0	1050
48	HCO + H => CO + H2	4.0E+13	0	0
49	HCO + M => CO + H + M	1.60E+14	0	14700
50	CH3 + O2 => CH3O + O	5.0E+12	0	25652
51	CH3O + H => CH2O + H2	2.00E+13	0	0
52	CH3O + M => CH2O + H + M	2.40E+13	0	28812
53	HO2 + HO2 => H2O2 + O2	8.0E+12	0	0
54	H2O2 + M => OH + OH + M	1.30E+17	0	45500
55	OH + OH + M => H2O2 + M	9.86E+14	0	-5070
56	H2O2 + OH => H2O + HO2	1.00E+13	0	1800
57	H2O + HO2 => H2O2 + OH	2.86E+13	0	32790
58	OH + H + M => H2O + M	2.20E+22	-2	0
59	H + H + M => H2 + M	1.80E+18	-1	0
60	CH3 + OH => CH2 + H2O	7.6E+06	2.0	5000
61	CH2 + O => CO + H2	3.0E+13	0	0
62	CH2 + OH => CH + H2O	1.13E+07	2.0	3000
63	CH + O => CO + H	5.7E+13	0	0
64	CH + OH => HCO + H	3.0E+13	0	0
65	CH + O2 => HCO + O	3.3E+13	0	0
66	CH + CO2 => HCO + CO	2.5E+14	0	50

Table A.2: Mechane-air mechanism [2]

Num.	Reaction	A_r	β_r	$E_{a,r}$ [cal mol^{-1}]
1	H + O2 <=> O + OH	1.09E+14	0.	15310
2	O + H2 <=> H + OH	3.82E+12	0.	7950
3	O + H2 <=> H + OH	8.79E+14	0.	19180
4	OH + H2 <=> H + H2O	2.16E+08	1.51	3437
5	OH + OH <=> O+H2O	3.35E+04	2.42	-1928
6	H2 + M <=> 2H + M	4.58E+19	-1.4	104390
	H2:2.5/CO:1.9/CO2:3.8/H2O:12/CH4:2/CH2O:2.5			
7	O + H + M <=> OH + M	4.71E+18	-1.0	0
	H2:2.5/CO:1.9/CO2:3.8/H2O:12/CH4:2./CH2O:2.5			
8	H2O + M <=> H + OH + M	6.06E+27	-3.322	120800
	O2:1.5/H2:3/CO:1.9/CO2:3.8/H2O:0/CH4:7./CH2O:2.5			
9	H2O + H2O <=> H + OH + H2O	1.01E+26	-2.44	120200
10	H + O2 (+M) <=> HO2 (+M)	4.65E+12	0.44	0
	O2:0.78/H2:2/CO:1.9/CO2:3.8/H2O:14/CH4:2/CH2O:2.5			
	Low/ 1.91E+21, -1.72, 525.			
	Troe/ 0.5, 30, 90000, 90000.			
11	HO2 + H <=> H2 + O2	3.68E+06	2.087	-1455
12	HO2 + H <=> OH + OH	7.08E+13	0	300
13	HO2 + H <=> O + H2O	1.45E+12	0	0
14	HO2 + O <=> OH + O2	1.63E+13	0	-445
15	HO2 + OH <=> H2O + O2	7.00E+12	0	-1093
16	HO2 + OH <=> H2O + O2	4.50E+14	0	10930
17	CO + O (+M) <=> CO2 (+M)	1.06E+13	-0.308	6943
	H2:2.5/CO:1.9/CO2:3.8/H2O:12/CH4:2/CH2O:2.5			
	Low/ 1.40E+21, -2.1, 5500			
18	CO + O2 <=> O + CO2	2.53E+12	0	47700
19	CO + OH <=> H + CO2	8.46E+04	2.053	-356
20	CO + OH <=> H + CO2	8.64E+12	-0.664	332
21	CO + HO2 <=> OH + CO2	1.57E+05	2.18	17944
22	CH4 + H <=> CH3 + H2	3.07E+06	2.5	7588
23	CH4 + O <=> OH + CH3	2.31E+08	1.56	8485
24	CH4 + OH <=> CH3 + H2O	1.00E+06	2.182	2446
25	CH3 + H (+M) <=> CH4 (+M)	1.41E+14	0	0
	H2:2/CO:1.5/CO2:2/H2O:6/CH4:2./CH2O:2.5			
	Low/6.35E+35, -5.57, 3818			
	Troe/0.37, 3315, 61, 90000			

Num.	Reaction	A_r	β_r	$E_{a,r}$ [cal mol^{-1}]
26	CH3 + O <=> H + CH2O	1.08E+14	0	0
27	CH3 + O => H + H2 + CO	2.31E+13	0	0
28	CH3 + HO2 <= O2 + CH4	1.16E+05	2.350	-1522
29	CH3 + HO2 => OH + CH2O + H	2.08E+13	0	-590
30	CH3 + O2 => O + CH2O + H	2.51E+12	0	28297
31	CH3 + O2 <=> OH + CH2O	2.28E+01	2.53	9768
32	CH3 + CH2O => H + CO + CH4	1.06E+01	3.36	4310
33	CH2O (+M) <=> H2 + CO (+M)	3.70E+13	0	71976
	H2:2/CO:1.5/CO2:2/H2O:6/CH4:2./CH2O:2.5			
	Low/4.40E+38, -6.1, 94000			
	Troe/0.932, 197, 1540, 10300			
34	CH2O + H => H2 + CO + H	5.67E+12	0.361	4609
35	CH2O + H => H2 + CO + H	1.14E+13	0.582	14395
36	CH2O + O => OH + H + CO	4.16E+11	0.57	2762
37	CH2O + OH => H + CO + H2O	7.82E+07	1.63	-1055
38	CH2O + O2 => HO2 + H + CO	2.44E+05	2.5	36460