Approximate Bayesian Computation for Parameter Estimation in Complex Thermal-Fluid Systems

by

Jason D. Christopher

B.S. United States Air Force Academy, 2007

M.S. Rice University, 2009

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Mechanical Engineering

2018

This thesis entitled:
Approximate Bayesian Computation for Parameter Estimation in Complex Thermal-Fluid Systems
written by Jason D. Christopher
has been approved for the Department of Mechanical Engineering

Prof. Peter E. Hamlington

Prof. Gregory B. Rieker

Prof. Ian Grooms

Prof. Daven K. Henze

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Christopher, Jason D. (Ph.D., Mechanical Engineering)

Approximate Bayesian Computation for Parameter Estimation in Complex Thermal-Fluid Systems

Thesis directed by Professors Peter E. Hamlington and Gregory B. Rieker

A major challenge in computational fluid dynamics (CFD) simulations of real-world flows is the accurate assignment of boundary, initial, and geometric conditions, as well as fluid and material properties. Despite advances in experimental techniques, however, acquiring the information necessary to simultaneously set each of these conditions and properties remains a considerable challenge. As a potential solution to this difficulty, recent advances in data-driven parameter estimation techniques have provided flexible and increasingly sophisticated methods for improving the fidelity of simulation configurations using experimental data. This dissertation applies, for the first time, a technique called approximate Bayesian computation (ABC) to complex thermal-fluid flows in order to determine numerical simulation parameters from experimental or other reference data.

In this dissertation, the ABC approach is demonstrated for several engineering test cases to demonstrate its efficacy at determining unknown parameters in a wide variety of settings. As a simple initial case, the logistics equation is used to demonstrate the technique. This is followed by the case of a two-dimensional turbulent buoyant jet with variable inlet velocity. The jet is modeled using a large eddy simulation (LES), and reference data is obtained from ensembles of both LES (serving as a benchmark for the technique) and direct numerical simulation (DNS) cases. The reference parameters are correctly identified based on either velocity or temperature measurements at various heights. Using a similar setup, but now with a lightly forced helium plume, the puffing frequency of the jet is identified and used to match experimental observations to predict inlet composition.

Moving on to industrial engineering applications, a three-dimensional turbulent buoyant jet with unknown temperature conditions is simulated using a Reynolds-averaged Navier Stokes (RANS) simulation. In this application, reference observations come from the same RANS case

with known parameters. The ABC procedure correctly identifies the inlet temperature boundary conditions. In the next case, a rotating cylinder above a high-temperature turbulent buoyant jet is investigated. Here the reference observations come from a two-dimensional RANS simulation. In particular, the initial reference case has known jet inflow and cylinder rotational velocities, and the ABC approach is shown to correctly identify the reference values of these parameters using sparse temperature statistics within the domain. In an additional test using the two-dimensional rotating cylinder case, the reference case has known species concentrations at the jet inflow and we show that ABC can correctly identify the reference concentrations using sparse species and temperature measurements within the domain.

The ultimate application of ABC in this dissertation uses the technique to determine three-dimensional LES parameters based on comparisons with experimental observations. The experimental temperature data are obtained above an industrially-relevant catalytic burner using laser absorption spectroscopy. This final application identifies parameters that are not able to be measured experimentally, including inlet velocity and heat addition due to continued combustion within the flow field.

These successes indicate that ABC can be extended to additional real-world engineering systems, even when only sparse observational data is available. Using ABC and reference data, one can accurately drive the selection of boundary conditions, as well as model parameters, in numerical simulations. Furthermore, ABC can provide insights into quantities that are not easily measured experimentally.

Acknowledgements

Helpful discussions with Profs. Ian Grooms, John Daily, Daven Henze and Melvyn Branch, as well as researchers Mark Strobel and Aniruddha Upadhye, are gratefully acknowledged.

This work utilized the Janus and Summit supercomputers, which are supported by the National Science Foundation (award number CNS-0821794), the CU Boulder, CU Denver, and the National Center for Atmospheric Research (NCAR). Computing resources were also provided by DoD HPCMP under a Frontier project award and by the Air Force Research Laboratory.

This work would not have been possible without help from my friends and colleagues at CU. My advisers, Peter Hamlington and Greg Rieker, have been *incredibly* helpful. Their technical expertise has greatly influenced the content of this work. They have also provided valuable mentoring about life as an academic and provided helpful perspective on having a family while being successful. From the Turbulence and Energy Systems Laboratory (TESLa), Caelan Lapointe, Nick Wimer, Sid Nigam, Olga Doronina, and Colin Towery, along with the other members of TESLa, have contributed greatly to my understanding of CFD concepts. From the Precision Laser Diagnostics (PLD) Laboratory, Torrey Hayden and Dan Petrykowski have provided great insight and perspective into obtaining experimental data, as has the rest of the team.

Finally, I thank my family! To my wife, Shannon, I am eternally grateful for your support. Your smile and love provided solace after challenging periods of research and studying. To my sons: you can't read this yet, but hopefully one day you will take a look and understand why I couldn't always be there for you. Your endless energy provides inspiration for me; your youth propels me to try to make the world a better place in any way I can.

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

Contents

Chapter

1 Introduction and Background Information **1**

1.1 Characterization, Analysis, and Design of Thermal-Fluid Systems 3

1.1.1 Numerical Simulation Overview 6

1.1.2 Experimental Overview 7

1.2 Combining Simulations and Experiments Using Data-Driven Techniques 8

1.2.1 Overview of Data Assimilation Techniques 10

1.2.2 Approximate Bayesian Computation (ABC) 13

1.3 Completing the Engineering Tasks of Design and Optimization 16

1.3.1 Bayesian Emulation 18

1.4 Overview and Contribution of this Dissertation 19

2 Approximate Bayesian Computation (ABC) Methodology **24**

2.1 Approximate Bayesian Computation Background and Overview 25

2.2 Approximate Bayesian Computation Algorithm 29

2.3 Selection of Rejection Distance, ε 30

2.4 Markov-chain Monte Carlo ABC 32

2.5 ABC Summary 33

3 Demonstration of the ABC Approach **35**

3.1 Approximate Bayesian Computation for the Logistics Equation 35

3.1.1 Results of Logistics Equation: Random Parameter Generation 38
3.1.2 Results of Logistics Equation: MCMC Parameter Generation 40
3.2 Parameter Estimation in a Turbulent Buoyant Jet Using ABC 44
3.2.1 Demonstration Using Computational Reference Data 45
3.2.2 Demonstration of ABC Using Experimental Reference Data 55
3.3 ABC Methods: Additional Considerations 58
3.3.1 ABC Implementation with "Basic" Parameters 58
3.3.2 Impact of Rejection Distance (ε) on Posterior Convergence 59
3.3.3 Statistical Convergence of Posterior Profiles 63
3.3.4 Regression 65
3.3.5 Weighting Parameters to Form Posterior 67
3.3.6 ABC in the Presence of Model Bias 68
3.4 Conclusions 71

4 Parameter Estimation for Industrial Applications using ABC **73**

4.1 Approximate Bayesian Computation for a Turbulent Buoyant Jet 74
4.1.1 Turbulent Buoyant Jet Simulation Setup 74
4.1.2 ABC Statistic Selection 76
4.1.3 Turbulent Buoyant Jet Simulation Results: Iteration 1 77
4.1.4 Turbulent Buoyant Jet Simulation Results: Iteration 2 80
4.2 Approximate Bayesian Computation for a Turbulent Buoyant Jet with Rotating Cylinder 84
4.2.1 Approximate Bayesian Computation Algorithm 87
4.2.2 Determination of Jet and Cylinder Velocities 89
4.2.3 Simulation Setup 89
4.2.4 Approximate Bayesian Computation Setup 91
4.2.5 Selection of Simulation Statistics and Metrics 93

4.2.6 Results: Finding Inlet and Rotational Velocities 96
4.3 Determination of Jet Inflow Composition 101
4.3.1 Simulation Setup 102
4.3.2 Approximate Bayesian Computation Setup 104
4.3.3 Selection of Simulation Statistics 104
4.3.4 Results: Finding ϕ Using H_2O Concentration Measurement 105
4.3.5 Results: Finding ϕ Using Temperature Measurement 106
4.4 Conclusions 108

5 Parameter Estimation in Turbulent Simulations using Approximate Bayesian Computation with Laser Spectroscopy **110**
5.1 Introduction for ABC using Laser Absorption Spectroscopy 110
5.1.1 ABC applied to LES Simulations and Laser Absorption Spectroscopy Observations 113
5.2 Case Setup 114
5.2.1 Experimental Approach: Laser Absorption Spectroscopy 115
5.2.2 Computational Approach: Large Eddy Simulation (LES) 121
5.2.3 ABC Statistic and Metric Selection 124
5.2.4 ABC Prior Distribution Selection 126
5.2.5 Comparison of Computational and Experimental Results 127
5.3 Results 128
5.3.1 Posterior Distributions 129
5.3.2 Temperature Profile and Field Comparisons 132
5.3.3 Observing System Experiment 133
5.4 Conclusions 137

6 Conclusions and Future Research **138**
6.1 Conclusions 139

6.2 Future Work 140

Bibliography **142**

Tables

Table

4.1 ABC setup for baseline, refined, and further refined cases. This shows the range and increment for each parameter along with the total number of simulations in each case. Note that moving down in the table increases the refinement within the parameter space so that parameters simulated are closer together. 92

5.1 Experimental conditions 120

5.2 Experimental conditions and parameters from posterior mode 132

Figures

Figure

2.1 Algorithm depicts the steps required for the basic rejection algorithm described in Section 2.2 30

3.1 Converged solution mean of the logistics equation taken after 1,000 simulation steps. Note there are three μ regions (of which one value is shown per region) that result in mean values (and associated variance values shown in figure 3.2) within an ε value of 0.01; thus all three indicated μ values would be accepted with $\varepsilon = 0.01$. 39

3.2 Converged solution variance of the logistics equation taken after 1,000 simulation steps. Note there are three μ regions (of which one value is shown per region) that result in variance values (and associated mean values shown in figure 3.1) within an ε value of 0.01; thus all three indicated μ values would be accepted with $\varepsilon = 0.01$. . . 41

3.3 Histogram showing the number of accepted simulations versus values of μ. Simulation was run with 1 million values of μ randomly distributed between $0.85 - 1$, 1,000 steps, $\varepsilon = 0.01$. Reference μ value shown in red. 41

3.4 Histogram showing the number of accepted simulations versus values of μ. Simulation was run with 1 million values of μ randomly distributed between $0.85 - 1$, 1,000 steps, $\varepsilon = 0.001$. Reference μ value shown in red. 42

3.5 Histogram showing the number of accepted simulations versus values of μ. Simulation was run with 1 million values of μ chosen according to MCMC-ABC procedure, starting at $\mu = 0.9652$, 1,000 steps, $\varepsilon = 0.01$. Reference μ value shown in red. 42

3.6 Histogram showing the number of accepted simulations versus values of μ. Simulation was run with 1 million values of μ chosen according to MCMC-ABC procedure, starting at $\mu = 0.9652$, 1,000 steps, $\varepsilon = 0.001$. Reference μ value shown in red. . . . 43

3.7 Representative fields of speed v (a) and temperature T (d) from DNS for the periodically forced turbulent buoyant jet. Panels (b) and (c) show reference probability density functions (PDFs) and power spectral densities (PSDs), respectively, of v at heights of 5, 10, 15, and 20 cm. PDFs and PSDs at different heights are shifted vertically for clarity. Panels (e) and (f) show corresponding PDFs and PSDs of T. Solid black lines in panels (b,c,e,f) show DNS reference data and dash-dot blue lines show LES reference data. PSDs are computed using Thomson's multitaper estimate [1]. . 48

3.8 Schematic of the general ABC approach corresponding to Method D from Marjoram *et al.* [2], and of the specific ABC implementation used in the present study. 49

3.9 Marginal posterior distributions from ABC (visualized using Gaussian kernel estimation) for the frequency (a,b,g,h), amplitude (c,d,i,j), and mean (e,f,k,l) of the periodically forced turbulent buoyant jet. Posteriors are calculated using measurements at heights from 0–20 cm (indicated by colors from blue to yellow). Results are shown for speed-based reference data from LES (a,c,e) and DNS (g,i,k), and for temperature-based reference data from LES (b,d,f) and DNS (h,j,l). True parameter values are shown by vertical black dashed lines. Horizontal dash-dot lines and gray regions show the priors. 52

3.10 Vertical profiles of 95% Bayesian confidence intervals, C_B, for the marginal posterior distributions in Figure 3.9 (cases shown in each panel are described in the caption of Figure 3.9). Blue dash-dot lines and gray shading indicate the 95% Bayesian confidence regions, and the solid red lines show the mean values of the posteriors. True parameter values are shown by vertical black dashed lines. 53

3.11 Vertical profiles of Kullback-Leibler (KL) divergences for the marginal posterior distributions in Figure 3.9. Divergences for the frequency (black solid lines), amplitude (red dash-dot lines), and mean (blue dashed lines) posteriors are shown for LES and DNS speed (v) based reference data (a,c) and for LES and DNS temperature (T) based reference data (b,d). Each of the KL divergences are normalized by their respective maximum values. 53

3.12 Representative 2D fields of density (kg/m^3) from LES at six different times (a-f) for the steadily forced plume described in Section 3.2.2. The fields shown begin at an arbitrary time t_0 late in the LES and are separated by a time interval spanning one complete period, t_P ($t_P = 1/f = 0.22$ s). 57

3.13 Relationship between Strouhal number St $= fw/v_i$ and inlet Richardson number Ri $= (1 - \rho_i/\rho_\infty)gw/v_i^2$ for the steadily forced helium-air plume. Experimental results from Cetegen *et al.* [3] are indicated by open circles, and the empirical fit St $= 0.55\text{Ri}^{0.45}$ is shown as a blue dash-dot line. ABC results are indicated by filled red circles located at the mean Ri of the posterior distribution. Uncertainty bars show the minimum and maximum values of Ri in the posteriors. 57

3.14 Confidence interval half-width plotted against normalized rejection distance for the amplitude parameter posterior. 60

3.15 Confidence interval half-width plotted against normalized rejection distance for the joint posterior. 61

3.16 Posterior distributions for all three parameters created using "basic" statistics at 1 cm above the inlet. Posteriors are shown for two normalized rejection distances corresponding to valleys in Figure 3.15. 61

3.17 Kullback-Leibler (KL) divergence from the prior to the posterior distributions. Each KL divergence is normalized by it maximum value. High values of KL divergence indicate high information gain, while values close to zero indicate a posterior that is similar to the prior distribution. 63

3.18 Posterior distributions for three inlet velocity parameters. Posteriors created by accepting 100 closest simulations while increasing the size of the total simulations considered (from 1% of total $\approx 10,0000$ available to 100% of them); the lines are colored by total sample size considered. 64

3.19 Posterior distributions for three inlet velocity parameters. Posteriors created by accepting 2% of the total simulations considered that are closest to the reference DNS ensemble. Each posterior has a unique number of the total simulations considered (from 1% of total $\approx 10,0000$ available to 100% of them); the lines are colored by total sample size considered. 65

3.20 Regression schematic. Each blue dot represents a simulation output at its original location; this data is used to compute the original posteriors. The green line shows the regression line which is a best fit of the data. Each simulation is regressed along the green regression line until the it reaches the reference observation data from DNS simulation at which point the posterior is recomputed. An identical approach is applied for each parameter independently. 67

3.21 Posteriors before and after regression are presented. Regression causes each posterior to narrow. 67

3.22 Posteriors before and after weighting parameters by rejection. Weighting causes each posterior to narrow. 68

3.23 Vertical profiles of simulation speed are shown. One-hundred member ensembles of DNS and LES were computed. Each ensemble member was temporally averaged. Then the ensemble members were averaged together for either the DNS or LES ensemble. These means are shown in red and blue, respectively. The difference in profiles at each height is reported as the bias in the LES model and shown in green; the bias is close to 0 m/s for much of the domain. 69

3.24 Histograms showing posterior estimations for inlet mean velocity parameter. The uniform prior is indicated by the shaded gray box (normalized for the appropriate number of accepted simulations). The true parameter value used to drive the reference DNS ensemble shown with solid red line. Dashed blue line shows the mean of the posterior parameters. Dotted magenta line indicates mode of distribution. Subfigure (a) was created using original speed profiles, while subfigure (b) was crated using profiles after correcting for the LES bias. 70

3.25 Histograms showing posterior estimations for inlet velocity amplitude parameter. The uniform prior is indicated by the shaded gray box (normalized for the appropriate number of accepted simulations). The true parameter value used to drive the reference DNS ensemble shown with solid red line. Dashed blue line shows the mean of the posterior parameters. Dotted magenta line indicates mode of distribution. Subfigure (a) was created using original speed profiles, while subfigure (b) was crated using profiles after correcting for the LES bias. 71

4.1 (a) Schematic showing the setup of the RANS 3D domain. (b) Temperature field showing a single snapshot in time of a simulation with inlet mean temperature of 1460 K and temperature variability of 0.19. This 2D plane is a slice from the 3D volume; it is parallel to the y-z plane (the normal is in the x-direction) and located at the center of the jet ($x = 0.1875$). 75

4.2 Probability distribution showing the marginalized posterior of the inlet temperature mean parameter. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1400 K and 1800 K. The inlet temperature variations were between 0.0 and 0.2. Rejection distances of 20 K and 20 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. Accepted mean temperature values shown by green circles. Reference temperature mean shown in red at 1600 K. The mode of accepted temperatures is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values. 79

4.3 Probability distribution showing the marginalized posterior of the inlet temperature variation parameter. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1400 K and 1800 K. The inlet temperature variations were between 0.0 and 0.2. Rejection distances of 20 K and 20 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. Accepted temperature variation values shown by green circles. Reference temperature variation shown in red at 0.1. The mode of accepted temperature variabilities is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values. 80

4.4 Probability distribution showing the marginalized posterior of the inlet temperature mean parameter. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1400 K and 1800 K. The inlet temperature variations were between 0.0 and 0.2. Rejection distances of 26 K and 26 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. *Additionally*, rejection distances of 39 K and 13 K were applied in ABC method to the mean temperature and temperature standard deviation measured farther above the jet (with height equal to 11.5 mm) and centered in the width and length directions. Accepted mean temperature values shown by green circles. Reference temperature mean shown in red at 1600 K. The mode of accepted temperatures is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values. 81

4.5 Probability distribution showing the marginalized posterior of the inlet temperature variation parameter. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1400 K and 1800 K. The inlet temperature variations were between 0.0 and 0.2. Rejection distances of 20 K and 20 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. *Additionally*, rejection distances of 39 K and 13 K were applied in ABC method to the mean temperature and temperature standard deviation measured farther above the jet (with height equal to 11.5 mm) and centered in the width and length directions. Accepted temperature variation values shown by green circles. Reference temperature variation shown in red at 0.1. The mode is not shown as there is no single mode. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values. 82

4.6 Probability distribution showing the marginalized posterior of the inlet temperature mean parameter for *Iteration 2*. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1550 K and 1650 K. The inlet temperature variations were between 0.08 and 0.12. Rejection distances of 12.5 K and 6.5 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. Accepted mean temperature values shown by green circles. Reference temperature mean shown in red at 1600 K. The mode of accepted temperatures is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values. 84

4.7 Probability distribution showing the marginalized posterior of the inlet temperature variation parameter for *Iteration 2*. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1550 K and 1650 K. The inlet temperature variations were between 0.08 and 0.12. Rejection distances of 12.5 K and 6.5 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. Accepted temperature variation values shown by green circles. Reference temperature variation shown in red at 0.1. The mode of accepted temperature variabilities is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values. 85

4.8 The plot's vertical axis shows the prescribed inlet mean temperature that is assigned and held constant for each simulation. The horizontal axis shows the observed temperature; this value is the time averaged value of the temperatures measured at the center of the inlet (width and length wise) at a height of 2 mm over the last 20 seconds of the simulation. The reference simulation, with a prescribed mean of 1600 K, is indicated with the red circle. 86

4.9 (a) Schematic showing the setup of the RANS simulation. Measurements taken along the three green vertical dashed lines. (b) Temperature field averaged over the last 15 s of the simulations with jet velocity 1 m/s and rotational velocity of 12.5 rad/s. 90

4.10 Plot shows height above the hot jet versus the temporal average of simulation temperature at the jet centerline with fixed rotational velocity for the cylinder and fixed jet inlet conditions. The red circles correspond to LES while the blue squares are from RANS data; both simulations have the same boundary conditions and were run using OpenFOAM version 2.2.1. 91

4.11 Temperature versus time along the jet centerline at a height of 4.1 mm above the jet. The dash-dot blue line indicates the instantaneous temperature, while the solid red line shows the moving average. 93

4.12 Temporal average (over 15 s) of temperature as a function of height along the jet centerline. Panels show: (a) Variations in jet velocity, where blue circles correspond to a jet velocity of 0.3 m/s while red diamonds correspond to a jet velocity of 0.7 m/s, with a cylinder angular velocity of 12.5 rad/s in both cases; (b) Variations in cylinder velocity, where blue circles correspond to a cylinder angular velocity of 3 rad/s while red diamonds correspond to an angular velocity of 28 rad/s, with a jet velocity of 0.5 m/s in both cases. 95

4.13 Number of accepted simulations versus jet inflow velocity for 'baseline' case. Here 1071 simulations are performed over uniform priors with widths 0.0–1.0 m/s and 3–28 rad/s. The reference velocity at 0.5 m/s is shown in red with the mean of accepted values using Statistic 1 combined with Statistic 2 (all measurement locations) shown in green at 0.501 m/s, while the mean of the accepted values using Statistics 1 and 3 (two measurement locations) is shown in blue at 0.502 m/s (both green and blue lines are obscured by the red line). 96

4.14 Probability density functions (pdfs) for the number of accepted simulations versus values of cylinder rotational velocity. Solid black lines show Gaussian curves with the same mean and standard deviation as the accepted values. Here in the 'baseline' case 1071 simulations are performed over uniform priors with widths 0.0–1.0 m/s and 3–28 rad/s. Reference rotational velocity shown in red at 12.5 m/s with the mean of the accepted values shown in green at 12.2 rad/s for case (a) and in blue at 12.6 for case (b). 97

4.15 Probability density functions for the number of accepted simulations versus all simulated values of jet inlet velocity each with a pair of unique uniform velocity applied across the jet and cylinder rotational velocity; this is the 'refined' case. Here 8081 simulations are performed over uniform priors with widths of 0.3–0.7 m/s and 7.5–17.5 rad/s. Reference velocity shown in red at 0.5 m/s with the mean of the accepted values using only two measurement locations for metric 2 is shown in blue at 0.5003 m/s. Reference rotational velocity shown in red at 12.5 m/s with the mean of the accepted values shown in blue at 12.42 rad/s. 99

4.16 Probability density functions for the number of accepted simulations versus all simulated values of jet inlet velocity each with a pair of unique uniform velocity applied across the jet and cylinder rotational velocity; this is the 'further refined' case. Here 4141 simulations are performed over uniform priors with widths of 0.48–0.52 m/s and 10–15 rad/s. Reference velocity shown in red at 0.5 m/s with the mean of the accepted values shown in blue at 0.500001 m/s. Reference rotational velocity shown in red at 12.5 m/s with the mean of the accepted values shown in green with the mean of the accepted values shown in blue at 12.29 rad/s. 100

4.17 Joint probability distribution function (pdf) showing the accepted values for the 'further refined' case. Of the 4141 cases, 975 are accepted. The peak of this joint pdf occurs at 0.5005 m/s and 12.45 rad/s. 101

4.18 Plot shows height above the hot jet versus the temporal average of simulation temperature at the jet centerline with fixed rotational velocity for the cylinder and fixed jet inlet conditions. The red circles correspond to LES while the blue squares are from RANS data; both simulations have the same boundary conditions and were run using OpenFOAM version 4.1. 103

4.19 Plot shows height above the hot jet versus the temporal average of simulation H_2O mass fraction at the jet centerline with fixed rotational velocity for the cylinder and fixed jet inlet conditions. The ϕ values used to set the inflow conditions were evenly distributed from 0.3 to 1.0. Note that several of the profiles cross higher in the domain indicating a complex behavior exists. Thus, an analytical solution is not practical and simulations are necessary to understand the behavior of the fluid enabling accurate parameter estimation. 106

4.20 (a) Confidence interval half width plotted against ε (rejection distance) values for *H_2O concentration* measurements. This value of ε strikes a balance between having high precision in the posterior (i.e. low standard deviation) and high confidence in the posterior (i.e. a large number of accepted samples). (b) Discrete PDF showing the number of accepted simulations versus values of inlet fuel-to-air ratio, ϕ, along with a Gaussian curve with the same mean and standard deviation as the accepted values. Here ∼10,600 simulations are performed over a uniform prior with ϕ 0.3–1.0. Statistic used is temporal mean *H_2O concentration* at one location in the domain. Reference ϕ shown in red at ∼0.65 with the mean of the accepted values shown with a dashed blue line. 107

4.21 (a) Confidence interval half width plotted against ε (rejection distance) values for *temperature* measurements. This value of ε strikes a balance between having high precision in the posterior (i.e. low standard deviation) and high confidence in the posterior (i.e. a large number of accepted samples). (b) Discrete PDF showing the number of accepted simulations versus values of inlet fuel-to-air ratio, ϕ, along with a Gaussian curve with the same mean and standard deviation as the accepted values. Here ∼10,600 simulations are performed over a uniform prior with ϕ 0.3–1.0. Statistic used is temporal mean *temperature* at one location in the domain. Reference ϕ shown in red at ∼0.65 with the mean of the accepted values shown with a dashed blue line. 108

5.1 Experimental setup for laser absorption spectroscopy above catalytic burner 116

5.2 Vertical temperature profiles corresponding to three experimental cases. Experimental uncertainty shown based on instrument validation 121

5.3 (a) Schematic showing the setup of the 3D LES domain. (b) Temperature field (volume rendering within an isosurface at 500 K and additional volume rendering at higher locations) a single snapshot in time of a simulation with inlet mean temperature of 1515 K and inlet speed of 0.38 m/s with no additional heat added above the burner surface. (c) Speed field corresponding to temperature shown in part (b) . . . 123

5.4 Posterior distributions for cases 1, 2 and 3 (top, middle and bottom row, respectively) for inlet speed, height height, and power (left, middle and right columns, respectively). Line colors correspond to percentage of accepted simulations as determined by ε parameter. Inlet speed posteriors show range of predicted velocities in the experimental results 134

5.5 Marginal posterior distributions for cases 1, 2 and 3 (top, middle and bottom row, respectively) for two parameters taken when ε is set to accept 20% of simulations. Colors represent density of posterior with yellow being highest density and blue being lowest density 135

5.6 Temperature profiles for experimental data are shown with empty circles. Error bars denote experimental uncertainty. Solid lines denote simulation data. Data colored according to experimental case as determined by power influx 136

5.7 2D slice of temperature fields (taken in the middle of the burner in the x-direction) corresponding to three simulations, each with with parameters taken from the mode of the marginal posterior distributions in Figure 5.4 and listed in Table 5.2 for each of the three cases. 136

Chapter 1

Introduction and Background Information

Computational simulations are widely used to design and analyze systems involving complex thermal-fluid flows. From microscale heat transfer in thermal management devices to liquid propulsion in heavy-lift rockets, nearly all such simulations are intended to provide three-dimensional (3D) spatially and temporally resolved numerical solutions of physics-based governing equations for realistic geometries, boundary conditions, and material properties. Prior to making simulation-based design and operational decisions, however, computational accuracy must be demonstrated by validating against (typically experimental) reference data for realistic or canonical test cases. As simulation fidelity has improved over the past decade with the development of higher-order and geometry-resolving numerical techniques, as well as with the increasing use of petascale computing resources to achieve very fine spatial resolutions and improved physical realism, a long-standing difficulty in such validation efforts has come into sharper focus.

Namely: *Even if a simulation is able to solve physically-realistic governing equations with high accuracy, unavoidable uncertainties in real-world boundary conditions, material properties, and other parameters result in ambiguity as to whether the computational and real-world systems are actually equivalent.* That is, there are few assurances that discrepancies between computational and experimental results during a validation test are not simply due to differences in the initial or boundary conditions between the real-world and simulated systems. This scenario can occur, for example, in cases where knowledge of a real-world system is limited, perhaps in experiments that were not originally intended for validation purposes, systems with limited access, or parameters

that are difficult to measure directly [4]. Discrepancies may also be due to model parameters that need to be optimized (see [5]). Moreover, agreement between simulations and experiments may also be hampered by inadequate physics modeling (e.g., unresolved flow features due to discretization or insufficient turbulence modeling, lack of combustion, lack of multiple present phases such as soot or condensation). Thus, it is important to address as many of these simulation limitations as is computationally affordable before trusting simulation results for the design and operation of real-world systems.

In broad terms, this difficulty can be addressed by proposing distributions of likely parameter values, performing simulations with parameters sampled from those distributions to determine the spread of outcomes in a particular quantity of interest within a flow, and then using statistical inference to determine likely distributions for unknown parameter values [4, 6–8]. This approach can, in principle, be attempted using full Bayesian analyses, which have recently gained popularity for parameter estimation in engineering applications [9–12]. However, in nearly all such analyses, the requisite components of Bayes' theorem (specifically, the likelihood function) may be unknown or enormously costly to compute and, consequently, non-physics-based reduced-order surrogate models have often been used to sample the unknown parameter space [10, 11, 13–15]. Optimization techniques have also been used for parameter estimation (e.g., [16–20]), but these methods seek to provide single values of unknown parameters, with no intrinsic measure of uncertainty when using potentially imperfect computational models and real-world data.

In this dissertation, approximate Bayesian computation (ABC) is used, for the first time, to estimate unknown parameters in thermal-fluid flows. The power of ABC lies in the fact that far fewer simulations are required than in full Bayesian analyses since ABC does not require a likelihood function, thus permitting the use of physics-based models. A wide variety of reference data can also be used to drive the estimation, including measurements that are only indirectly related to parameters of interest. The technique naturally provides probability distributions for unknown parameters, allowing Bayesian confidence intervals to be obtained along with parameter estimates.

This dissertation is one of a growing number of efforts to use data-driven methods with either experimental or higher-fidelity computational reference data to improve accuracy and quantify uncertainty in turbulent flow simulations. For example, ensemble Kalman filtering (EnKF) approaches have been used to assimilate reference data in simulations [21–24] and to infer turbulence model discrepancies [25]. In such approaches, reference data are used to update the state of a simulation or parameters in the simulation as the simulation progresses. Generally, however, EnKF methods require extensive and high quality reference data to provide reliable state and parameter estimates. The maximum a posterior (MAP) method has also been used to infer both parameter [26] and model [27] uncertainties. This method maximizes the posterior using either an analytical function or a Monte Carlo approach, but the analytical function used to represent the posterior (or the likelihood function) is often only known approximately, and the number of simulations required in even Markov chain Monte Carlo approaches can be enormous. Recently, the ABC method has been used to infer Arrhenius parameters for chemical kinetics rate coefficients in the context of combustion [28]. The present dissertation is specifically targeted at problems for which the reference data are spatially and temporally sparse, statistical in nature, or otherwise lacking in sufficient detail to update the model simulations at all locations and times (precluding the use of EnKF methods) and for which the true posterior is either intractable or prohibitively expensive to compute (precluding the use of MAP methods).

It should be noted that this dissertation focuses on demonstrating the use of ABC for estimating *physical* parameters that are difficult to obtain for real-world systems, as opposed to calibrating *model* parameters used in reduced-order or engineering models of more complex phenomena. Such model parameter calibrations are also possible using ABC and are the subject of future research [5].

1.1 Characterization, Analysis, and Design of Thermal-Fluid Systems

Numerical simulations and real-world experiments play key roles in understanding the types of engineering systems that are ubiquitous in modern society. Numerical simulations provide a wealth of data, often including fully resolved spatial and temporal domains with information on

many key quantities of interest such as velocity, temperature, and species concentrations [29]. However, numerical simulations are plagued by a key limitation; namely, persistent questions regarding the accuracy and reliability of simulation results and conclusions [4]. Experimental data derived from measurements of physical systems largely avoid this problem by offering highly accurate results for some of those same quantities. However, experiments are often limited in their dimensionality, only recording data from a small portion of the domain at any given time and for a small subset of the quantities of interest [30]. Thus, combining observations of the "truth" (which will be defined various ways throughout this dissertation and usually referred to as the "reference" data) with simulations provides a possible path to improving and overcoming these shortfalls. Specifically, the reference measurements bring credibility to the numerical simulations, while simulations provide additional insights into the system of interest. This dissertation demonstrates an innovative application of approximate Bayesian computation (ABC) to unify numerical simulations and experimental measurements. Before experimental observational data are used, however, reference data is drawn from either the same or higher fidelity simulations in order to benchmark how well ABC performs in situations with complex thermal-fluid flows, with a particular focus on systems involving coupled interactions between fluid mechanics and combustion.

In this dissertation, the term 'combustion engineering' is used in a general and broad sense to include systems that may or may not have active reactions in the region of interest (e.g., above a catalytic burner there are theoretically no additional significant reactions assuming complete combustion within the burner, or only limited amounts of additional reactions if the catalyst becomes saturated). The flow fields of interest, however, generally require combustion to provide the high temperature inlet conditions. The desired outcome of combustion engineering usually relates to answering a question regarding the performance of a system in one (or more) of three regimes: in everyday settings, in extreme conditions, and/or throughout its lifetime. Within these three performance areas, combustion engineering focuses on two primary tasks: characterization and analysis, as well as design and optimization. Each of these tasks encompasses a vast array of unique challenges, yet together they yield solutions for systems ranging from rocket and jet engines, to internal

combustion engines, to power generation and industrial processing techniques.

When completing the characterization and analysis of a combustion engineering system, it is important to consider relevant quantities of interest for the task. The choice of pertinent quantities of interest will often guide which method to use, depending on whether experiments, simulations or analytic approximations can better provide the desired information. Often for combustion performance, temperature and pressure are key drivers as they quantify an available potential to perform work or drive a process. To achieve the necessary temperature and pressure, engineering economics must address the efficiency and cost of a system. Pollution generation is also studied in combustion-driven systems, and can be linked directly to their performance (i.e., certain pollution reduction mechanisms also reduce performance). In this case, one needs to consider the concentration of product species with a particular focus on radicals. Practical considerations also bound engineering systems; combustion subsystems are often integrated into larger systems such as aircraft and thus the weight and size are key design limitations. Another practical consideration for the system focuses on its reliability, maintainability, availability, durability, and stability; if a system has exceptional performance but does not operate when needed, it is of little use to the end user.

Ultimately, the tasks of characterization and analysis focus on understanding existing systems. This includes modeling and measuring their performance across a wide array of operating conditions and durations to accurately predict how systems will operate once fielded. Two primary methods to characterize and analyze a system include simulations and experiments. Note that theoretical or analytic approaches are not generally feasible for advanced fluid dynamics [29] or combustion engineering [31] due to the complexities of the systems involved (e.g., due to complex problem-specific geometries, the presence of nonlinear mixing by turbulent flow fields, heat release and radiative effects due to complex chemical processes, highly-coupled physics spanning many spatial and temporal scales).

1.1.1 Numerical Simulation Overview

Numerical simulations of engineering systems span wide ranges of physical fidelity and computational cost. Most often, these two attributes are proportional. On the low cost, low fidelity end of the spectrum are reduced-order models based on simplified equations, geometries, and conditions. On the other end of the spectrum are extremely expensive but much more accurate computer models built from 3D computer-generated geometries that involve the solution of physics-based coupled partial differential equations. Such CFD approaches [29] include direct numerical simulations (DNS) (see for example, [32–35]), large eddy simulations (LES) (see for example [36–40]), and Reynolds-averaged Navier-Stokes simulations (RANS) (see for example [41, 42]). Simulations typically provide a wealth of information at comparatively affordable costs. Simulations are generally considered "cheaper" to run than experiments because once capital expenditures are made up front (which usually support a wide swath of computational challenges and, while high, are often paid for by agencies besides the end user), the incremental costs to keep mainframe computers operational are minor and shared by many users [43]. Simulations are often cheaper in terms of time, as well, because models execute in minutes or hours, can be run in parallel, and can be modified and rerun with little effort.

Simulations can provide data with extremely rich temporal and spatial resolutions, limited only by the amount of computing power available. However, with these benefits come certain limitations. In particular, many simulations of real-world systems are plagued by inaccuracies that stem from the simplifications and physical approximations used to model such systems (e.g., discretization ensures that the system will not be modeled at every point in space and can thus fail to capture certain phenomena, and the geometry of a model may not accurately represent a complicated physical setup), as well as uncertainties in the boundary and initial conditions that drive and bound the equations. Many of the physical parameters in the simulations might be difficult to determine or are unknown (e.g. material properties, transport coefficients, chemical composition, etc.), driving further potential inaccuracy into the results. These uncertainties lead

to significant doubt about the wisdom of exclusively using models in engineering and, indeed, force a complementary reliance on experimental data.

1.1.2 Experimental Overview

Experimental methods are highly robust in answering a variety of engineering analysis questions [30]. Trust in experimental data is an almost innate quality in humankind; each child starts his or her own series of experiments early in life dropping an item off of a table to see what will happen and forming conclusions about the way the world operates from such observations. The history of academically accepting experimental evidence stems back to Galileo, widely known as the first experimental scientist [44]. Experiments are so revered because they are based on the truth. Though they may contain error, they are routed in actual phenomena capturing all of the physics and chemistry that are extremely difficult to model at every scale. Data are often extremely accurate based on known calibrations. Experimental methods can answer a variety of engineering analysis questions and typically cannot avoid involving all of the relevant physics. For thermal-fluid systems there exists an enormous variety of techniques to obtain data experimentally, depending on the quantity one desires to measure. To measure pressure, for instance, one can use a fluid barometer or manometer, a mechanical pressure gauge such as a Bourdon-tube gauge, a pressure transducer such as a piezoelectric device, or even pressure sensitive paint [30]. Devices to measure temperature are equally varied; one could use a traditional liquid thermometer, a thermocouple, infrared thermography, or laser spectroscopy [30, 45]. Each of these experimental techniques has associated strengths and weaknesses.

In general, experiments of thermal-fluid systems provide highly accurate information, but at a limited number of locations and times, and only for a limited number of parameters (as compared to simulations). Experiments are generally more expensive to conduct because they require costly measurement equipment and require a mock-up of the system to be measured. Building and configuring both of these components takes a considerable amount of time, in addition to the duration of the experiments themselves and subsequent post processing and analysis. Lastly, once

running, experiments can be inherently more dangerous than computer simulations, prohibiting tests of extreme conditions (especially in the combustion realm). Thus, a key challenge existing within combustion engineering is to utilize sparse experimental information to validate and improve simulation accuracy. Tight coupling of experiments and simulations, to leverage the benefits of each and minimize their weaknesses, will be the focus of subsequent efforts in this dissertation.

1.2 Combining Simulations and Experiments Using Data-Driven Techniques

All of the aforementioned strengths and weaknesses of experiments and simulations set the stage for combining the two methods into an even more powerful tool. The benefits of experiments can improve on the shortfalls of simulations (primarily, their uncertainty), while simulations are able to address limitations of experiments (primarily, the restricted information content of measurements). There are two aspects of overcoming simulation limitations: improving model fidelity (including accurately replicating the physics represented by the model as well as the initial and boundary conditions), and improving the speed of the model with minimal impact on its fidelity (as will be discussed later, this is typically for optimization purposes). The challenge, then, is to fuse simulations and experiments into a robust tool that provides useful, improved quantities of interest relevant to the particular engineering system.

Within combustion research, an emerging area of interest combines experimental observations with computational models typically developed by discretizing governing equations. There are several methods to link these two unique sets of data depending on the final goal. Possible objectives when connecting experiments with computations include estimating parameters, matching flow-field statistics, or matching the field temporally and throughout the spatial domain. Looking at each of these more specifically, parameter estimation aims to better understand boundary (or initial) conditions (for instance, species concentration, thermodynamic properties such as a burner temperature or pressure, or kinematic quantities such as an inlet velocity) and physical characteristics of the problem (for instance, radiative heat loss or transport properties such as convection coefficients). Studies considering flow-field statistics aim to match the observation and model domains during a

prescribed time window on a purely statistical basis. This means that the temperature, velocity, pressure, or species concentration, for example, are measured over a prescribed time duration in both the model and observational domains, then statistics (e.g. mean, variance, skewness, etc.) of time-averaged results are compared. Finally, matching the field step-by-step entails matching the temperature, velocity, species concentrations, pressure, flame front position, etc. at each location in the field at each time step corresponding to an observation. Depending on which of these objectives one aims to achieve, the optimal method of combining simulations and experiments may differ.

First, consider the objective of parameter estimation. This task involves the determination of correct input parameters for a model such that the model is then able to progress forward and produce a field "similar" to that observed experimentally. "Similar" is, of course, a word that may have different meanings in different contexts. This leads to the next objective of matching statistics of the field versus matching the field point by point and from time to time. Here the methods are focused on the statistics of the computational field based on the parameters chosen; thus, in this context "similar" fields will have matching statistics. This focus on statistics is driven by the nature of the system being modeled. The stochastic behavior of most combustion applications means that the slightest anomaly in boundary or initial conditions can lead to a different time-dependent solution; this is commonly referred to as the butterfly effect. Thus, since one can not know the boundary or initial conditions exactly, the aim is instead to understand the general behavior of the system through its statistics. This goal is most applicable to systems where long-term performance is of ultimate interest; that is, when the primary mode of system operation is steady-state over a long time duration. In such cases, the system performance over a longer duration of time is best described by its statistical behavior. This description might apply to, for instance, an industrial burner used to continuously affect a given process such as heat treating a film, cooking a food item, or a jet engine operated with a constant throttle setting. With this objective in mind, one should note that a field obtained computationally is averaged over time, and thus time dependence is of little relevance as long as a time-dependent solution would provide the desired statistics. Since the statistics are of value in the computational realm, the statistics are also valuable on the

observational side. That is to say, the statistics of the observations are of prime importance when linking experimental results to a computational model.

1.2.1 Overview of Data Assimilation Techniques

Data assimilation has been widely used in different fields to link observations with model data, especially when a time-dependent, spatially accurate solution is of interest. Data assimilation is commonly used for dynamical control systems in engineering application such as robotics or aviation, where it is often accomplished through Kalman filtering [46]. It is also a prominent technique in meteorology, where data assimilation is the preferred tool to connect a multitude of observations (for example, space-based radiance measurements, radar measurements, wind stations, buoys, weather balloons, etc. [47]) with highly complex three-dimensional models for forecasting temperature, precipitation, and wind behavior, among other physical properties [48]. The assimilation process ingests real observations to update the model and provide more accurate forecasting at each location. The basic process relies on the use of Bayesian statistics to compare a prior distribution (i.e., the initial state of the system, which could come from a model, empirical observations, previous combinations of model and observations, etc.) to a likelihood distribution (i.e., the observations) in order to produce a posterior (i.e., the state of the system given the available model and observation data, each weighted according to its expected accuracy as captured in the variance of each data set) [49, 50]. These steps can be performed in real-time as data becomes available in order to continually update the model state and provide better forecasts based on forward runs of the model. This specific and narrow definition is what is meant by "data assimilation" in this dissertation (as opposed to more broad interpretations that include virtually any combination of observations and models).

Two of the primary techniques to conduct data assimilation include ensemble Kalman filtering (EnKF) [50] and 4D-VAR [51]. These techniques both attempt to combine observation data with models to improve predictions of the system's state and allow for improved forward runs of simulators. EnKF achieves this by acting as a particle filter running several parallel models simulta-

neously and combining their results to account for the sample covariance in the model. Approaches that use 4D-VAR assimilate data over a wide time window to improve agreement between the model forecast and observations within the assimilation time window. Future efforts may continue to combine these two approaches as seen in [52].

With this application in mind, it becomes readily apparent that the focus of data assimilation is on applications wherein a time-dependent state estimation is desired. That is to say, data assimilation is helpful when one desires to know each specific detail of a field (i.e., the entire state vector) at each time step. For instance, in weather prediction it is desirable to know the temperature, wind speed, and precipitation at each location at each moment in time – that is, the information one seeks when looking up the weather at a particular place on the map for a particular date. This approach also implies that the model itself, along with its boundary and initial conditions and driving parameters, are less important as long as the prediction is accurate for a short time forward (which is achieved by continually assimilating newly acquired data). Note that weather forecasts 12 hours out are much less reliable than forecasts for the same time period made 6, 3, or even 1 hour out from newly assimilated data [53]. Weather forecasting benefits drastically in the near term due to data assimilation, as do other fields such as radar tracking and guidance, navigation and control [46]. Data assimilation has also been applied to parameter estimation as outlined in [24].

Extending these characteristics to combustion modeling, one can draw several conclusions about the applicability of data assimilation. For combustion, and turbulent flows in general, one is often primarily interested in flow-field statistics. This focus is driven by the stochasticity and nonlinearity of the governing equations. A small perturbation in boundary or initial conditions, for instance, can lead to a different solution. Thus, it is difficult to correctly predict each eddy and perturbation of the flow field at a given instant in time. Instead, one is generally concerned with the statistics of the flow field, e.g., the temporal temperature mean and standard deviation at a particular location in the domain. The mean and standard deviation could come from running a model for a long time or from executing multiple iterations of the model (giving ensemble statistics).

These statistics would then allow a designer to know the characteristic behavior of the flow field and plan accordingly when deciding how to construct an industrial system or machine.

As a result, data assimilation may not be the primary tool of interest when attempting to utilize sparse observations in order to improve combustion engineering model performance (when model performance is characterized by its ability to correctly match observation statistics without being continually nudged by observations). However, if the time-dependent (instantaneous), spatially accurate (local) solution to the model is desired, and an abundance of spatially and temporally rich data is available, then data assimilation is an appropriate method. This latter performance is indeed the primary goal of several existing combustion studies, including estimating the location of a wild-fire flame front using a temperature field obtained with an aircraft mounted infrared camera [54], determining the location of a 2D flame front using synthetically generated temperature measurements from a model [55], or predicting the behavior of a flame in a scramjet combustor using data from stereoscopic particle image velocimetry and coherent anti-Stokes Raman spectroscopy [56]. These problems all focus on cases where it is important to know the behavior of a single implementation of a model versus the general behavior applicable to many executions of the model. In the case of [55], the author uses data assimilation to drive a model to match observations of a frame front location; from this match, the flame speed parameter can then be predicted.

Thus, data assimilation tools can be used to estimate parameters by augmenting the state vector. However, this approach has its limitations. First, to be successful in thermal-fluid applications, predicting parameters using data assimilation typically requires extensive observations. For example, in [55], the observations are synthesized from another run of the same model as the one used in the assimilation, but with a different flame speed parameter chosen; in [57] many velocity measurements throughout the domain are used to reduce model-form uncertainties in RANS. Second, data assimilation is a less direct method of determining unknown parameters, particularly in comparison to ABC, since parameter estimation using data assimilation requires state estimation at each time step [24]). For these reasons, data assimilation has not been used as the tool to determine unknown parameters in this work.

1.2.2 Approximate Bayesian Computation (ABC)

A technique that appears ready to solve the problem of parameter estimation using flow field statistics is approximate Bayesian computation (ABC). This method seeks to directly find a set of parameters that would produce the appropriate statistics of the observations. ABC is introduced in this section, and expanded more fully in Chapter 2.

The ABC technique [58, 59], like other Bayesian techniques, are based on their namesake, Bayes' theorem,

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}\,, \tag{1.1}$$

to solve for the probability of the parameters, θ, given the observations or data (D) available from experimentats or, alternatively, a similar or more refined model [i.e., the goal is to solve for $P(\theta|D)$, also known as the posterior]. The probability density function of the parameter, independent of observations, is known as the prior and is denoted $P(\theta)$. The probability density function of the experimental observations (data) is known as $P(D)$; this is obtained through analyzing experimental results, but it is often neglected since it serves merely as a normalization constant in Bayes' theorem. Lastly, the probability of the data occurring given the particular parameter, $P(D|\theta)$, is known as the likelihood function.

Traditional *full* Bayesian techniques either approach the problem of finding the posterior analytically (although this is often not feasible for complex models) or through Monte Carlo approaches. A standard Monte Carlo approach to find the posterior, called rejection sampling, is presented in [60] for observations that can take countably many values (i.e., discrete and not continuous variables). In this method to find the exact posterior, many parameter values are chosen, then the simulator is run forward to produce data. This data is compared to the observations and if it exactly matches then the parameters are retained. These accepted parameters are then used to construct the posterior. This method is not practical for many engineering systems as there are too many degrees of freedom and variables are typically continuous; thus, the probability for an exact match between observations and simulation data is extraordinarily small. Consequently,

two approximations are made in order to form the basis of *approximate* Bayesian computation. Namely: 1. the data are represented by their statistics and 2. the statistics of the simulation and observations need not match exactly but rather are only required to agree within some prescribed distance.

Next we examine these two approximations and ABC nomenclature in more detail. In ABC, the observation data, D, [1] is typically represented by its statistics, $\mathcal{S}$ (for the approximate data coming from a simulation, $\hat{D}$, the statistics are $\hat{\mathcal{S}}$). Since the dimensionality of the data can be extremely high in combustion and fluid dynamics systems, representative statistics of the data are used throughout this dissertation. The likelihood function, $P(D|\theta)$, is usually very difficult to obtain either due to its mathematical intractability or its computational in-feasibility. To quantify this statement, consider the example of full Bayesian analysis given in [60]; to produce the posterior and predict a single parameter based on observations with relatively few degrees of freedom, 20 million data sets were simulated. Indeed, in a complex combustion model or fluid dynamics model one can safely assume the likelihood will not be available in the majority of cases. The lack of an available likelihood function is where the ABC method reveals its value; instead of needing to compute the exact likelihood function [i.e., $P(D|\theta)$], the ABC method instead computes many approximate realizations of the data ($\hat{D}$). The method accomplishes this by first sampling values of the parameter from the anticipated parameter space [i.e., from the prior, $P(\theta)$]. The prior could be a uniform distribution of each parameters or a Gaussian distribution, for example. The method next takes a value of θ (i.e., a parameter) and runs the model forward to come up with model data ($\hat{D}$).

ABC then compares each approximate realization ($\hat{D}$) with the experimental data, D, and either keeps the parameters used to produce that data set, or rejects the parameters depending on a predefined distance function $\delta(\cdot,\cdot)$, and associated threshold (ε) [i.e., $\delta(D,\hat{D}) < \varepsilon$]. The distance function could, for instance, compare the mean (a statistic of the data) of the reference data to

[1] In this dissertation, "observation data" D may also be referred to as "reference data" (since it is the data ABC *refers* to for comparisons) or "truth" data (since at times a value of parameters is chosen a priori then ABC attempts to find these known *true* parameters).

the mean of each simulation using an L1 norm (the metric). Once a comparison is made, the resulting distance is compared to a threshold value. If the distance is less than the threshold, the parameter is considered viable and it is added to the candidate pool. However, if the distance exceeds the threshold, then the parameter is discarded because it did not produce a realistic model outcome compared to the available observation. From all of the accepted parameter values, one can then construct a distribution of likely parameters that would result in the given observations; this distribution is representative of the true parameters assuming negligible bias in the data or model. The distribution thusly created, $P_{ABC}(\theta|D) = P(\theta|\delta(D, \hat{D}) < \varepsilon)$, which is an approximation of the full Bayesian posterior, $P(\theta|D)$, is known as the posterior and is indeed the very item of interest.

Looking more closely at the particular application of a catalytic burner (an example case later in this work), ABC is used to determine unknown boundary parameter values. Laser spectroscopy methods are used to obtain line-of-sight (LOS) averaged, absorption-weighted temperature measurements at various locations within the flow field. These measurements have very rich temporal data that can be used to generate a probability density function (pdf) with relevant statistics (e.g., time-averaged temperature and associated variance or a complete pdf of temperature values or a power spectral density estimate such as an FFT of the data); this is $P(D)$ in Bayesian terms. Next, a computational model simulates the same setup as was measured experimentally. The values for parameters of interest (e.g., boundary conditions including inlet velocity at the burner in addition to heat addition within the domain) are chosen according to a prior distribution, $P(\theta)$. The prior distribution for these parameters is based on mass and energy balances as well as physical intuition about the system. Each draw of parameters produces a separate solution (by computing a forward solution using those parameters) whose statistics are analyzed and compared to the experimental data. As described previously, if the statistics agree according to a predefined threshold, the parameters will be kept, or otherwise rejected. The resulting set of parameters makes up the posterior distribution, $P_{ABC}(\theta|D) = P(\theta|\delta(D, \hat{D}) < \varepsilon)$.

In summary, ABC is a powerful tool to develop estimates for parameters given a set of statistical observations. The method works by sampling a large set of parameters from the anticipated

parameter space and running a model forward using one of those parameters. It then compares the resulting state space with the observed state space. If the results are close, then it keeps the parameters; if the results are not close, then it discards the parameters. This process repeats until many sets of parameters have been modeled and an adequate number have been accepted. These parameters make up the desired posterior distribution.

1.3 Completing the Engineering Tasks of Design and Optimization

Returning to the overarching concept that combustion engineering aims to study the performance of systems in several different realms, we next consider design and optimization. [2] This area focuses on two main time frames within the development of a system. First, the design process looks at a system from its genesis. It focuses on what a system will do that makes it unique and worth creating versus using an existing system. Additionally, design might focus on modifying an existing system to make it better match a new application or to improve its performance. In both instances, design happens early in the life cycle of a system, whether it be during its creation or during its modification. Because of the fast pace of creation, experiments are often not readily executable due to a lack of existing hardware and constantly changing specifications. Instead, the theory of operation and simulations can serve as primary tools early in the design process to flush out ideas that will later be built and realized physically. Theoretical approaches usually provide the basis for initial designs. They are based on idealized, simplified conditions that form the starting point for more complex subsequent iterations. Simulations are particularly useful for design work due to their flexibility in implementing new geometries and physical parameters.

The second portion of this task, optimization, is a highly versatile tool that provides deep insight about the performance of a system and offers methods to improve the performance. A variety of optimization tools exist, including linear and nonlinear, gradient-based or gradient-free,

[2] This dissertation does not directly attempt to design or optimize systems; indeed this section is presented for completeness for the engineering process. One interpretation of ABC is that it can be a form of optimization that attempts to minimize the distance between a model and reference data. However, this description is not within the intended definition of "optimization" presented here.

and adjoint techniques [61]. Optimization techniques calculate a cost function to compare the performance of a system to a desired outcome (which could be reference data from an experiment or to maximize or minimize some performance characteristic) and a method to explore the parameter space to find local and global minimum to the cost function. Optimization typically occurs during the design process to affect the final design and lead to a better outcome. Because of this timing, optimization is typically relegated to the same two methodologies as design, namely theory and simulation. Additionally, optimization tools often require a large number of iterations making experimental approaches unfeasible.

To successfully complete the task of designing and optimizing a system, one requires informative quantities of interest. These are often the same quantities investigated during characterization and analysis of a system. For instance, when designing a combustion system it typically needs to provide a given output of power or heat to serve the purpose for which it is being created. A step to help quantify the output is to measure the designed operating temperature and pressure of the system. Cost is a prime driver in most designs, so it must be tracked through a careful understanding of the price to procure parts, manufacture the device, and operate it. Securing long term operations through improvements in the reliability, maintainability, and availability of a system is a goal common to most industrial-grade engineering tasks for items that will be used for long durations. The weight of the system must also be considered, as it will drive how a system can be used (for transportation applications) or how much it costs to transport and deliver initially as well as potentially impacting future maintenance activities (e.g., if an item needs to be elevated to work on a certain component or shipped back to a factory for calibration or repairs). Additionally, the pollution profile of a system will become increasingly important as more combustion systems penetrate the market and emission regulations become more difficult to meet.

Optimization seeks to improve on any one of these parameters (or possibly several of them) by changing design inputs, tracking results, and iteratively seeking to minimize a predefined cost function that describes the performance of a system in terms of desired parameters. This is the goal of many combustion engineering efforts. For instance, the National Renewable Energy Laboratory's

Co-Optimization of Fuels and Engines (Co-Optima) program [62] seeks to simultaneously optimize engine fuel efficiency and the fuel they use, instead of a stove-piped approach to only improve one or the other. With this task in mind, it is clear that a fast model will be required to enable optimization techniques. One such technique is Bayesian emulation, along with other machine learning techniques.

1.3.1 Bayesian Emulation

Bayesian emulation is a powerful tool used to create a simpler, cheaper alternative to an expensive model [63, 64]. The basic premise of this technique is that a model simulation can be very resource intensive to run. Thus, it can be helpful to develop an emulator to run in place of the simulator (i.e, the original expensive model) to get a reasonable estimate for the state space that would result from the simulator given a set of parameters/inputs. Note that in this thesis Bayesian emulation is presented for context, however it is not implemented because the base RANS and LES cases are computationally affordable enough to use directly with ABC without the needs to create reduced order models.

As the name implies, this technique is also based on Bayesian statistics. The primary objective of this technique is to develop a cheap statistical representation of an expensive model. To do this, in a manner similar to ABC, a set of input parameters are chosen to run in the simulator. These parameters should cover a wide range so as to simulate a large set of anticipated future parameters of interest. The output of the simulator is then recorded for each set of inputs. The Bayesian emulator is created from this training data and serves as a statistical model to predict, for a given input, what the expected output would be. The emulator is, in statistical terminology, the posterior distribution of the model outputs given the training data set [65]. An emulator must satisfy two criteria [63]. First, it must return the exact simulator output with no uncertainty for inputs used as training data. Second, it must give reasonable approximations of the output (i.e., close to what the simulation would provide based on extrapolating or interpolating the training data) for another input not used as training data with realistic uncertainty associated with those outputs. Bayesian

emulation typically requires fewer executions of a model to produce a set of training data from which the emulator can be developed when compared to a Monte Carlo based method such as ABC (O'Hagan [63] suggests 100s of runs for emulation vs many thousands for MC). This comes from the efficiency gained in the emulation process by extrapolating the data obtained from each training point to the surrounding region [63].

Note that the primary objective of Bayesian emulation is not to determine a distribution of parameters that match a given set of observations, but rather to come up with a cheaper alternative to the simulation. This means that once the emulator is developed, additional steps would be required to ascertain what parameters would produce results close to the observations. This could potentially revert back to a rejection sampling method, such as ABC, using the emulator instead of the simulator to produce each approximate realization of the data needed to compare to the reference observations. Additionally, optimization techniques could benefit from the emulator by allowing the designer to tweak many parameters and using the emulator to quickly realize the outcome of these modifications. The emulator outputs would then provide feedback to the optimization process. It is clear, though, that creating the emulator is an extra unnecessary step if the simulations are cheap enough to run many thousands of times when trying to ascertain parameters. In this dissertation, therefore, emulation was deemed unnecessary as simulation costs were kept low allowing the base simulator to be utilized hundreds or thousands of times in the examples that follow.

1.4 Overview and Contribution of this Dissertation

To date, ABC has been used primarily for biological problems [66–70] and is becoming more widely used in other fields such as the geophysical sciences [71–76], but is relatively unexplored for engineering systems except for a few applications to structural engineering [77, 78]. Several studies have investigated parameter estimation for engineering problems using Bayesian methods [9–12], although such approaches require knowledge of a likelihood function. Calculating the likelihood function in these applications requires assumptions that have varying degrees of accuracy depending

on the source of measurement data and the complexity of the model.

Thus, instead the approaches that have been applied in the past, ABC is investigated herein to allow more robust, physics-based simulations to accomplish parameter estimation. This dissertation presents an overview of work completed to introduce ABC to the engineering community through a variety of applications. The initial application in Chapter 3 is focused on the logistics equation, analogous to a zero-dimensional Navier Stokes equation, that is intended to serve as a preliminary test of the ABC approach in a fluids-relevant problem. Next, a two-dimensional (2D) turbulent buoyant jet is demonstrated using DNS and its parameters are estimated using LES. A similar 2D LES case, but now with a forced buoyant helium plume, is used to predict inlet composition based on comparisons with experimental measurements of puffing frequency from Cetegen *et al.* [3]. In Chapter 4, several industrially relevant applications are explored, including a turbulent buoyant jet with unknown temperature mean and variation, a turbulent buoyant jet with a rotating cylinder above it with an unknown rate of rotation and an inlet with unknown velocity and species composition. Finally, in Chapter 5, we explore the use of ABC to predict boundary and internal conditions for a turbulent buoyant jet 3D LES using experimental measurements of temperature made using laser absorption spectroscopy. This dissertation demonstrates how effectively ABC works as a tool for correctly estimating parameters in each of these cases, and thus provides evidence that ABC could be applicable in a wide variety of engineering applications.

The following papers and conference presentations contain work I have contributed to (as either a first author or a contributing author) while working on this dissertation:

Journal Publications:

(1) J. D. Christopher, N. T. Wimer, C. Lapoint, T. R. S. Hayden, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter estimation for complex thermal-fluid flows using approximate bayesian computation. **Physical Review Fluids**, 2018 *(in review)*

(2) J. D. Christopher, D. J. Petrykowski, T. R. S. Hayden, C. Lapointe, N. T. Wimer, S. P. Nigam, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter estimation in turbulent

simulations using approximate bayesian computation with laser spectroscopy. **Experiments in Fluids**, 2018 *(in prep)*

(3) T. R. S. Hayden, D. J. Petrykowski, A. Sanchez, S. P. Nigam, C. Lapointe, J.D. Christopher, N.T. Wimer, A. Upadhye, M. Strobel, P.E. Hamlington, and G.B. Rieker. Characterization of oh, h 2 o, and temperature profiles in industrial flame treatment systems interacting with polymer films. **Proceedings of the Combustion Institute**, 2018

(4) T. R. S. Hayden, N. T. Wimer, C. Lapoint, J. D. Christopher, S. P. Nigam, A. Upadhye, M. Strobel, P. E. Hamlington, and G. B. Rieker. Characterization of the buoyant jet above a catalytic combustor uinsg wavelength modulations spectroscopy. **International Journal of Heat and Mass Transfer**, 2018 *(in prep)*

Conference Publications:

(1) J. D. Christopher, D. J. Petrykowski, T. R. S. Hayden, C. Lapointe, N. T. Wimer, S. P. Nigam, I. Grooms, P. E. Hamlington, and G. B. Rieker. Parameter estimation using wavelength modulation spectroscopy temperature measurements and approximate bayesian computation. In **Optics and Photonics for Energy and the Environment**. Optical Society of America, 2018 *(in review)*

(2) J. D. Christopher, N. T. Wimer, T. R. S. Hayden, C. Lapointe, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter Estimation for a Turbulent Buoyant Jet Using Approximate Bayesian Computation. **AIAA Paper**, AIAA-2017-0531, 2017

(3) J. D. Christopher, C. Lapointe, N. T. Wimer, T. R. S. Hayden, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter Estimation for a Turbulent Buoyant Jet with Rotating Cylinder Using Approximate Bayesian Computation. **AIAA Paper**, AIAA-2017-3629, 2017

(4) T. R. S. Hayden, N. T. Wimer, C. Lapointe, J. D. Christopher, A. S. Makowiecki, P. E. Hamlington, and G. B. Rieker. Characterization of the output from a catalytic combustor

using wavelength modulation spectroscopy. In **47th AIAA Fluid Dynamics Conference**, page 4424, 2017

(5) T. Hayden, N. Wimer, C. Lapointe, J. Christopher, P.E. Hamlington, and G. B. Rieker. Characterization of a jet above a catalytic combustor using wavelength modulation spectroscopy. **Proceedings of the 10th U.S. National Meeting on Combustion**, 2017

(6) T. R. S. Hayden, N. T. Wimer, C. LaPointe, J. D. Christopher, S. Nigam, P. E. Hamlington, and G. B. Rieker. Wavelength modulation spectroscopy of OH radical in an industrial flame. In **Optics and Photonics for Energy and the Environment**, pages ETh2A–4. Optical Society of America, 2017

(7) C. Lapointe, J. D. Christopher, N. T. Wimer, T. R. S. Hayden, G. B. Rieker, and P. E. Hamlington. Optimization for internal turbulent compressible flows using adjoints. In **23rd AIAA Computational Fluid Dynamics Conference**, page 4115, 2017

(8) N. T. Wimer, C. Lapointe, T. R. S. Hayden, J. D. Christopher, A. Y. Poludnenko, G. B. Rieker, and P. E. Hamlington. Near-and far-field properties of high-temperature turbulent buoyant jets. In **47th AIAA Fluid Dynamics Conference**, page 4423, 2017

(9) O. A. Doronina, J. D. Christopher, C. A. Z. Towery, P. E. Hamlington, and W. J. A. Dahm. Autonomic Closure for Turbulent Flows Using Approximate Bayesian Computation. **AIAA Paper**, AIAA-2018-0594, 2018

(10) O. A. Doronina, C. A. Z. Towery, J. D. Christopher, I. Grooms, and P. E. Hamlington. Turbulence Model Development Using Markov Chain Monte Carlo Approximate Bayesian Computation. **AIAA Paper**, AIAA, 2019 *(submitted)*

Conference Presentations:

(1) J. D. Christopher, N. T. Wimer, C. Lapointe, T. R. S. Hayden, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter estimation for a pulsating turbulent buoyant jet using approximate Bayesian computation. **Bulletin of the American Physical Society**, 2017

(2) J. D. Christopher, N. T. Wimer, T. R. S. Hayden, C. Lapointe, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter Estimation for a Turbulent Buoyant Jet Using Approximate Bayesian Computation. **Bulletin of the American Physical Society**, 61, 2016

(3) O. Doronina, J. D. Christopher, P. E. Hamlington, and W. J. A. Dahm. Autonomic closure for turbulent flows using approximate Bayesian computation. **Bulletin of the American Physical Society**, 2017

(4) S. P. Nigam, C. Lapointe, J. D. Christopher, N. T. Wimer, T. R. S. Hayden, G. B. Rieker, and P. E. Hamlington. Flame structure and dynamics for an array of premixed methane-air jets. **Bulletin of the American Physical Society**, 2017

(5) N. Wimer, C. Lapointe, T. R. S. Hayden, J. D. Christopher, G. B. Rieker, and P. E. Hamlington. Effects of exit variability on near-field statistics for turbulent buoyant jets. In **Bulletin of the American Physical Society**, 2016

(6) C. Lapointe, N. T. Wimer, T. R. S. Hayden, J. D. Christopher, G. B. Rieker, and P. E. Hamlington. Scaling analysis of temperature variability between a rotating cylinder and a turbulent buoyant jet. In **Bulletin of the American Physical Society**, 2016

Chapter 2

Approximate Bayesian Computation (ABC) Methodology

Even for real-world flow and combustion problems where the underlying governing equations can be solved with a high degree of accuracy (i.e., assuming a physically and numerically accurate computational approach), the reliability and predicative capability of numerical simulations are heavily determined by the accuracy with which boundary, initial, and geometric conditions are represented. Fluid and material properties such as transport and heat transfer coefficients must also be accurately represented in order to obtain simulation results that correspond closely to reality. In many cases, simulations are designed to provide good agreement with data from an experiment, and some of the necessary conditions and properties may be known with high accuracy based on the setup of the experiment and other physical constraints. However, there are often many other parameters that are *not* known with sufficient accuracy, to the extent that, despite our best efforts, a computational simulation may model a fundamentally different problem than that studied experimentally.

In order to overcome this knowledge gap and improve simulation accuracy, approximate Bayesian computation (ABC) is used in this dissertation to reliably estimate boundary conditions in simulations of a high-temperature turbulent buoyant jet in various configurations. This work is one of the first attempts to apply ABC to a complex engineering-relevant compressible flow problem.

2.1 Approximate Bayesian Computation Background and Overview

The ABC method allows unknown parameters in a model or simulation to be determined by comparing simulation results with the full data or statistics from experimental, or other "reference", observations. In addition to these optimal parameter estimates, ABC also provides meaningful estimates of the uncertainty associated with each estimated parameter value. For additional details regarding the choice to use ABC over other data-driven techniques, see Chapter 1.

The ABC framework is a relatively new approach for linking reference data to physical and model parameters [2, 58, 59, 77, 84, 85, 98]. ABC was first conceptualized by Rubin in 1984 [99] as a parameter estimation method for models that are sufficiently computationally inexpensive that multiple simulations can be performed, but that are intractable with regards to calculating likelihood functions. More than a decade passed, however, before the technique was applied to problems of practical interest in the population genetics community [100, 101]. Since these early implementations of ABC, many papers have reviewed ABC efforts and suggested improvements to the method (e.g., see [58–60, 98]). ABC has increased in popularity in biological applications [66–70, 102, 103], in particular, and is becoming more widely used in other fields such as the geophysical sciences [71–76, 104]. ABC has also recently spread to less traditional domains, including financial modeling [105] and psychology [106]. Although ABC has been used recently in the context of chemical kinetics modeling for combustion [28] and structural engineering [77, 78], it remains relatively unexplored for most engineering applications, particularly complex thermal-fluid systems, and this dissertation provides one of the first attempts to apply ABC in an engineering context.

From a fundamental perspective, the ABC method produces a posterior distribution of unknown parameters given some reference data. This occurs by comparing observations or summary statistics from the reference data to corresponding data or summary statistics from lower-cost physics-based 'model' simulations (e.g., simulations with coarse spatial resolutions or lower-order numerical schemes). The model simulations are repeated many times, with each simulation using

parameters drawn from a prior distribution. The prior distribution is the best guess at the span of the unknown parameter space; it must be wide enough to contain the true parameter values, but narrow enough to keep the task computationally manageable. Parameter values are retained if the model data or summary statistics are similar to the reference observations. Once many such candidates are obtained, a posterior distribution is formed, providing estimates for the most likely parameter values as well as their uncertainties, given the reference data.

Bayes' theorem, $P(\theta|D) = P(D|\theta)P(\theta)/P(D)$ [see also Eq. (1.1)] provides the basis for the ABC technique [2, 58, 59]. This theorem gives the probability, $P(\cdot)$, of obtaining different values of an unknown model parameter, θ, given the observations or reference data, D. The model parameter, θ, could be a boundary condition, material property, transport property, model parameter (e.g. a turbulence model coefficient [5] or a parameter needed to adjust another model such as a heat source term), or another physical property, for example. In this dissertation, we will investigate how to solve for various parameters, including inlet temperature and velocity means (Chapters 3, 4, and 5), inlet temperature and velocity variations (Chapters 3 and 4), cylinder rotational velocity (Chapter 4), inlet composition (Chapters 3 and 4), and parameters to model the addition of heat into the domain from combustion (Chapter 5).

The probability density function of the unknown parameter, independent of any observations, is known as the 'prior' and is denoted $P(\theta)$. The set of observations is denoted D and could be, for example, a time series of 10,000 temperature, velocity and/or concentration measurements at one or many points in space. Note that the data can also be represented by its statistics, $\mathcal{S}$ [e.g., the mean value from a longer time series of data, the probability density function (pdf) or a power spectral density (psd) estimate]. The data may come from experiments or, alternatively, another implementation of a model which may be more refined. The probability density function of the data is $P(D)$; this is obtained through analyzing experimental or other reference results, but is often neglected in the equation since it can be treated as a normalization constant. The goal of Bayes' theorem is to solve for $P(\theta|D)$, also known as the Bayesian posterior (alternatively, it may be called the "true" Bayesian posterior or the "full" Bayesian posterior because it does not contain

the approximations made here); this provides the estimate for the unknown parameter conditioned on the observed data. Lastly, the probability of the observed time series of data, D, occurring given a set of parameter values, θ, $P(D|\theta)$, is known as the likelihood function. The likelihood function is often difficult to obtain, either because of analytical intractability or due to the high cost of its computational calculation. Indeed, in a complex fluid mechanics model, one can safely assume that the likelihood function will not be available in the majority of cases.

The ABC method aims to calculate an approximation to the full Bayesian posterior, $P(\theta|D)$ without explicitly evaluating the likelihood function. The ABC posterior distribution is denoted $P_{ABC}(\theta|D) = P(\theta|\delta(D, \hat{D}) < \varepsilon)$ (this is also what is being referred to in subsequent uses of the word "posterior"). The ABC posterior distribution is not the same as the full Bayesian posterior, but it approximates it whenever $\mathcal{S}$ is a set of 'sufficient' statistics as described below. ABC is most relevant and useful when there is no available likelihood function. Instead of requiring the exact likelihood function $P(D|\theta)$, the ABC method instead replaces the likelihood function with an *ad hoc* distance function, $\delta(\cdot, \cdot)$ to compare statistics from each simulation, $\hat{\mathcal{S}}$, with statistics from the reference observation, $\mathcal{S}$.

In order to generate each simulation used in the comparison, the method first samples values of the parameter θ from the anticipated parameter space [i.e., from the prior, $P(\theta)$]. For example, the prior could be a uniform distribution of each parameter or a Gaussian distribution. The choice of prior is based, as the name implies, on prior information a practitioner has about the possible values a parameter might take. The posterior will mainly reflect the observational data as long as the prior is sufficiently uninformative. A prior that is very dissimilar to the posterior will increase computational costs as many of the simulated parameter values will be rejected. Once a set of parameters, θ, are chosen, the method runs the model forward using those parameters to come up with a set of synthetic observations, $\hat{D}$ (whose statistics are denoted $\hat{\mathcal{S}}$). The ABC method then compares statistics from each approximate realization, $\hat{\mathcal{S}}$, with the experimental statistics, $\mathcal{S}$, and either keeps the parameters used to produce that realization, or rejects the parameters depending on a predefined distance function, $\delta(\cdot, \cdot)$, and associated threshold, ε. The distance function is simply

a metric to compare simulation results to an observation; the distance function must be a metric (i.e., symmetric, positive, and satisfy the triangle inequality) but is otherwise fairly arbitrary. The distance function could, for instance, take the absolute value of the difference between the mean of the observations to the mean of the simulation results generated with a set of parameters. Another common choice for metrics when comparing probability distributions is the Hellinger distance; this will be used extensively in Chapter 3.2.

When performing the comparison, the statistics (i.e., $\mathcal{S}$) must accurately represent the data. Ideally, ABC should be conducted with 'sufficient' statistics that represent all of the information about θ contained in the data, D [58]. If the statistics are 'sufficient' to describe the data, then ABC approximates the full Bayesian posterior. Put another way, if the statistics are 'sufficient,' then the difference between the full Bayesian posterior, $P(\theta|D)$, and the ABC posterior, $P_{ABC}(\theta|D)$, is controlled by the threshold, ε (i.e., the error goes to zero with ε). Often, however, 'sufficient' statistics are not available and lesser summary statistics must be used; these statistics must still adequately describe the data and are therefore useful when comparing simulations and the reference observations. Choosing an appropriate statistic to analyze the flow field is a complicated task because many statistics will not contain the desired information about the flow field. In other words, if one changes the parameter of interest, there may be only a negligible change in the statistic since it has low identifiability [107]. Thus, choosing an appropriate and informative statistic is a key task in successfully implementing ABC algorithms.

Once a comparison is made between each simulation and the reference observations, the resulting distance is compared to a threshold value. If the distance is less than the threshold, the parameter is considered viable and it is added to the candidate pool. However, if the distance exceeds the threshold, then the parameter is discarded since it did not produce a realistic model outcome compared to the available observation. Using all of the accepted parameter values, one can construct a distribution of likely parameters that would result in the given observations; this distribution is representative of the true parameters. The distribution thusly created, $P_{ABC}(\theta|D)$, which can now be equivalently denoted as $P_{ABC}(\theta|\delta(\mathcal{S},\hat{\mathcal{S}}) < \varepsilon)$, is known as the posterior and is

the item one sought in the parameter estimation.

2.2 Approximate Bayesian Computation Algorithm

The basic idea underlying the ABC approach is that, based on system intuition, physical constraints, previous experience, etc., a researcher typically has a range of probable values for a parameter. These upper and lower limits bound the cases for which simulations must be run. The 'prior' set of parameters must be large enough to ensure that it contains the 'true' solution, however more simulations are required as the prior becomes wider, thereby significantly increasing the computational cost. Once the reference case and a subsequent case, with parameter values chosen from the prior distribution are complete, the next step is to study the flow fields of each case. The statistics of each simulation are compared to statistics from the reference case. If the results are similar, the simulated parameters are stored as a possible candidate values. If the results are not similar, then the set of parameters used for that simulation is discarded. Another simulation is then run with a new set of parameters and the process repeats until a sufficiently high number of accepted parameter values exists from which conclusions can be made about the results.

This ABC algorithm, known as method D from Marjoram *et al.* [2], and herein as simply algorithm 1 is summarized as follows:

(1) Generate parameter θ from the prior distribution $P(\theta)$.

(2) Simulate approximate data $\hat{D}$ using parameter θ, and compute the corresponding statistics $\hat{\mathcal{S}}$.

(3) Calculate the distance $\delta(\mathcal{S}, \hat{\mathcal{S}})$ between a reference statistic $\mathcal{S}$ and statistic from the simulator $\hat{\mathcal{S}}$.

(4) Accept θ if $\delta \leq \varepsilon$ (where ε is the 'rejection distance'), and return to step 1 until sufficient θ values have been accepted.

The distribution of accepted θ values comprise the ABC posterior, $P_{ABC}(\theta|D)$. An overview of the

algorithm is shown in Figure 2.1.

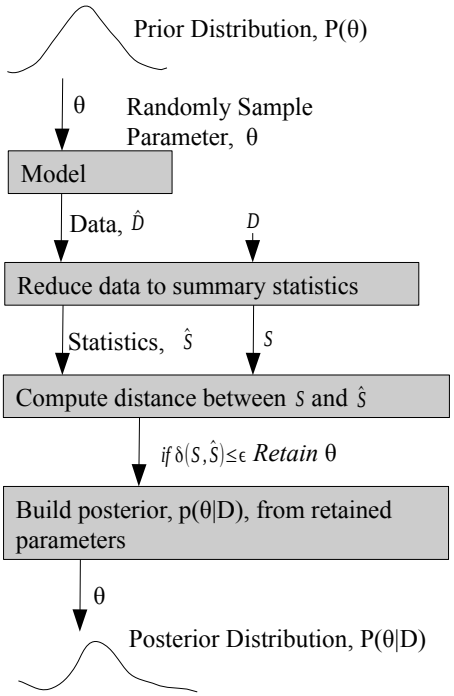


Figure 2.1: Algorithm depicts the steps required for the basic rejection algorithm described in Section 2.2

2.3 Selection of Rejection Distance, ε

In ABC, once the summary statistic, $\mathcal{S}$, is chosen and a distance calculated between each simulation and the reference observation, the next task is to determine how small the 'rejection distance,' ε, should be to ensure that each result is sufficiently close enough to the 'true' case in order to be accepted. Recall that as ε goes to infinity the ABC procedure provides no additional information compared to the initial prior distribution used for the Bayesian inference. Also, assuming sufficient statistics, as ε decreases to zero the posterior converges to the full Bayesian posterior. Hence, the preference for smaller ε values (to produce better posteriors) must be balanced against the increasing computational cost required to reduce the Monte Carlo error [60].

The task of choosing an appropriate value for ε can be addressed by a variety of techniques,

often relying on expert judgment to balance the spread in accepted parameter values with the number of accepted samples. In general, as ε decreases from infinity to zero the ABC method goes from accepting all of the simulations (i.e., the posterior would match the prior) to accepting the one "true" simulation, which may not exist in most instances and hence zero cases would be accepted since there is some finite difference between the best simulated case and the reference case when continuous variables are used. As a result, often a relatively small (nonzero) value of ε is chosen based on the behavior of the posterior as ε decreases. A more specific approach to select ε requires a predetermined percentage of samples being accepted [66, 69] or choosing multiple ε values and reporting the results simultaneously [2,58,59,104]. Marin *et al.* [59] argue that a smaller ε is better generally, but the size is largely determined by the available computational resources. In subsequent chapters we choose values for ε based on the simplest approach of balancing the precision in the solution (low standard deviation) against having sufficient confidence in the results (high number of accepted samples). Since the 'true' parameter values are known, this is a reasonable approach.

Once the true value of the parameter is unknown, however, it becomes less clear how well this approach converges. Therefore, in Section 4.3 of Chapter 4, we introduce a novel approach of choosing an ε value that minimizes the confidence interval of the posterior found by sweeping across a wide range of ε values. The confidence interval is proportional to the standard deviation of accepted parameter values divided by the square root of the number of accepted values. The range of ε values must be wide enough so that the smallest value accepts very few of the proposed parameter values while the largest ε accepts nearly all of the proposed parameter values. For very small values of ε, the majority of the cases are rejected, resulting in a very small standard deviation of the posterior. As ε increases, the number of accepted samples increases, initially reducing the confidence interval width (assuming a unimodal posterior distribution). As ε continues to increase, however, the standard deviation starts to increase faster than can be compensated for by the larger sample size (recall that the sample size is to the 1/2 power), so the confidence interval width increases until all of the simulations are accepted, at which point it remains constant.

The minimum confidence interval width observed corresponds to a ε value that strikes a

balance between having high precision in the posterior (i.e., low standard deviation) and high confidence in the posterior (i.e., a large number of accepted samples). Note that this approach is not used exclusively in this dissertation because the standard deviation is susceptible to large, fast initial growth when the first few samples are not tightly grouped. In this case, the "best" ε is determined as the value required to only accept a single case. This is not desirable, however, as a single case provides little confidence and little information about the possible distribution of likely parameters.

2.4 Markov-chain Monte Carlo ABC

Markov-chain Monte Carlo approximate Bayesian computation (MCMC-ABC) is another algorithm that can be used to find unknown parameters, and includes an additional level of complexity. In particular, the MCMC-ABC algorithm is similar to method D from Marjoram *et al.* [2] outlined above, but applies an additional constraint to generate each successive parameter value (instead of choosing them all independently at random). The MCMC-ABC algorithm, known herein as algorithm 2 and presented as method F in Marjoram *et al.* [2] is as follows:

(1) If now at θ propose a move to $\hat{\theta}$ according to the transition kernel $q(\theta \to \hat{\theta})$.

(2) Generate $\hat{D}$ using model $\mathcal{M}$ with parameters $\hat{\theta}$.

(3) If $\delta(\mathcal{S}, \hat{\mathcal{S}}) \leq \varepsilon$, go to step 4, and otherwise stay at θ and return to step 1.

(4) Calculate

$$h = h(\theta, \hat{\theta}) = \min\left[1, \frac{P(\hat{\theta})q(\hat{\theta} \to \theta)}{P(\theta)q(\theta \to \hat{\theta})}\right] \tag{2.1}$$

(5) Accept $\hat{\theta}$ with probability h and otherwise stay at θ, then return to step 1.

The MCMC-ABC algorithm has the benefit of utilizing past acceptable parameters to inform its decision about the next parameter to choose. This approach can significantly reduce the number of required simulations by more efficiently sampling the parameter space and focusing on regions with higher probability of acceptable parameters. Instead of randomly or blindly choosing the next

parameter from the prior distribution, as is done in algorithm 1, the MCMC-ABC method uses an accepted parameter as a starting point and then moves a specified distance away from that accepted parameter using the transition kernel, q, described in step 1. A transition kernel is simply some function that describes how to move to the next parameter given a starting parameter. An example transition kernel is to fit a Gaussian distribution with the current parameter as the mean and a prescribed variance defined within the context of the particular problem; the next parameter is then determined by generating a random parameter value from this Gaussian distribution.

The next difference of the MCMC-ABC algorithm is step 4; this equation calculates the product of the prior distribution evaluated at the new parameter times the probability of moving from that new parameter to the original parameter, all divided by the product of the prior evaluated at the original parameter times the probability of moving from the original parameter to the new parameter. For a Gaussian transition kernel, the probability of going from one parameter to the next is the same as moving in the opposite direction, since each probability distribution function is centered on the starting parameter and has the same variance. Thus, in the case of a Gaussian transition kernel, these terms cancel out and the driving function becomes the prior; this means if a new parameter is toward the center of a Gaussian prior distribution, for example, then the new parameter is more likely to be accepted. Further, note that if the prior is uniformly distributed, then the contribution from the evaluation of the prior will cancel out as well. Thus, with a uniform prior and a Gaussian transition kernel, step 4 will always return $h = 1$, meaning that step 5 will automatically accept any new parameter that passes the distance criteria of step 3 [2]. Note that, in this dissertation, the MCMC-ABC algorithm is only used in Section 3.1.2 (i.e., not in the other chapters) due to its inherent serial nature; the cost of subsequent simulations necessitates parallel computation to achieve timely results.

2.5 ABC Summary

In summary, ABC is a versatile tool that can be used to develop estimates for parameter values, including uncertainties in those values, given a set of observations. The method first samples

a large set of parameters from the likely parameter space and runs a model forward using one of those parameters. Next, ABC compares the data from the simulation against the reference observations. If the simulation and observation are close, then ABC keeps the parameters; if the results are not close, then it rejects the parameters. This process continues until many parameters have been modeled and an adequate number have been accepted. These accepted parameters make up the desired posterior distribution. A central tendency of the posterior, such as the mean, mode, or median, can be used to obtain a single estimate for the unknown parameter, and a measure of the posterior width, such as the standard deviation, can be used to estimate the uncertainty in the parameter value.

Chapter 3

Demonstration of the ABC Approach

As mentioned in Chapter 1, approximate Bayesian computation (ABC) has, to date, been mostly applied in fields outside of engineering, and has yet to be applied to parameter estimation for systems involving complex thermal-fluid flows. This chapter will provide an illustrative example of how ABC can be of great utility for fluid dynamics and combustion related engineering. First the logistics equation, described in Section 3.1 as a zero-dimensional Navier Stokes equation, will serve as an illuminating example for determining accurate values for unknown parameters. Then, a two-dimensional turbulent buoyant jet will serve as a canonical flow case for testing the accuracy of ABC by determining unknown inlet velocity conditions. Lastly, experimental measurements will be used with this turbulent buoyant jet to predict inlet composition.

3.1 Approximate Bayesian Computation for the Logistics Equation

As an initial case to test the ABC method, the logistics equation has been examined. This equation, given as

$$v_{t+1} = 4\mu v_t(1 - v_t)\,, \tag{3.1}$$

has been chosen for its complex behavior and for its relevance to fluid mechanics. Although Eq. (3.1) is deterministic, it has stochastic-like behavior for μ values greater than approximately 0.89 (i.e., a small change in μ leads to an entirely different response) and it is nonlinear. Additionally, the logistics equation can also be considered a zero-dimensional (0D) Navier-Stokes equation with μ

playing the role of Reynolds number, and is thus a representative equation to investigate fluid mechanics behavior.

The logistics equation has several regions of behavior corresponding to different μ values. In region 0, for μ less than 1/4, the solution always goes to zero in the limit as $t \to \infty$. In region 1, for $1/4 < \mu < 3/4$, the solution goes to a specific value, v_∞, for each μ selected. In region 2, $3/4 < \mu < \mu_\infty$ (where $\mu_\infty \approx 0.89248$) the equation's solution first oscillates between 2 values and then increasingly more values. This behavior occurs until finally, for $\mu > \mu_\infty$ the solution is non-periodic, highly sensitive to initial conditions, covers the full range $[0, 1]$, and has long-term behavior independent of the initial conditions.

An ABC algorithm was implemented to predict an unknown μ value. This was performed by first simulating a solution to the logistics equation using a μ value sampled between 0 and 1; this first parameter chosen will serve as μ_{ref} for subsequent simulations to try to find. The mean and variance of the resulting solution were then stored. Next, assuming the original μ was unknown, a new μ value was randomly chosen and a simulation was run. Its resulting statistics were compared to the stored values coming from the reference μ value, μ_{ref}. If the results were similar, the new μ value was stored as a possible candidate μ value; if the results were not similar, that μ value was discarded. Another simulation was then run with a new, randomly chosen μ value and the process repeated until a sufficiently high number of accepted μ values existed to draw conclusions about the results. This first ABC algorithm, known as method D in Marjoram *et al.* [2], is presented in Chapter 2, and repeated here for convenience:

(1) Generate parameter θ from the prior distribution $P(\theta)$.

(2) Simulate data $\hat{D}$ from model $\mathcal{M}$ with parameter θ, and compute the corresponding statistics $\hat{\mathcal{S}}$.

(3) Calculate the distance $\delta(\mathcal{S}, \hat{\mathcal{S}})$ between $\mathcal{S}$ and $\hat{\mathcal{S}}$.

(4) Accept θ if $\delta \leq \varepsilon$ (where ε is the rejection distance), and return to step 1.

Note that the generic parameter, θ, that one seeks to find is μ when studying the the logistics equation.

A Markov chain Monte Carlo (MCMC) approach can be utilized to increase parameter acceptance rates [2] as discussed in Chapter 2. With an MCMC method, instead of choosing parameters θ independently to run each simulation, if a parameter is accepted then a new parameter is chosen with a value close to that accepted parameter. From many simulations a posterior distribution will emerge indicating which parameters (θ) are probable given the data (D); this is, again, the posterior, $P(\theta|D)$. The primary limitation of an MCMC approach is its serial nature requiring each simulation to complete before the next can begin. Indeed for more expensive simulations parallel computing is a key factor enabling their implementation in ABC. Though parallel MCMC algorithms using multiple chains have been proposed (see, e.g., [108–110]) here the focus is on techniques that are embarrassingly parallel to enable the user to scale with more processors as needed regardless of the length of the simulation time available (recognizing that many high performance computing resources are set up to allow many thousands of simulations to run in parallel but restrict reservation lengths to allow only a few simulations to run in sequence).

The MCMC-ABC algorithm, known herein as algorithm 2 and presented as method F in Marjoram *et al.* [2], as well as in Chapter 2, is repeated here for convenience as follows:

(1) If now at θ propose a move to $\hat{\theta}$ according to the transition kernel $q(\theta \to \hat{\theta})$.

(2) Generate $\hat{D}$ using model $\mathcal{M}$ with parameters $\hat{\theta}$.

(3) If $\delta(\mathcal{S}, \hat{\mathcal{S}}) \leq \varepsilon$, go to step 4, and otherwise stay at θ and return to step 1.

(4) Calculate

$$h = h(\theta, \hat{\theta}) = \min\left[1, \frac{P(\hat{\theta})q(\hat{\theta} \to \theta)}{P(\theta)q(\theta \to \hat{\theta})}\right] \tag{3.2}$$

(5) Accept $\hat{\theta}$ with probability h and otherwise stay at θ, then return to step 1.

The MCMC-ABC algorithm has the benefit of utilizing past acceptable parameters to inform its decision about the next parameter to choose. More details can be found in Chapter 2. Note that

in this dissertation the MCMC-ABC algorithm is only used in this section (i.e. not in the other chapters) due to its inherent serial nature; the cost of subsequent simulations necessitate parallel computation to achieve timely results.

3.1.1 Results of Logistics Equation: Random Parameter Generation

The initial case studied investigates the behavior of the logistics equation using the first algorithm above. Each execution of the logistics equation is run forward 1,000 steps so that the statistics have approximately converged. One million μ values are simulated to ensure sufficient accepted cases; this high number of cases helps fill in the resulting histograms of accepted values for μ.

In region 0, the solution always converges to 0 regardless of the μ value chosen. Thus, the method correctly predicts that $0 < \mu_{ref} < 1/4$, but cannot provide any additional specificity. In region 1, this method can easily predict values of μ_{ref} since, again, the solutions behave linearly. Note that decreasing the acceptance distance, ε, results in a narrower range of possible μ_{ref} values but decreases the number of accepted values potentially requiring more simulations to be run to maintain high confidence in the resulting parameter. Thus, the decision for which ε value to use must balance accuracy against computational burden. In region 2, the solution behavior is also entirely linear with the solution oscillating between two unique values for each μ chosen. Thus, this method is able to quickly determine the correct μ_{ref} value.

In region 3, the task of finding μ_{ref} becomes much more challenging due to the increasingly nonlinear and stochastic behavior of the logistics equation. However, for sufficiently small ε values for the mean and variance, the correct μ_{ref} value emerges. When higher values of ε are chosen (e.g., 0.01), the parameter space starts to include other values of μ that produce very similar statistics, as shown in Figures 3.1 and 3.2. In Figure 3.1 each of the highlighted points has a converged solution mean value (plotted on the vertical axis) within 0.01 of the solution's mean value of 0.5765 found with $\mu_{ref} = 0.989$. In figure 3.2 each of the highlighted points has a converged solution variance (plotted on the vertical axis) within 0.01 of the solution's variance of 0.0991 found with

$\mu_{ref} = 0.989$. Of the three points highlighted, only the right-most point ($\mu = 0.9898$) represents a solution coming from a μ value very close to μ_{ref} while the other two points are substantially (i.e., more than 1%) different μ values.

Next, Figure 3.3 is a histogram of the accepted parameters found by implementing algorithm 1 with 1 million values of μ randomly drawn between 0.85 and 1 (i.e., this assumes a uniform distribution of values in the prior with bounds closely aligned with region 3), with a rejection distance, ε, of 0.01. This result clearly demonstrates the inability of the procedure to correctly identify μ_{ref} with this relatively large ε value. Figure 3.4 shows a simulation with identical parameters except the rejection distance, ε, has been reduced to 0.001. With $\varepsilon = 0.001$, the two "false-peaks" have been eliminated. This narrowly confined solution correctly identifies the value of μ_{ref} with the mean of the histogram equaling 0.9886 (0.04% error), however it rejects the vast majority of the simulations accepting only 0.17% of the simulated μ values.

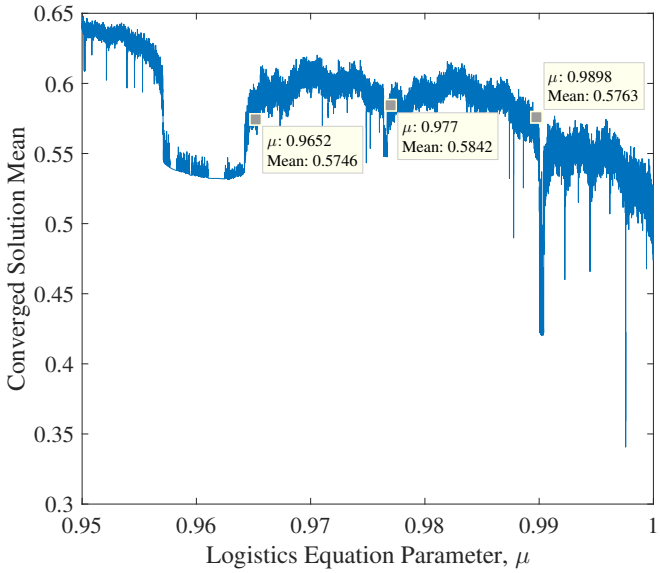


Figure 3.1: Converged solution mean of the logistics equation taken after 1,000 simulation steps. Note there are three μ regions (of which one value is shown per region) that result in mean values (and associated variance values shown in figure 3.2) within an ε value of 0.01; thus all three indicated μ values would be accepted with $\varepsilon = 0.01$.

3.1.2 Results of Logistics Equation: MCMC Parameter Generation

Note that since there is only one parameter of interest in the logistics equation, and because it is very inexpensive to compute, the basic ABC algorithm was sufficient to determine μ_{ref}. However for more complicated problems with higher-dimensional parameter space this may not be the case. Hence, the MCMC-ABC algorithm was next implemented for the logistics equation to understand how it complements the results. Following algorithm 2, a transition kernel $q(\theta \to \hat{\theta})$ was chosen to start at the current θ value then add a noise value drawn from a zero-mean Gaussian distribution with a specified standard deviation ($\sigma = 0.01$) (where the variance is $\sigma^2 = 0.01^2$). The value chosen for the transition kernel's variance is very important. A variance that is too narrow means the MCMC-ABC method will have a difficult time wandering around far enough to fully map out the parameter space. A variance that is too wide means the MCMC-ABC method will revert to $q(\theta \to \hat{\theta}) = q(\hat{\theta})$, meaning the next parameter value will essentially have no memory of the previous parameter. Next, the prior distribution in each case presented herein is Gaussian with a mean at 0.9652 and a standard deviation of 0.01. Figure 3.5 shows a histogram produced by a simulation run with the same conditions as above in Section 3.1.1, but with an MCMC-ABC method to determine each successive μ value instead of a random draw from a uniform distribution. The prior is centered at 0.9652 and hence values of μ closer to that mean are favored resulting in a histogram with a bias toward the $\mu = 0.9652$ peak instead of the right-most $\mu = 0.989$ peak. Figure 3.6 shows an identical procedure but with a reduced rejection distance, ε. Note that this results in a very dense histogram around $\mu_{ref} = 0.989$; the mean of the accepted parameters is 0.9885 (0.05% error). This highlights a benefit of using MCMC-ABC in that it fills in the parameter space close to the reference parameter(s).

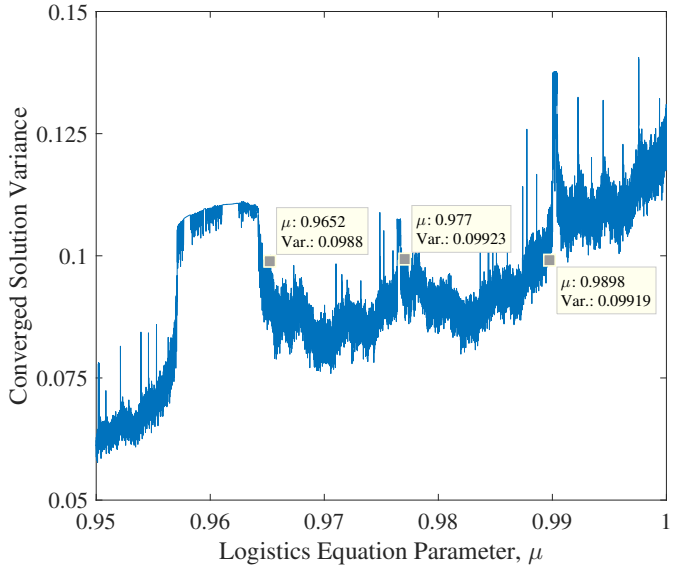


Figure 3.2: Converged solution variance of the logistics equation taken after 1,000 simulation steps. Note there are three μ regions (of which one value is shown per region) that result in variance values (and associated mean values shown in figure 3.1) within an ε value of 0.01; thus all three indicated μ values would be accepted with $\varepsilon = 0.01$.

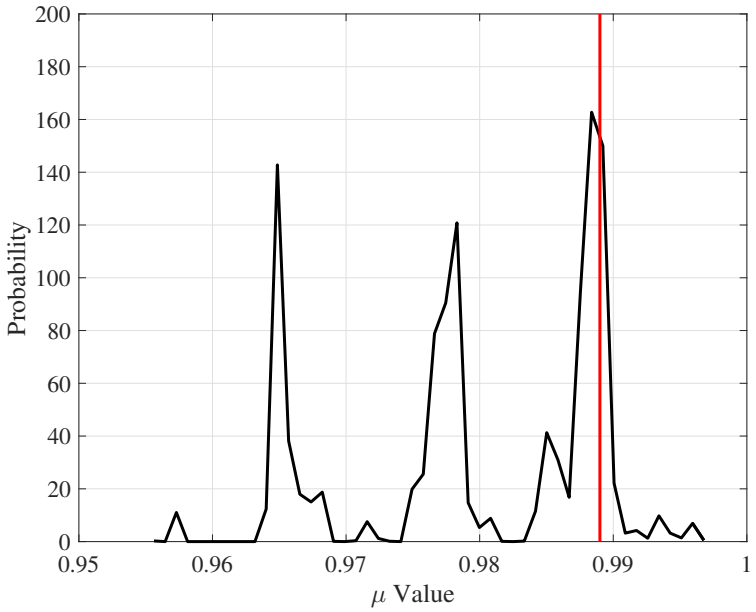


Figure 3.3: Histogram showing the number of accepted simulations versus values of μ. Simulation was run with 1 million values of μ randomly distributed between $0.85 - 1$, 1,000 steps, $\varepsilon = 0.01$. Reference μ value shown in red.

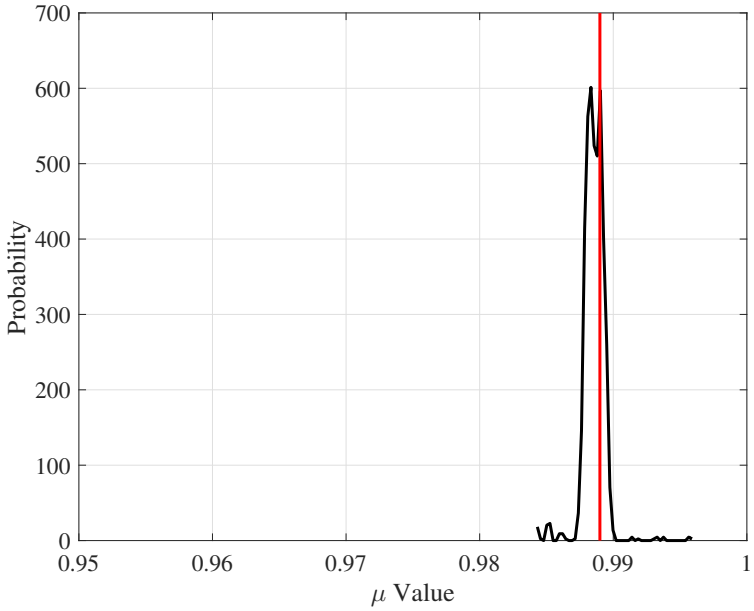


Figure 3.4: Histogram showing the number of accepted simulations versus values of μ. Simulation was run with 1 million values of μ randomly distributed between $0.85 - 1$, 1,000 steps, $\varepsilon = 0.001$. Reference μ value shown in red.

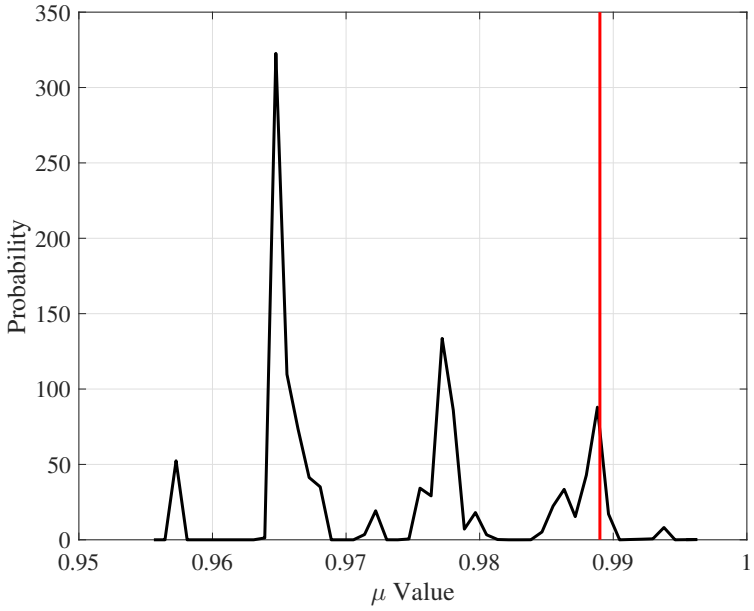


Figure 3.5: Histogram showing the number of accepted simulations versus values of μ. Simulation was run with 1 million values of μ chosen according to MCMC-ABC procedure, starting at $\mu = 0.9652$, 1,000 steps, $\varepsilon = 0.01$. Reference μ value shown in red.

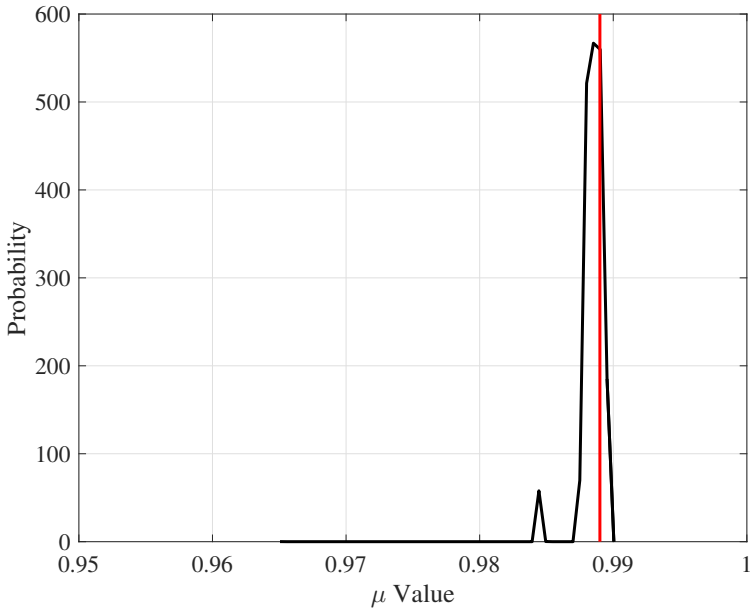


Figure 3.6: Histogram showing the number of accepted simulations versus values of μ. Simulation was run with 1 million values of μ chosen according to MCMC-ABC procedure, starting at $\mu = 0.9652$, 1,000 steps, $\varepsilon = 0.001$. Reference μ value shown in red.

3.2 Parameter Estimation in a Turbulent Buoyant Jet Using ABC

This section introduces ABC as an effective tool for estimating parameters in fluid mechanics research, especially with an engineering focus. Building upon the initial success with the logistics equation [84], this work demonstrates ABC is effective when the reference data comes from a either an LES case or a DNS case that uses a different mesh and no modeling. This section is closely tied to a paper under review with Physical Review Fluids [79].

Approximate Bayesian computation (ABC) is a data-driven technique that uses many low-cost numerical simulations to estimate unknown physical or model parameters (e.g., boundary conditions and material properties), as well as their uncertainties, given reference data from real-world experiments or higher-fidelity numerical simulations. In this study, ABC is used to estimate unknown parameters in complex thermal-fluid flows, and the technique is demonstrated on a periodically forced high-temperature jet and a steadily forced helium-air plume. In the first case, computational reference data are used to assess the accuracy of ABC when estimating the frequency, amplitude, and mean of the periodic velocity forcing at the jet inlet. These tests show that ABC provides accurate and reasonably certain estimates of inflow parameters even when the model simulations imperfectly represent the physics underlying the reference data. These tests also show that measurements far from the inlet can be used to perform the estimation, and that temperature measurements can be used to infer velocity inflow parameters. For the second case, ABC is used to estimate the inlet Richardson number and its uncertainty given experimental measurements of the Strouhal number within the plume. Once again, the approach is able to accurately estimate unknown parameters with relatively low uncertainty. As a result, ABC is shown to be a versatile technique for estimating unknown physical parameters when knowledge of a real-world system is limited or incomplete.

In the present study, approximate Bayesian computation (ABC), a data-driven Bayesian technique adopted from the biological and geophysical sciences, is used to estimate unknown parameters for complex, turbulent thermal-fluid flows. The power of ABC lies in the fact that far

fewer simulations are required than in full Bayesian analyses since ABC does not require a likelihood function, thus permitting the use of physics-based models. Rather than attempting to match the observational data at all locations and times, as in full Bayesian analyses, the ABC method develops an approximate posterior distribution for unknown parameters through comparison of summary statistics from the reference data and model simulations. A wide variety of reference data can be used to drive the estimation, including measurements that are only indirectly related to parameters of interest. Because the technique naturally provides posterior distributions for unknown parameters, Bayesian confidence intervals can be obtained along with parameter estimates. Once parameter estimates are obtained, they can either be used in higher-fidelity numerical simulations of the same system, or they can be considered as part of the description of the real-world system itself (with appropriate uncertainty caveats).

This section outlines the estimation of inflow boundary conditions for two compressible turbulent flows: (*i*) a periodically forced high-temperature jet, and (*ii*) a weakly but steadily forced helium-air plume. Both cases are two-dimensional (2D), but are physically complex due to compressibility and the corresponding coupling between temperature, density, buoyancy, and forced advection. The first case is intended to demonstrate the utility and breadth of ABC by using computational reference data to estimate the frequency, amplitude, and mean of the forced inflow. Since these parameters and the underlying physical model are precisely known for the computational data, the success of the ABC approach can be quantified. The second case is a demonstration of ABC using experimental measurements in an engineering context.

3.2.1 Demonstration Using Computational Reference Data

As a demonstration of ABC, inflow parameters for a periodically forced high-temperature turbulent buoyant jet [111–115] are estimated using computational reference data. This case is examined due to the simplicity of the geometry combined with the complexity of the high-temperature unsteady compressible flow physics. Large eddy and direct numerical simulations (LES and DNS, respectively) are used as reference data, and model simulations are performed using LES.

The use of computational reference data allows the success of ABC to be assessed when the physics underlying the reference data, as well as all system parameters, are known exactly. Three questions, in particular, are addressed: (*i*) How accurately can ABC estimate unknown parameters when the reference data physics are exactly reproduced by the model simulations? (*ii*) How accurately can ABC estimate unknown parameters using model simulations that imperfectly represent the physics of the reference case? (*iii*) How accurately can ABC estimate unknown parameters using reference data that are only indirectly connected to the parameters of interest?

The first question represents a 'best case scenario' where there is zero model error and is examined using both model and reference data from LES. The second question pertains to a more realistic application of ABC where the physics governing the reference data are not exactly represented by the model simulations. This question is examined here using LES model simulations and DNS reference data, and is further addressed in Section 3.2.2 using experimental reference data. The third question addresses the identifiability of unknown parameters using different types of reference data, and is examined by performing ABC at different heights above the inlet using either velocity or temperature measurements.

It should be noted that the test where both the model and reference data are obtained from LES most closely resembles an observing system simulation experiment (OSSE), which is a common technique used for the validation of data assimilation methods [116,117]. In the present context, the OSSE represented by this test allows the choices of prior, summary statistic, distance function, and rejection distance to be evaluated in the absence of model error. In this sense, these tests indicate whether the ABC approach can ever be expected to succeed, and to what extent. The following analysis shows that, even in such a 'best case scenario', the ABC parameter estimation does not perfectly recover the true parameter values due to the presence of stochasticity in the model and reference data. That is, the model and reference data are never in perfect local and instantaneous agreement because each of the simulations are initialized with different random temperature fields (even when using identical boundary parameter values), and thus the true parameter values are only ever recovered approximately, even when the model error is zero.

3.2.1.1 Physical and Numerical Setup

The reference buoyant jet has a sinusoidally varying inlet velocity with a frequency of 4 Hz, an amplitude of 0.2 m/s, and a mean of 0.5 m/s. The inflow velocity is spatially uniform across the inlet. The physical domain is 1×2 m and the inlet is 0.1 m wide. The ambient and inlet temperatures are 300 K and 1700 K, respectively, resulting in substantial density differences and buoyancy-driven flow. Instantaneous fields of velocity magnitude (i.e., speed) and temperature are shown in Figure 3.7.

All of the computations were performed using the FireFoam solver [118] within OpenFOAM [119,120]. The simulations were restricted to 2D to minimize computational cost, although the ABC method is equally applicable in 3D. For the LES, the compressible filtered Navier-Stokes equations were solved with second-order accuracy in space and time using the one-equation eddy viscosity model [121] for closure of sub-grid scale stresses. The LES domain was discretized using $61,088$ cells, with grid stretching in the vertical direction and two levels of refinement near the jet inlet. For the DNS computations, the compressible Navier-Stokes equations were solved in conjunction with the total energy equation at a high spatial resolution, again using a second-order accurate numerical scheme in space and time. The DNS domain was discretized using $745,472$ cells with two levels of grid refinement, giving a grid that was over an order of magnitude larger than the LES grid.

For both the LES and DNS, the ideal gas equation was used to relate state variables, and fluid viscosity and specific heat varied with temperature according to the Sutherland model [122] and JANAF tables [123], respectively. The LES and DNS numerical schemes were also the same, with Crank-Nicolson time stepping and Gauss integration with linear interpolation for spatial derivatives. Pressure-velocity coupling was accomplished using the PIMPLE algorithm, which combines the pressure-implicit split-operator (PISO) and the semi-implicit method for pressure-linked equations (SIMPLE). An adaptive time step was used to limit the maximum CFL condition to 0.5 for the LES and 0.2 for the DNS.

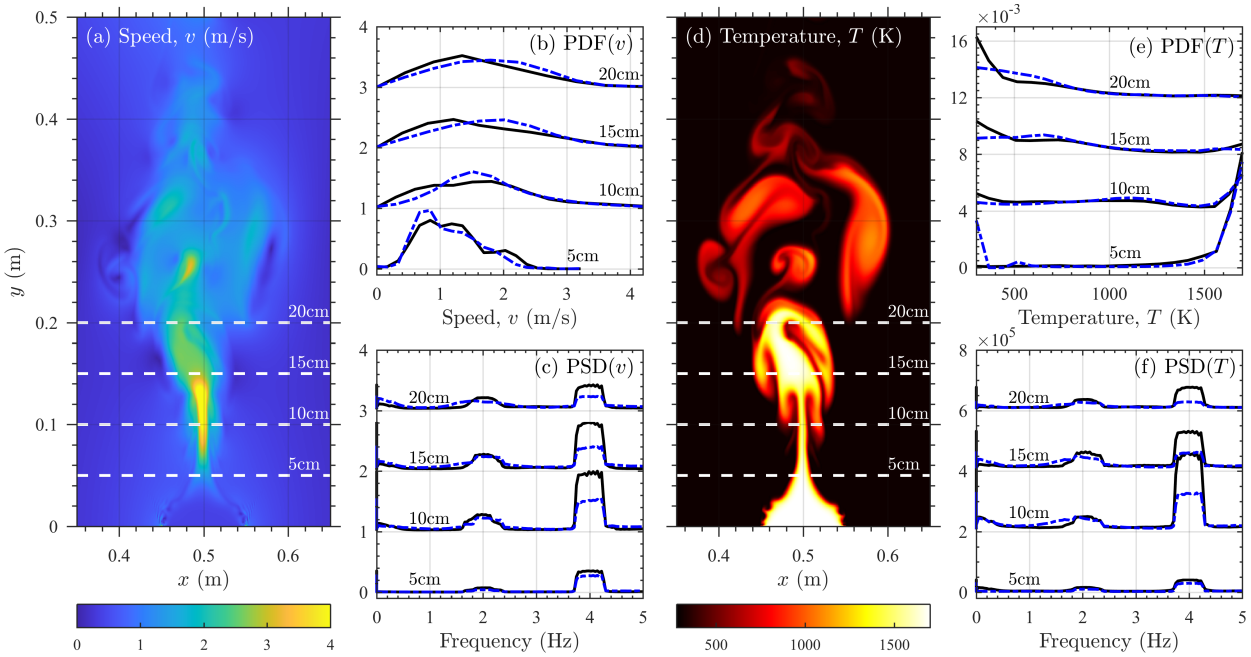


Figure 3.7: Representative fields of speed v (a) and temperature T (d) from DNS for the periodically forced turbulent buoyant jet. Panels (b) and (c) show reference probability density functions (PDFs) and power spectral densities (PSDs), respectively, of v at heights of 5, 10, 15, and 20 cm. PDFs and PSDs at different heights are shifted vertically for clarity. Panels (e) and (f) show corresponding PDFs and PSDs of T. Solid black lines in panels (b,c,e,f) show DNS reference data and dash-dot blue lines show LES reference data. PSDs are computed using Thomson's multitaper estimate [1].

3.2.1.2 ABC Overview

The specific ABC algorithm used in this work is Method D from Marjoram *et al.* [2]. Given the summary statistic $\boldsymbol{\mathcal{S}}$ obtained from reference data $\boldsymbol{D}$, this algorithm involves the following steps (see also the schematic in Figure 3.8):

(1) Generate parameters $\boldsymbol{\theta}$ from the prior distribution $P(\boldsymbol{\theta})$.

(2) Simulate approximate data $\hat{\boldsymbol{D}}$ using parameters $\boldsymbol{\theta}$, and compute the corresponding statistic $\hat{\boldsymbol{\mathcal{S}}}$.

(3) Calculate the distance $\delta(\boldsymbol{\mathcal{S}}, \hat{\boldsymbol{\mathcal{S}}})$ between the reference statistic $\boldsymbol{\mathcal{S}}$ and the simulated statistic $\hat{\boldsymbol{\mathcal{S}}}$.

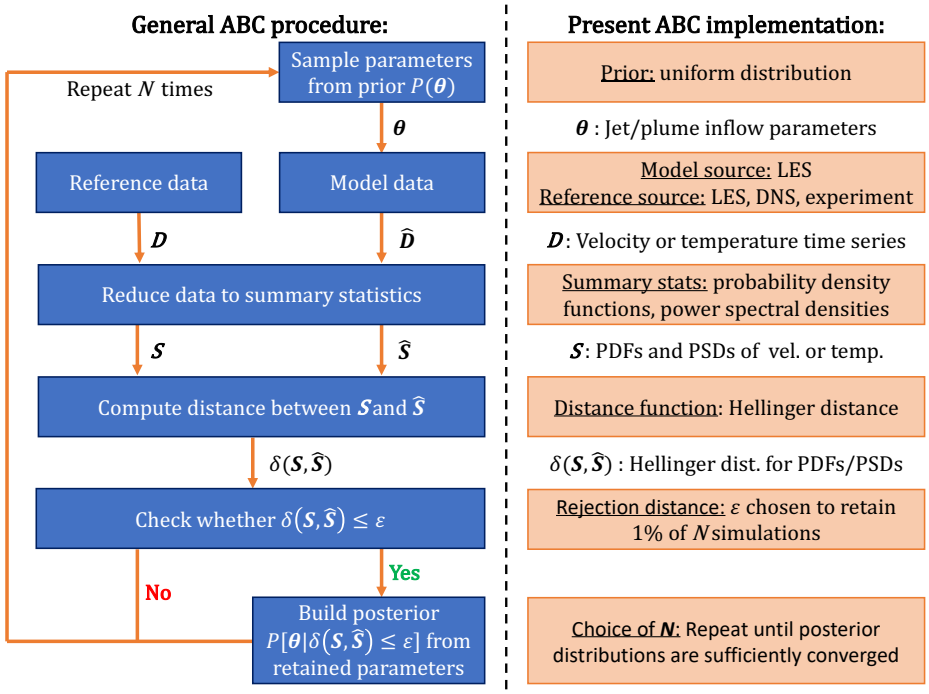


Figure 3.8: Schematic of the general ABC approach corresponding to Method D from Marjoram *et al.* [2], and of the specific ABC implementation used in the present study.

(4) Accept $\boldsymbol{\theta}$ if $\delta(\boldsymbol{\mathcal{S}}, \hat{\boldsymbol{\mathcal{S}}}) \leq \varepsilon$ (where ε is the 'rejection distance') and build the posterior distribution of accepted parameters, denoted $P[\boldsymbol{\theta} \mid \delta(\boldsymbol{\mathcal{S}}, \hat{\boldsymbol{\mathcal{S}}}) \leq \varepsilon]$.

(5) Return to step 1 and repeat a total of N times until a reasonable estimate is obtained for the posterior distribution.

In general, the parameters $\boldsymbol{\theta}$, data $\boldsymbol{D}$ and $\hat{\boldsymbol{D}}$, and summary statistics $\boldsymbol{\mathcal{S}}$ and $\hat{\boldsymbol{\mathcal{S}}}$ are all multi-dimensional and are correspondingly indicated by boldface notation.

In broad terms, the prior distribution is simply the initial guess for the values of the true parameters. Greater confidence in the values of unknown parameters permits a more concentrated prior distribution (e.g., a Gaussian instead of a uniform distribution). The only requirement on the prior is that its range spans the true values of the unknown parameters. In many practical cases a wide prior can be used initially to gain an approximate idea of the true parameter values, and

then a narrower prior can be used to determine parameter values with greater precision.

The use of summary statistics is intended to reduce the overall computational cost of the parameter estimation. Although it would be ideal to compare all of the available data $\boldsymbol{D}$ to all of the simulation data $\hat{\boldsymbol{D}}$, this is typically a very high-dimensional problem and is not computationally feasible in general. Thus, instead, summary statistics are used. A key challenge in successfully implementing ABC is to identify relevant summary statistics that significantly reduce the dimensionality of the data, while still maintaining sufficient identifiability of unknown parameters [107]. Examples of summary statistics include averages, standard deviations, probability distribution functions (PDFs), and power spectral densities (PSDs).

The choice of distance function is typically based on the form of the summary statistic. For example, the root mean square error can be used to compare spatial profiles of average quantities and the Hellinger distance [124] or Kullback-Leibler divergence [125] can be used to compare PDFs. The specific choice of rejection distance, ε, is typically determined by a balance between computational cost and parameter estimate uncertainty. Smaller values of ε lead to higher rejection rates and reduced uncertainty in estimated parameter values [59], but also increase the number of model simulations necessary to generate a sufficiently converged posterior.

With respect to evaluating the success of the ABC parameter estimation, "accuracy" can be assessed either by comparing the approximate posteriors generated using ABC to the true posteriors, or by comparing the estimated parameter values to the true parameter values, along with a measure of certainty in the estimated values. In the present study, however, the true posteriors are not known analytically and they are prohibitively expensive to compute numerically. As a result, the accuracy of the ABC approach is quantified here by computing 95% Bayesian confidence intervals (sometimes called credible intervals), denoted C_B, from the approximate parameter posteriors, then determining whether C_B contains the true parameter value. The ultimate measure of accuracy in the present study is indicated by how closely a central tendency of the posterior (such as the mean) matches the true parameter value. The degree of certainty in the estimated parameter value is determined by the width of C_B, where narrower intervals indicate more certain parameter

estimates.

Additional details on the specific choices for the prior, summary statistics, distance function, and rejection distances used in the present ABC implementation are outlined in the following sections, and are also summarized in Figure 3.8.

3.2.1.3 ABC Implementation

In this demonstration, ABC is used to estimate the frequency, amplitude, and mean of the periodic forcing at the inlet of the buoyant jet given either LES or DNS reference data, with model simulations from LES. The reference data consists of time series of speed or temperature measured at nine heights above the inlet, ranging from roughly 0.5 mm up to 0.2 m. Time series at each height were recorded over approximately 14 s after a 1 s initialization period, and PDFs and PSDs were computed for each time series; these are the ABC summary statistics. The PSDs were calculated using Thomson's multitaper estimate [1, 126], which provides a robust estimate of the PSD by reducing energy leakage across frequencies and reducing variance [127].

To ensure statistical convergence of the reference data, ensembles of 100 statistically independent simulations were created using both LES and DNS. Each simulation had the same periodically forced inflow at a frequency of 4 Hz, an amplitude of 0.2 m/s, and a mean of 0.5 m/s, but was initialized using different stochastically generated temperature fields. The final reference data for both the LES and DNS cases were then created by averaging the PDFs and PSDs for each of the 100 simulations in each ensemble, at each of the nine heights, giving the reference datasets shown, for example, in Figure 3.7(b,c) for speed and in Figure 3.7(e,f) for temperature. Moreover, Figure 3.7 shows that the LES and DNS reference data are in reasonably good agreement.

Approximately 10^4 additional LES model cases were generated for the parameter estimation. Each simulation was executed with a unique set of inlet parameters randomly sampled from prior distributions for the frequency, amplitude, and mean of the sinusoidal velocity oscillations at the inlet. The priors were uniform with bounds of 3.6–4.4 Hz for the frequency, 0.0–0.4 m/s for the amplitude, and 0.0–1.0 m/s for the mean. Uniform priors were chosen to avoid the creation of

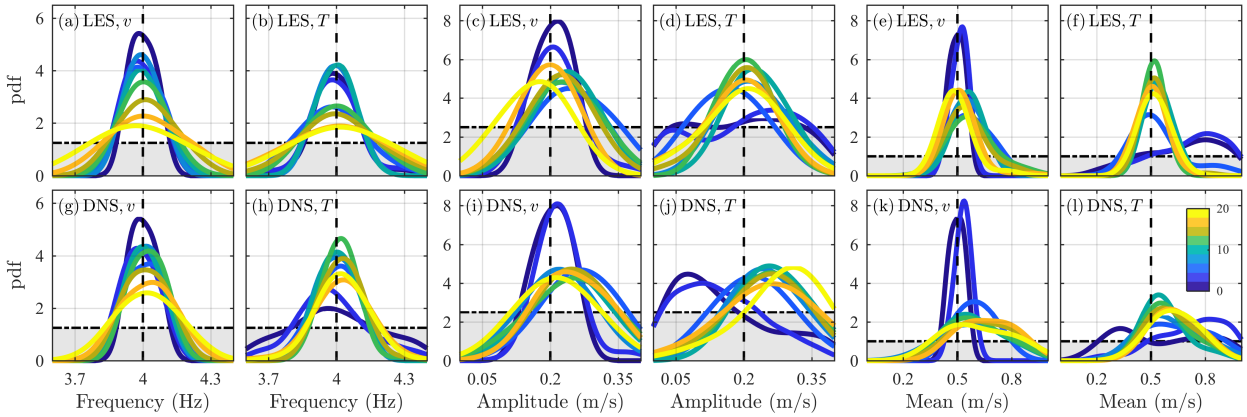


Figure 3.9: Marginal posterior distributions from ABC (visualized using Gaussian kernel estimation) for the frequency (a,b,g,h), amplitude (c,d,i,j), and mean (e,f,k,l) of the periodically forced turbulent buoyant jet. Posteriors are calculated using measurements at heights from 0–20 cm (indicated by colors from blue to yellow). Results are shown for speed-based reference data from LES (a,c,e) and DNS (g,i,k), and for temperature-based reference data from LES (b,d,f) and DNS (h,j,l). True parameter values are shown by vertical black dashed lines. Horizontal dash-dot lines and gray regions show the priors.

anomalous biases in the posterior distributions. Although wider priors would result in greater recovery of the edges of the posterior distributions, it will be seen in the following that nearly all of the posteriors peak within the range of values present in the priors.

The Hellinger distance [124] was used to quantify the agreement between summary statistics from the reference and model data. Values for ε were chosen such that a fixed percentage of the tested parameters were retained in the final posterior [66, 69]. Parameter rejection was first performed using the Hellinger distance for the PSD, where ε was selected to reject 80% of all parameters tested. Subsequently, rejection was performed using the Hellinger distance for the PDF, with ε chosen to reject 95% of the remaining parameters. With this sequential approach, only 1% of the parameters tested were included in the final posteriors.

3.2.1.4 Inflow Parameter Estimation

Figure 3.9 shows marginal posterior distributions for each inflow parameter using both speed and temperature reference data from LES and DNS at different heights. The posteriors are repre-

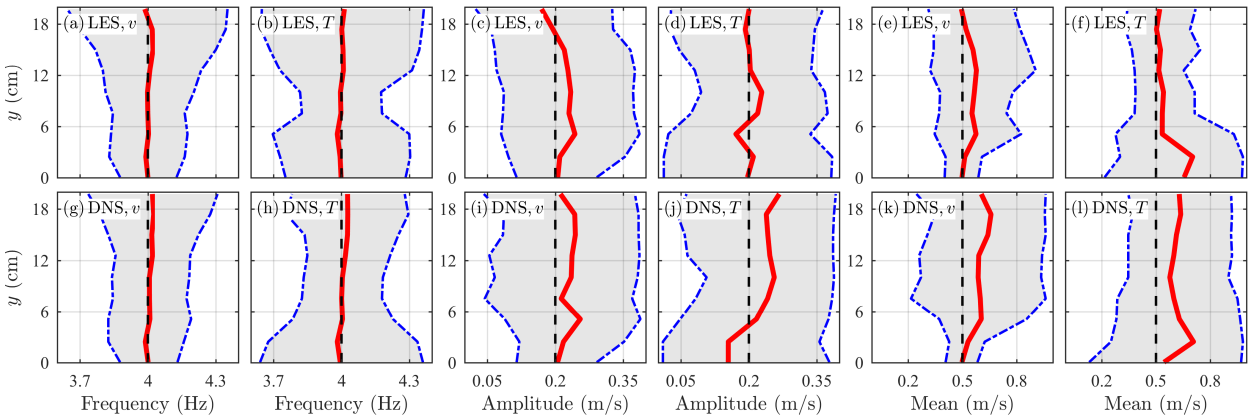


Figure 3.10: Vertical profiles of 95% Bayesian confidence intervals, C_B, for the marginal posterior distributions in Figure 3.9 (cases shown in each panel are described in the caption of Figure 3.9). Blue dash-dot lines and gray shading indicate the 95% Bayesian confidence regions, and the solid red lines show the mean values of the posteriors. True parameter values are shown by vertical black dashed lines.

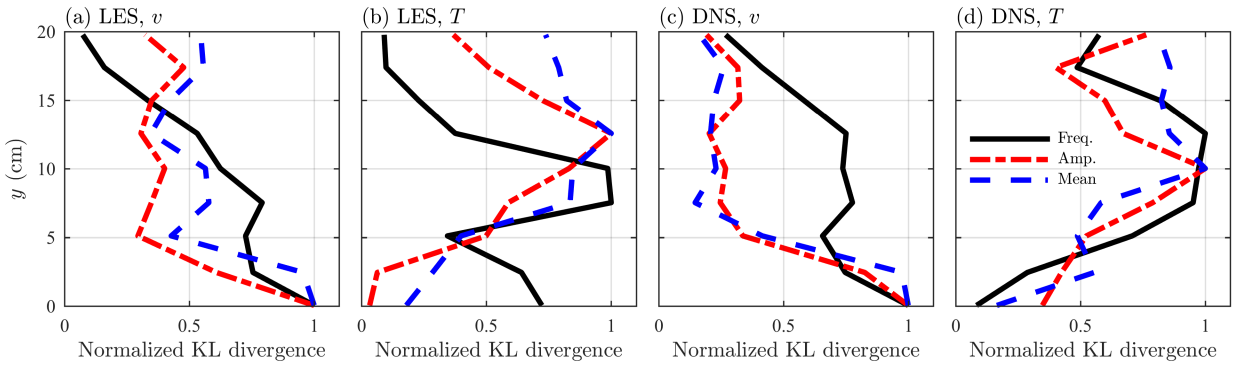


Figure 3.11: Vertical profiles of Kullback-Leibler (KL) divergences for the marginal posterior distributions in Figure 3.9. Divergences for the frequency (black solid lines), amplitude (red dash-dot lines), and mean (blue dashed lines) posteriors are shown for LES and DNS speed (v) based reference data (a,c) and for LES and DNS temperature (T) based reference data (b,d). Each of the KL divergences are normalized by their respective maximum values.

sented by Gaussian kernel estimations with bandwidths from Scott's normal reference rule [128]. For both the LES and DNS speed-based reference data, Figure 3.9 shows that the true parameter values are captured by the ABC posteriors. In general, the uncertainty is smallest at locations close to the inlet, although reasonable estimates of the true parameters are still obtained at higher locations.

For the speed-based measurements, Figure 3.10 shows that C_B is narrowest near the inlet when using either LES or DNS reference data. This indicates that identifiability of the unknown parameters is greatest near the inlet for speed-based reference data, although even higher in the domain it is still possible to estimate inflow parameters, albeit with less confidence. Figure 3.10 also shows that mean values from the posteriors are in good agreement with the true values, particularly near the inlet for the amplitude and mean, and at all heights for the frequency.

Figure 3.10 additionally shows that C_B is generally narrower for LES, as opposed to DNS, reference data. This greater certainty is consistent with the exact representation of the underlying physics by the model simulations when using LES reference data. However, even when using DNS reference data, reasonable parameter estimates are still obtained. As a result, ABC is able to provide predictions for unknown parameters using either exact or imperfect model simulations (as compared to the reference data), and for locations that are indirectly related to the parameters of interest (i.e., using measurements at locations far from the inlet).

The identifiability of unknown parameters using different measured quantities can be assessed by repeating the ABC procedure using PDFs and PSDs computed from temperature time series. Figure 3.9(c,d) shows that temperature-based frequency predictions are all centered on the true parameter value, and Figure 3.10 shows that the posterior means are in good agreement with the true value. The amplitude and mean posteriors are centered around the true parameter values in the region from approximately 0.5 to 1.5 jet widths above the inlet. Below these heights, however, there is not enough variation in the temperature field to aid in parameter estimation, given the uniform inflow temperature. Because the PDFs were nearly identical at such low heights, the ABC procedure was modified such that 15% of the simulated parameters were accepted based solely on the PSD comparison. Higher than 1.5 jet widths above the inlet, the connection between the temperature field and the inlet velocity becomes weaker, resulting in reduced identifiability of velocity boundary conditions using temperature measurements. Despite these limitations, as shown in Figure 3.10, all parameters are predicted for all heights within the C_B intervals for each posterior. Once again, the parameter estimation is generally more accurate using LES, as opposed to DNS,

reference data, reflecting the additional model error introduced when using the DNS reference data.

The amount of information gained about the unknown parameters during the ABC procedure can be quantified using the Kullback-Leibler (KL) divergence [125] between the prior and posterior distributions. High values of KL divergence indicate the posterior is significantly different than the prior (note that each KL divergence is normalized by its respective maximum value), while values close to zero indicate posterior similar to the prior. Figure 3.11 shows vertical profiles of the KL divergence for each of the different types of reference data, revealing that, in all cases, at least some information about the unknown parameters is gained at nearly all locations within the domain. For the speed-based reference data shown in Figures 3.11(a) and (c), the greatest information about the unknown parameters is obtained near the jet inlet, and the posteriors for the frequency generally provide more information higher in the domain than the posteriors for the amplitude and mean. By contrast, for the temperature-based reference data in Figure 3.11(b) and (d), the greatest information is obtained higher in the domain, close to 10 cm (corresponding to one jet width above the inlet). In general, only the temperature-based data near the jet inlet fail to provide significant information regarding the unknown parameter values.

3.2.2 Demonstration of ABC Using Experimental Reference Data

To demonstrate the utility of the ABC approach when the reference data are obtained from an experiment, the ABC technique is next applied to a steadily forced planar plume for which experimental data are available from Cetegen *et al.* [3]. In this experiment, a helium-air mixture is weakly forced into 300 K ambient air, resulting in a natural oscillatory behavior of the plume. By spanning a range of inflow conditions, Cetegen *et al.* [3] determined the empirical relation $\mathrm{St} = 0.55\mathrm{Ri}^{0.45}$ between the Strouhal, $\mathrm{St} = fw/v_i$, and Richardson, $\mathrm{Ri} = (1 - \rho_i/\rho_\infty)gw/v_i^2$, numbers, where ρ_i is inlet density, ρ_∞ is ambient density, g is gravity, w is inlet width, v_i is inlet velocity, and f is the frequency of the natural oscillation.

In this demonstration, ABC is performed assuming that the oscillation frequency above the jet inlet, f, has been measured experimentally and that the density of the inflow mixture, ρ_i/ρ_∞,

is unknown (where $\rho_i/\rho_\infty = 0.14$ for pure helium and 1 for pure air). All other properties of the inflow, including v_i and w, are assumed to be known. This case thus serves as the first experimental demonstration that ABC is able to accurately estimate unknown parameters in complex thermal-fluid flows, while also providing measures of uncertainty in the parameter estimates.

The model simulations were once again performed using LES in FireFoam [118]. The plume had a width of $w = 0.07$ m and a steady uniform inflow velocity of $v_i = 0.067$ m/s. The reference Strouhal number was assumed to be $\mathrm{St} = 4$, which is close to several values measured experimentally (see Figure 3.13). The inlet width and velocity were held fixed for 100 LES runs while the inlet density ratio varied from 0.14 to 0.6, corresponding to the published range of density ratios from [3]. The lowest density ratio corresponds to pure helium while the upper value was found in [3] to be the bounding case beyond which the characteristic natural oscillation was not observed. The Strouhal number was obtained from the peak frequency in a fast Fourier transform of vertical velocity a distance w above the inlet; this was the summary statistic. The distance function consisted of an L1 error norm between reference and model peak frequencies, and ε was chosen to retain 20% of the parameter values from the model simulations. The characteristic oscillation of the plume is captured by the LES, as shown in Figure 3.12.

Figure 3.13 shows that the estimated value of Ri from ABC for $\mathrm{St} = 4$ agrees closely with the experimental data. Additional tests for $\mathrm{St} = 2$ and 3 were also performed using $v_i = 0.1$ m/s and 0.15 m/s, respectively. These different inlet velocities were necessary to span a range of Ri while constraining the physically allowable values of ρ_i/ρ_∞ to between 0.14 and 0.6. Once again, Figure 3.13 shows that ABC provides estimates of Ri that are in close agreement with experiments. As a result, ABC correctly identifies the unknown Ri that would provide the desired natural oscillation frequencies for known inlet velocity and width.

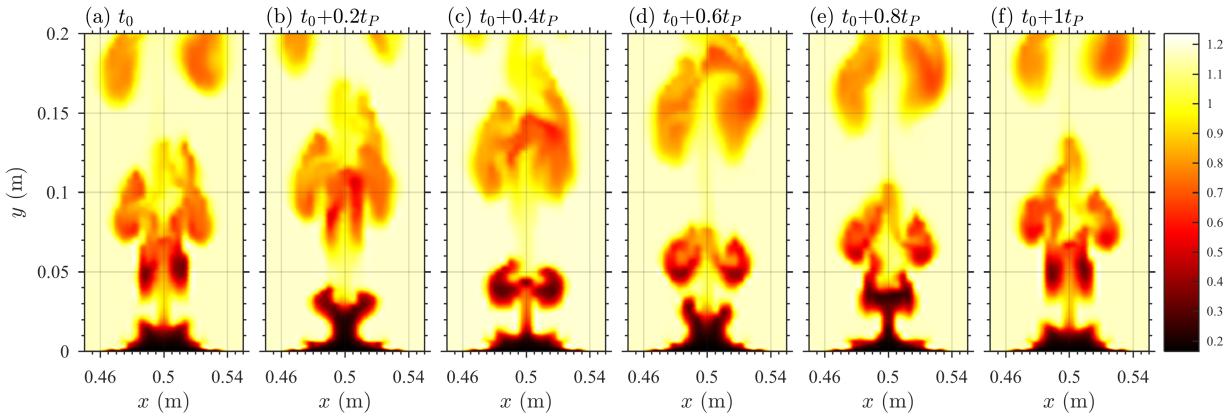


Figure 3.12: Representative 2D fields of density (kg/m^3) from LES at six different times (a-f) for the steadily forced plume described in Section 3.2.2. The fields shown begin at an arbitrary time t_0 late in the LES and are separated by a time interval spanning one complete period, t_P ($t_P = 1/f = 0.22$ s).

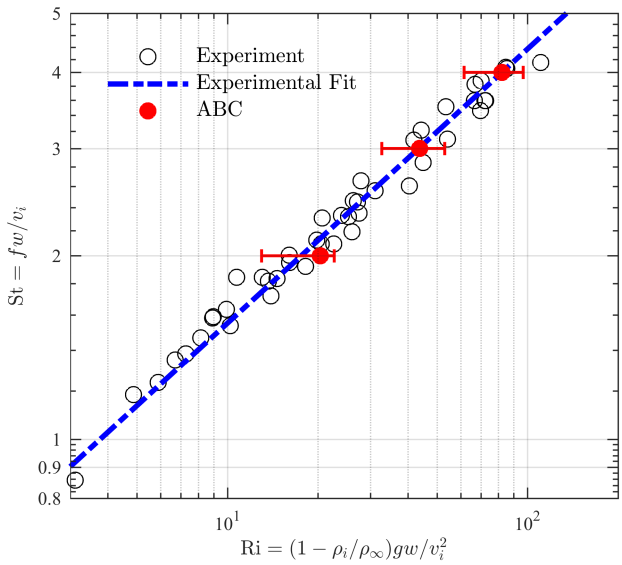


Figure 3.13: Relationship between Strouhal number St $= fw/v_i$ and inlet Richardson number Ri $= (1-\rho_i/\rho_\infty)gw/v_i^2$ for the steadily forced helium-air plume. Experimental results from Cetegen *et al.* [3] are indicated by open circles, and the empirical fit St $= 0.55\text{Ri}^{0.45}$ is shown as a blue dash-dot line. ABC results are indicated by filled red circles located at the mean Ri of the posterior distribution. Uncertainty bars show the minimum and maximum values of Ri in the posteriors.

3.3 ABC Methods: Additional Considerations

This effort builds upon the successes of Section 3.2.1.3. This section uses the same 2D DNS reference data (based on 100-member ensemble) and approximately 10,000 LES cases. While the results in those sections form a complete picture and successfully estimate several unknown parameters (namely inlet velocity mean, amplitude, and frequency), additional work is performed here to complement that analysis. Specifically, the following sub-sections will seek to address the impact of varying quantities intrinsic to the ABC methodology such as rejection distance (ε) and number of simulations computed. Additionally, techniques to improve the efficiency of the ABC method will be explored; these include regression and parameter weighting by distance. Finally, this section will also address remedies when a model contains known bias. This section is based on a paper draft to be submitted to *Journal of Computational Science*.

3.3.1 ABC Implementation with "Basic" Parameters

The ABC procedure here uses single valued statistics to infer the mean, amplitude and frequency at the inlet; specifically, this analysis finds the mean, standard deviation, and fast Fourier transform (FFT) peak location for the reference data and the simulation. These statistics are used instead of the the entire PDF and PSD used previously. This change in statistics simplifies (and in some cases enables) the analysis and isolates the impact of various techniques being implemented. The approach herein compares each simulations values to the reference value to obtain an absolute distance corresponding to the three statistics (mean, standard deviation, and peak location). Then to calculate an aggregate distance for each simulation, these values are combined together. Combination works by normalizing each of the three components of the distance (corresponding to mean, standard deviation and peak location) by its maximum value (across all simulations), then adding together the resulting normalized distances after weighting the frequency by 0.1. The low weight for frequency was chosen because the combined distance contribution from frequency can dominate the weight otherwise; this weighting scheme also illustrates the ability of various techniques below

to narrow the width of a posterior that initially has limited information gain relative to the prior.

3.3.2 Impact of Rejection Distance (ε) on Posterior Convergence

One of the challenges that emerges with ABC is choosing an appropriate rejection distance, ε. Generally it is acknowledged that a particular choice for rejection distance must balance bias in the posterior against confidence in the posterior based on how many simulations are retained. By this we mean that if the rejection distance is too large, then the posterior distribution will be biased toward the prior distribution; in the extreme case, as the rejection distance grows the posterior will identically match the posterior once all simulations are accepted. Conversely, as the rejection distance decreases the number of simulations decreases which also decreases the confidence in the parameter estimates coming from the posterior distribution; this is known as Monte Carlo error. Marin *et al.* [59] and Lintusaari *et al.* [60] argue that a smaller ε is generally preferred, but the size is largely determined by the available computational resources. Another approach proposed by Christopher *et al.* [85] is to calculate the confidence interval (CI) of the posterior, which goes as the standard deviation of the accepted simulations divided by the square root of the number of accepted simulations. Assuming a unimodal, tight distribution of initial parameters this technique suggests the CI will initially decrease due to accepting more and more simulations that are all tightly clustered. Then, as more simulations are accepted, they will start to spread out across the posterior increasing the variance of the posterior. This will increase the size of the CI until the width of the posterior approaches approximately the full width of the prior distribution. Once this happens, the CI will again decrease as more simulations are accepted while the width of the posterior remains approximately constant. Note that the progression outlined here is one of several possible evolutions of the CI as a function of rejection distance; how each system will evolve depends on the growth rate in the standard deviation of the posterior versus the number of accepted samples retained.

This behavior is illustrated in Figures 3.14 and 3.15. For this analysis, the LES simulations were compared to DNS reference measurements at a height of 1 cm. For the statistics, the three

"basic" parameters are used (mean, standard deviation, and FFT peak location) as described in Section 3.3.1. Figure 3.14 shows the CI of the amplitude parameter posterior. Note that the rejection distance is normalized so at a value of 1 we accept all parameter values. This figure shows that for very small rejection distances we accept very few simulations so the CI is relatively large. Then, as we increase the rejection distance we also increase the number of accepted simulations tightly clustered around the initial value which decreases the confidence interval width; this trend continues until a normalized rejection distance of 0.05 to 0.1. Beyond this range the posterior begins to widen at a rate faster than the number of accepted simulations; as the posterior becomes less concentrated, the CI increases. This occurs until a rejection distance of approximately 0.18; after this many more simulations are accepted causing the CI to decrease. This suggests an optimal ε corresponding to the normalized rejection distance of 0.05 to 0.1. The confidence interval of the joint PDF for all three parameters is shown in Figure 3.15. It displays similar behavior to the amplitude parameter posterior.

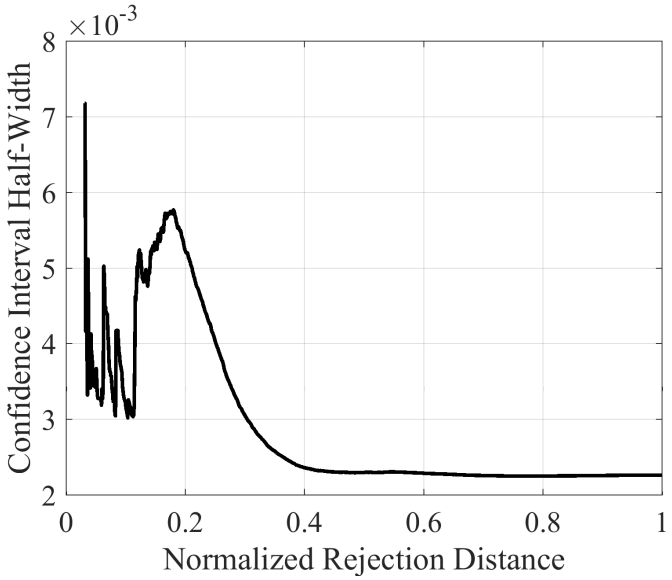


Figure 3.14: Confidence interval half-width plotted against normalized rejection distance for the amplitude parameter posterior.

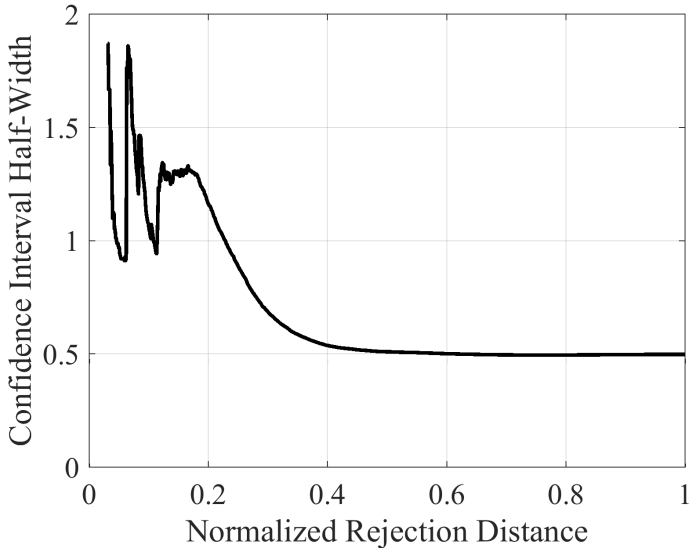


Figure 3.15: Confidence interval half-width plotted against normalized rejection distance for the joint posterior.

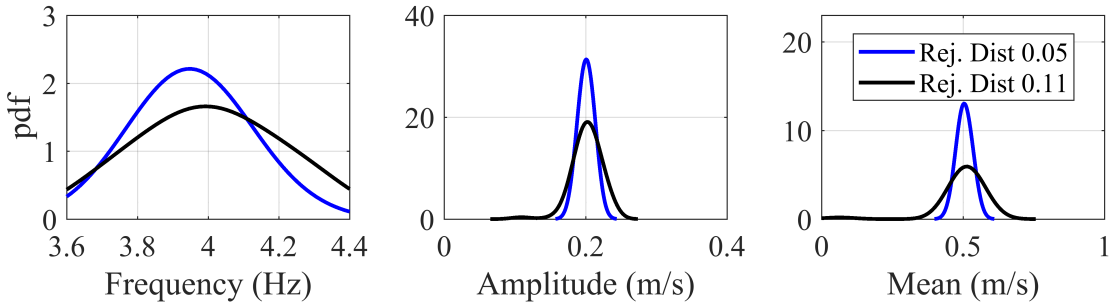


Figure 3.16: Posterior distributions for all three parameters created using "basic" statistics at 1 cm above the inlet. Posteriors are shown for two normalized rejection distances corresponding to valleys in Figure 3.15.

The posterior distributions for all three parameters at normalized rejection distances of approximately 0.05 and 0.1, corresponding to the dips in the above CI plots (Figs. 3.14 and 3.15) are shown in Figure 3.16. Notice that as the rejection distance is increased, the posterior widens slightly but has approximately the same confidence interval half-width due to the increased number of accepted samples.

Another approach to quantify how the posterior distributions change as a function of rejection

distance is to consider the Kullback–Leibler divergence [125] from the prior to the posterior. The Kullback–Leibler divergence is defined for a discrete probability distribution as

$$D_{KL}(P||Q) = \sum_{i=1} P(i) log\left(\frac{P(i)}{Q(i)}\right). \tag{3.3}$$

Equation 3.3 provides the Kullback–Leibler divergence from Q to P (which is different than the Kullback–Leibler divergence from P to Q as it is not symmetric). We define P(i) to be the posterior probability distribution resulting from ABC. We further define Q(i) to be the prior probability distribution from ABC. Thus, the Kullback–Leibler divergence tells us (in the Bayesian inference context) the information gained by going from the prior probability distribution Q to the posterior probability distribution P; thus, the higher the Kullback–Leibler divergence magnitude the more information gained by using the posterior. This is clearly seen in the plot in Figure 3.17. This shows the Kullback–Leibler divergence for each posterior as well as the joint posterior, each normalized by their peak values (which all occur for the smallest normalized rejection distance as is expected since that is the most peaked distribution). As more samples are accepted, the posteriors eventually spread out and return to the uniform prior distribution; as the rejection distance increases toward accepting all simulations the Kullback–Leibler divergence goes to zero. Note that because the frequency is given a smaller weighting in the distance function it appears to become the worst (i.e. returns to the prior distribution) the fastest with increasing rejection distance. This illustrates how one is able to improve the estimate of one parameter (mean) at the expense of the ability to simultaneously predict another parameter (frequency).

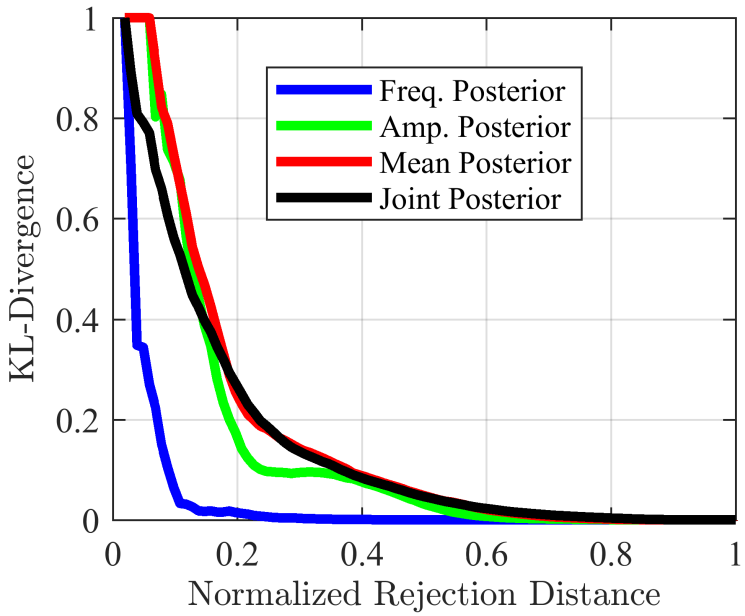


Figure 3.17: Kullback-Leibler (KL) divergence from the prior to the posterior distributions. Each KL divergence is normalized by it maximum value. High values of KL divergence indicate high information gain, while values close to zero indicate a posterior that is similar to the prior distribution.

3.3.3 Statistical Convergence of Posterior Profiles

Another avenue of study for ABC involves the efficiency of the method. A rough guideline thumb indicates that the number of simulations (N) required to estimate n parameters goes as ($N \sim \mathcal{O}(10^n)$); thus for three parameters, $10^3 = 1,000$, so on the order of 1,000 simulations are required. In this study, to ensure statistical convergence, approximately 10^4 simulations have been used. Thus it is reasonable to consider how the posterior distributions would change if fewer simulations are used. The resulting posteriors from the first approach considered are shown in Figure 3.18. In this approach, the best 100 simulations are accepted regardless of the number of total simulations performed. The quantity of total simulations is varied from 1% of all of the simulations (100 simulations) to all of the original simulations (approximately 10^4 simulations). As the total number of simulations increases, note how posterior distributions narrow and become more concentrated around true values. As anticipated, this indicates that more simulations generally

provide better posteriors. However, the question of how many simulations are needed has yet to be answered. The most obvious answer is to conduct as many simulations as possible based on computational budgets. However, it is generally helpful to have an upper limit in mind so as not to waste precious computing resources. Thus, one could consider the Kullback-Leibler divergence [125] from each prior distribution to the corresponding posterior and determine when it asymptotes; after this point additional simulations would add little to the parameter estimation. The mean and frequency parameter posteriors exhibit this behavior in Figure 3.18. As more simulations are added beyond 20% of the total simulation pool size (i.e. 2,000 of 10,000 available) the posterior does not improve substantially. The amplitude posterior does have a marked improvement between 30% and 100% of the simulations being used, however. Lintusaari *et al.* [60] take this approach of considering the Kullback-Leibler divergence of the posterior when comparing various rejection distances and summary statistics. They are able to also include the full Bayesian posterior, however, adding additional support to their conclusions of how many simulations are required before their approximate posteriors asymptote. In the present analysis, the full Bayesian posteriors are prohibitively expensive to compute and thus such a benchmark (converging to the full Bayesian posterior) is unfeasible.

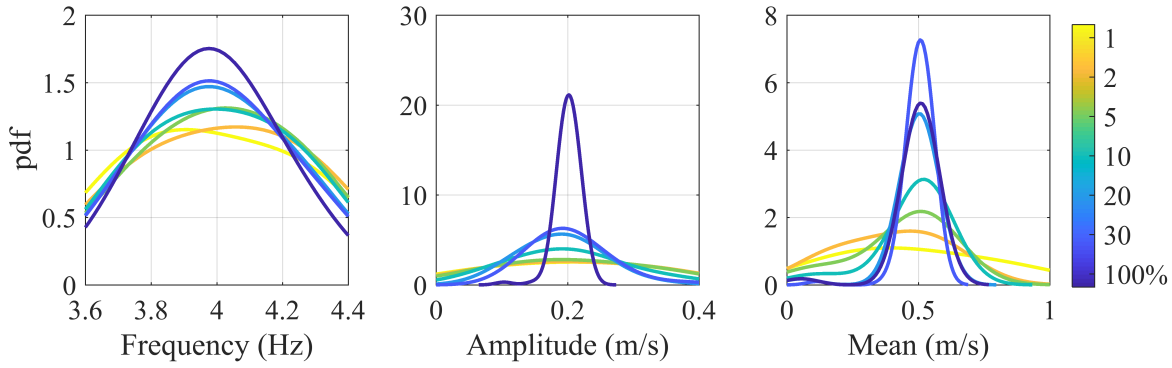


Figure 3.18: Posterior distributions for three inlet velocity parameters. Posteriors created by accepting 100 closest simulations while increasing the size of the total simulations considered (from 1% of total $\approx 10,0000$ available to 100% of them); the lines are colored by total sample size considered.

Another approach to decide when the posteriors have converged is to fix the number of

simulations accepted at 2%, then vary the total number of simulations considered (this equates to increasing the rejection distance as the number of total simulations considered increases). In this approach, far fewer simulations are accepted compared to the first approach, until the total number of simulations used for the analysis is 5,000 (i.e. 2% of 5,000 equals 100, which is the fixed number of accepted samples in the first approach). The small numbers of accepted simulations for this approach cause the posteriors, in general, to be narrower than the first case, as is seen in Figure 3.19. All of the posteriors end up closely centered on the true values and have tight credible intervals when approximately 1,000 or more simulations are used as the total pool size; this agrees with with the guideline ($N \sim \mathcal{O}(10^n)$) (before this the frequency posterior is quite spread out). These two approaches highlight the inexorable connection between rejection distance and total number of computations performed.

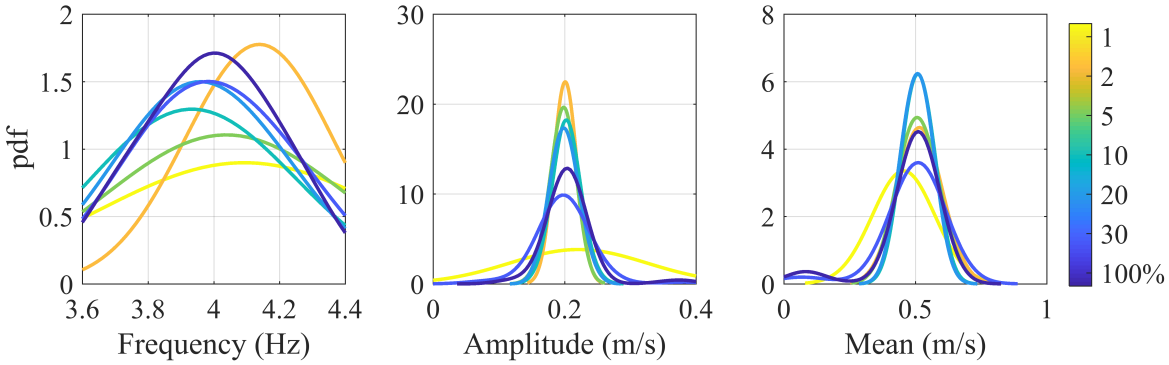


Figure 3.19: Posterior distributions for three inlet velocity parameters. Posteriors created by accepting 2% of the total simulations considered that are closest to the reference DNS ensemble. Each posterior has a unique number of the total simulations considered (from 1% of total $\approx 10,0000$ available to 100% of them); the lines are colored by total sample size considered.

3.3.4 Regression

A technique that is presented frequently in approximate Bayesian computation literature pertains to regression [66, 129]. In principle regression is supposed to increase the computational efficiency of ABC by allowing more simulations to be retained while still maintaining a relatively narrow posterior. This works by applying the standard ABC rejection method to obtain the

posterior, $P_{ABC}(\theta|D)$ (in the present case we accept $\approx 1.5\%$ of the total $\approx 10,000$ simulations and use the "basic" statistics and associated distance function from Section 3.3.1). However, after creating the posterior, an extra step is taken. The parameters that are retained for the posterior are plotted against their corresponding statistic. For this particular application, we show in Figure 3.20 the prescribed inlet mean velocity versus the mean velocity magnitude observed 1 cm above the burner surface. In red we show the observed velocity from the DNS reference ensemble data. Next, a linear regression of the parameters is performed along the best fit line for the simulation data. In this case a linear regression appears to fit well, but nonlinear choices may better suit other applications [130]. Once all of the parameters and their corresponding observations are regressed to the reference observation then a new posterior is made based on the new locations of the parameters (found by horizontally tracing each newly located simulation to the vertical axis on the regression plot of Figure 3.20. Figure 3.21 shows the impact of performing the regression; note how much more peaked the posteriors are after the regression. By concentrating the posteriors so dramatically, regression allows fewer simulations, that may be spread out, to be retained for the final posterior while still creating a posterior with narrow credible intervals and high information gain compared to the prior.

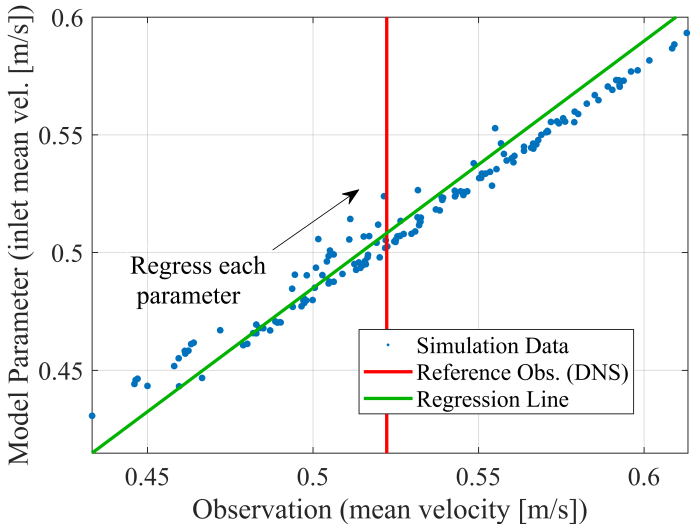


Figure 3.20: Regression schematic. Each blue dot represents a simulation output at its original location; this data is used to compute the original posteriors. The green line shows the regression line which is a best fit of the data. Each simulation is regressed along the green regression line until the it reaches the reference observation data from DNS simulation at which point the posterior is recomputed. An identical approach is applied for each parameter independently.

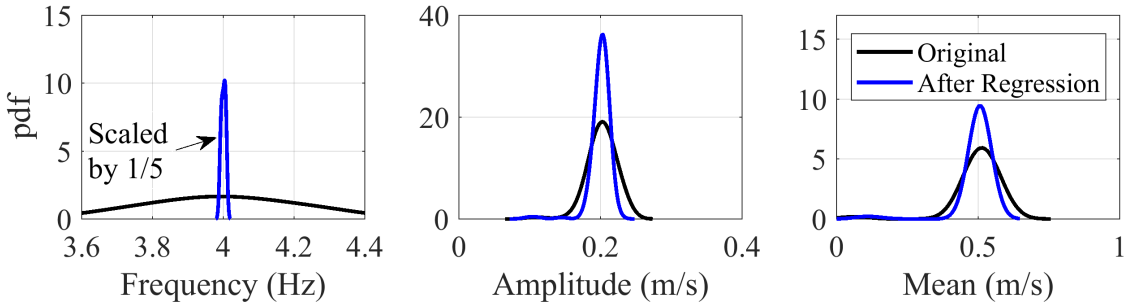


Figure 3.21: Posteriors before and after regression are presented. Regression causes each posterior to narrow.

3.3.5 Weighting Parameters to Form Posterior

Another technique we propose to reduce the computational cost of the ABC procedure is to keep all of the simulations (i.e. $\varepsilon \to \infty$) but then to use information from each of them to inform the posterior estimate. This is done by following the standard ABC procedure (given in Section 3.3.1) with $\varepsilon = \infty$. Before generating the posterior, however, an extra step is taken since

the posterior at this point would equal the prior distribution with all parameters accepted. Based on the distance obtained for each simulation, the parameters for that simulation are weighted accordingly. The simulation that is closest to the posterior receives the most weight, while the simulation that is most dissimilar from the reference observation receives the least weight. Finally, a posterior distribution is created based on the accepted parameters according to their weights. This is illustrated in Figure 3.22. Various functions may be used to weight each simulation, however, in this case to further differentiate the simulations that are close from those that are far from the reference simulation, a weighting function of $1/distance^2$ was used.

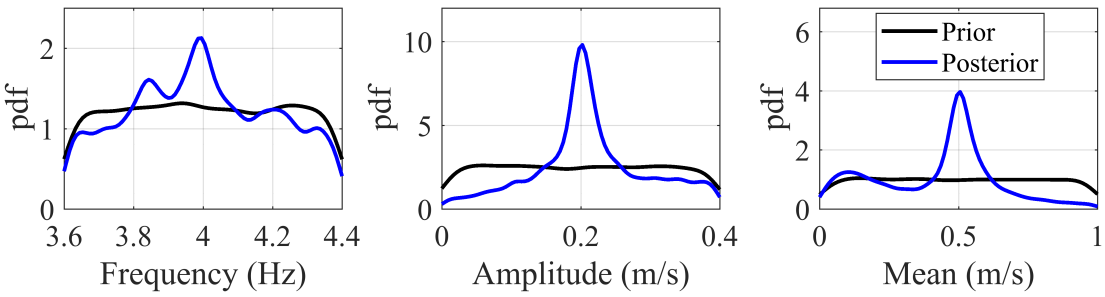


Figure 3.22: Posteriors before and after weighting parameters by rejection. Weighting causes each posterior to narrow.

3.3.6 ABC in the Presence of Model Bias

Model bias is a primary concern when performing ABC, because if a model is biased then the parameter estimations it produces may also be biased. Though not always the case, bias in a model may be known *a priori* due to calibration efforts with well-known conditions. In such a case, model outputs can be adjusted to account for model bias. Such an approach is proposed here. The DNS ensemble and LES ensemble serve as statistically-converged data sets representative of behavior of either simulation method, respectively. Thus, comparing the mean speed profiles produced by the DNS ensemble and LES ensemble provides information as to the biases in the LES data; note that here we use the 100-member ensemble average taken after temporally averaging each member's speed profile. These biases arise due to differences in the DNS and LES solutions,

primarily caused by turbulence modeling in the LES, higher resolution of the DNS grid, and tighter constraints on the DNS time stepping scheme (i.e. CFL = 0.2 compared to CFL = 0.5 for LES). Mean profiles for both the DNS and LES ensembles are shown in Figure 3.23 along with the bias of the LES simulation. Note that the LES simulation starts biased toward slower speeds, but then around 10 cm (i.e. 1 inlet width) above the inlet the bias flips signs and becomes positive (i.e. the LES predicts higher speeds than the DNS). Using this observed bias, all LES vertical profiles are corrected.

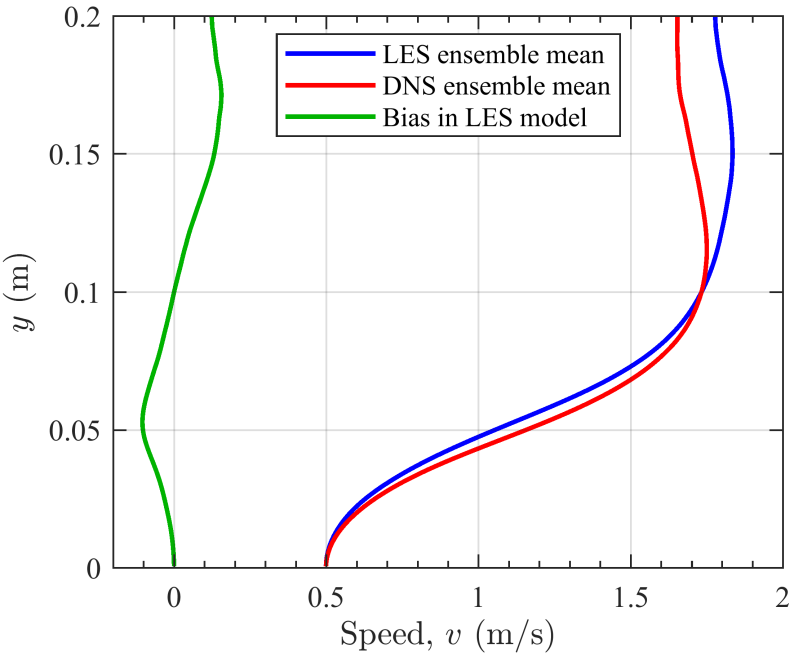
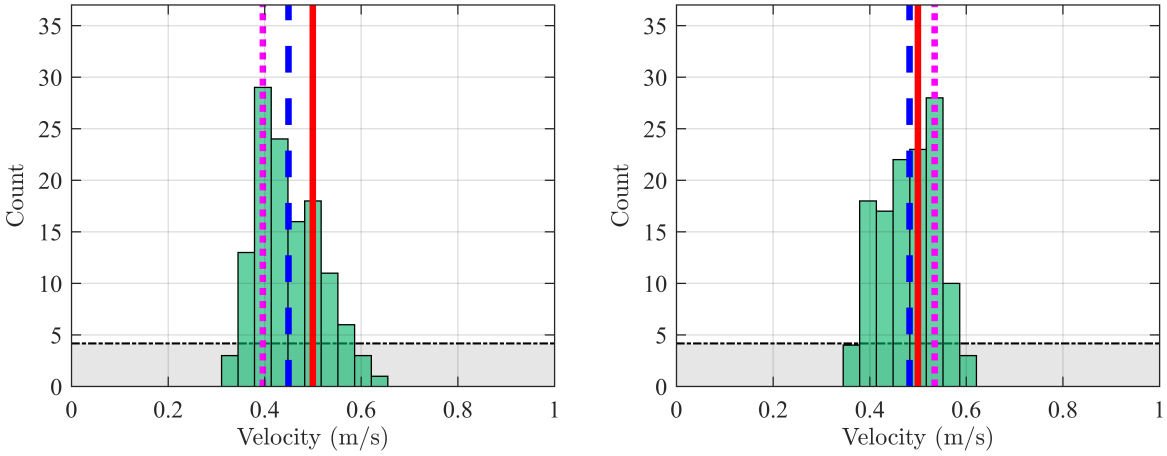


Figure 3.23: Vertical profiles of simulation speed are shown. One-hundred member ensembles of DNS and LES were computed. Each ensemble member was temporally averaged. Then the ensemble members were averaged together for either the DNS or LES ensemble. These means are shown in red and blue, respectively. The difference in profiles at each height is reported as the bias in the LES model and shown in green; the bias is close to 0 m/s for much of the domain.

Since the bias correction applied here is for the vertical profile of speed, it makes sense to use the vertical profile of speed as the summary statistic to represent each simulation and the DNS ensemble mean. To compare the vertical profile from each simulation to the DNS ensemble mean vertical profile, the root-mean-square error (RMSE) is used. Thus, a larger RMSE indicates more

disparity between a given simulation (and hence its parameters) and the DNS ensemble mean. Two posterior histograms are shown in Figure 3.24 representing the inlet mean velocity predicted first without, and then with, the bias correction applied. Note that the mean of the posterior improves from 0.45 m/s to 0.48 m/s after applying the the bias correction (the error goes from over 10% initially to less than 3.5% after fixing the bias in the LES simulations).



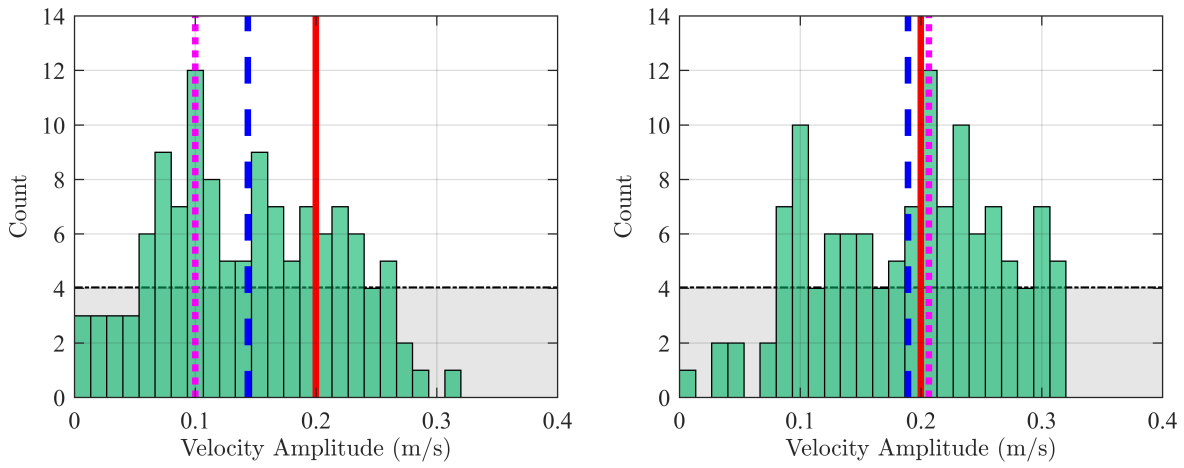
(a) Original profiles used

(b) Bias-corrected profiles used

Figure 3.24: Histograms showing posterior estimations for inlet mean velocity parameter. The uniform prior is indicated by the shaded gray box (normalized for the appropriate number of accepted simulations). The true parameter value used to drive the reference DNS ensemble shown with solid red line. Dashed blue line shows the mean of the posterior parameters. Dotted magenta line indicates mode of distribution. Subfigure (a) was created using original speed profiles, while subfigure (b) was crated using profiles after correcting for the LES bias.

Two additional posterior histograms are shown in Figure 3.25 representing the inlet velocity amplitude predicted without and then with the bias correction applied. Note that the mean of the posterior improves from 0.14 m/s to 0.19 m/s after applying the the bias correction (the error goes from over 28% initially to just over 5% after fixing the LES bias). Thus, it is clear the mean speed profiles provide useful information regarding both the inlet velocity mean and amplitude. The inlet velocity frequency posterior is less clear, however, indicating the mean speed profile may not contain sufficient information to accurately predict this parameter. This indicates the parameter

is either unidentifiable or noninfluential using this summary statistic [107].



(a) Original profiles used

(b) Bias-corrected profiles used

Figure 3.25: Histograms showing posterior estimations for inlet velocity amplitude parameter. The uniform prior is indicated by the shaded gray box (normalized for the appropriate number of accepted simulations). The true parameter value used to drive the reference DNS ensemble shown with solid red line. Dashed blue line shows the mean of the posterior parameters. Dotted magenta line indicates mode of distribution. Subfigure (a) was created using original speed profiles, while subfigure (b) was crated using profiles after correcting for the LES bias.

3.4 Conclusions

Approximate Bayesian computation (ABC) has, for the first time, been used to estimate unknown physical parameters in complex thermal-fluid flows. As a demonstration of the approach, ABC was used to estimate the frequency, amplitude, and mean of the velocity inflow in a periodically forced turbulent buoyant jet, as well as the inlet Richardson number for a steadily forced helium-air plume. In the former case, computational reference data were used to drive the parameter estimation, while in the latter case experimental reference data were used. For both cases, ABC provided accurate estimates of unknown physical parameters even when the model simulations did not exactly match the physics underlying the reference data (e.g., when using LES for the model simulations and reference data from either DNS or experiments), and when the reference data were

only indirectly connected to the unknown parameters of interest (e.g., when using measurements distant from the jet and plume inlets, or when using temperature measurements to infer velocity inlet parameters). The primary impacts of using imperfect model simulations or only indirectly connected measurements were observed in the width of the Bayesian confidence intervals. As the physics of the model simulations more closely matched that of the reference data, and as the connection between the measurements and unknown parameters improved, the confidence intervals were found to become more narrow.

Although the present demonstration indicates that ABC provides accurate estimates of parameter values with relatively little uncertainty, there are many directions for future research. Different choices for the prior, summary statistic, distance function, and rejection distance can all potentially lead to slightly different parameter estimates. Further work is required to determine the specific effects of each of these choices on the accuracy and uncertainty of ABC. Markov chains can also be used to significantly reduce the cost of ABC by limiting the number of parameter values that are rejected, even for relatively small rejection distances [2]. Such approaches have become more popular in recent years, but have yet to be applied in the present context. Finally, the true power of ABC lies in its ability to predict unknown parameters in real-world systems, and future work will more deeply explore the use of experimental reference data to drive parameter estimates, particularly taking into account experimental uncertainty in the context of validation efforts for simulations of complex thermal-fluid problems.

Chapter 4

Parameter Estimation for Industrial Applications using ABC

In this chapter, we use ABC to predict unknown temperature or velocity conditions in simulations of a three-dimensional (3D) turbulent, high-temperature buoyant jet, and then a two-dimensional (2D) turbulent, high-temperature buoyant jet with a rotating cylinder above it. For both test cases, reference data are obtained from corresponding simulations with known boundary conditions and problem parameters. Using spatially sparse temperature statistics from each reference simulation, we show that the ABC method provides accurate predictions of the temperature mean and variability at the jet inlet for the 3D case. We then show that similar spatially sparse temperature statistics can be used with ABC to accurately predict the inlet velocity and cylinder rotational velocity. The success of the ABC approach in the present tests suggests that ABC is a useful and versatile tool for engineering fluid dynamics research. As pointed out in Section 1.4, ABC has primarily been applied to biological and geophysical problems to date making this a unique and novel application of ABC.

It should be noted that, by contrast to prior studies [10,11,14,131] where simplified surrogate models have been used for parameter estimation in turbulent flows, here we instead perform more complete CFD simulations during the ABC process. This chapter is focused on the case of a turbulent buoyant jet due to the geometrically simple yet physically complex nature of this compressible flow problem, which has many analogs in both engineering systems and environmental flows. The case in Section 4.2.2 is an extension of this buoyant jet, but with an internal moving boundary above it; industrially this could represent a roller used to treat a thin film. The approach of using

the same model to generate the reference observations and subsequent simulations provides a clear indication of how well the ABC method can determine unknown parameters using only spatially sparse reference data; in the data assimilation community this is known as an Observing-System Simulation Experiment (OSSE) (see for example [116, 117]).

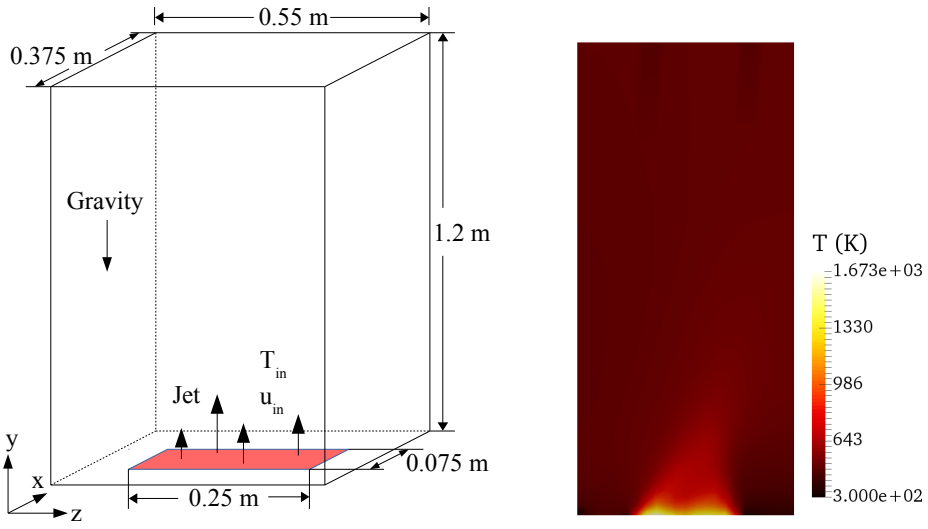
4.1 Approximate Bayesian Computation for a Turbulent Buoyant Jet

With the ABC method successfully applied to a 2D turbulent buoyant jet in Chapter 3, the method is next applied to CFD simulations of a *3D* turbulent buoyant jet as originally reported in [84]. This added complexity aims to demonstrate the efficacy of the method for a realistic engineering application. The open-source computational fluid dynamics software OpenFOAM, version 2.2.x, [119, 132] was used to design and run unsteady 3D RANS simulations for this study. The compressible RANS equations were solved in conjunction with the energy equation. Heat transfer mechanisms that were modeled in this simulation include advection of temperature by hot gases and radiation exchange between domain boundaries and the fluid within the domain. The standard k-ϵ turbulence closure model [133] was used. The RANS equations were solved with second order accuracy in space and time to obtain a solution. Limiters on velocity divergence and Laplacian schemes were implemented to aid convergence. Fluid viscosity and specific heat were assumed to be constant for the purpose of this study. The RANS simulation compares favorably to results from a 3D LES [96].

4.1.1 Turbulent Buoyant Jet Simulation Setup

The RANS equations were solved on a 3D grid with 6,270 hexahedral cells. A high temperature jet inflow is centered width- and length-wise at bottom of the domain. The mesh is more refined at the jet to accommodate higher gradients. A schematic of the setup is shown in Figure 4.1(a) and an example temperature field is shown in Figure 4.1(b).

The domain dimensions were chosen to allow the bottom-driven jet to exit the domain primarily vertically due to the vertical inflow velocity and buoyant forces. The jet inflow was centered



(a) Simulation domain

(b) 2D temperature slice (K)

Figure 4.1: (a) Schematic showing the setup of the RANS 3D domain. (b) Temperature field showing a single snapshot in time of a simulation with inlet mean temperature of 1460 K and temperature variability of 0.19. This 2D plane is a slice from the 3D volume; it is parallel to the y-z plane (the normal is in the x-direction) and located at the center of the jet ($x = 0.1875$).

at the bottom of the domain and is 0.25 m wide and is 0.075 m across. The jet velocity, equal to 0.5 m/s, is held constant for all time across all locations and for all simulations. Each computational grid cell of the jet inflow is prescribed a temperature that is allowed to vary spatially and temporally. The mean temperature and the temperature variation are fixed for the duration of each simulation; these two values, temperature mean and temperature variation, comprise the unknown parameters that are to be found using the ABC procedure in the present study. The prior distributions for these two parameters vary for each simulation, and are thus specified in the description of each simulation.

The jet inflow temperature is varied about a fixed mean at each computational grid cell on the jet surface by setting several parameters in OpenFOAM. Specifically, in OpenFOAM the temperature boundary condition at the jet is assigned a built-in type named “turbulentInlet” with a unique reference field and fluctuation scale for each simulation. These settings instruct

OpenFOAM to vary the temperature at each cell on the jet inlet around a set mean. The amount of variation in temperature is drawn randomly from a Gaussian distribution that adds or subtracts a value whose magnitude is determined by the prescribed temperature variability. The variability is defined as the standard deviation divided by the mean; for example, for a simulation with a mean temperature at the inlet of 1500 K, a variability of 0.1 will result in a Gaussian temperature distribution being applied at the inlet with a standard deviation of 150 K. Additionally, these temperatures fluctuate smoothly in time as specified by a correlation parameter α which was set to 0.1. A simulation is run with each set of inflow conditions. The resulting flow field is then compared to a reference simulation. The reference simulation is an execution of the same code with known inflow conditions. The comparison is made using statistics relevant to the flow field at hand, as described in the following section.

4.1.2 ABC Statistic Selection

The ABC method applied is given as algorithm 1 in Section 3.1; this is the basic ABC algorithm with no Markov-chain method (thus the parameters chosen for each run do not affect the parameters chosen for the next computation). The MCMC-ABC algorithm is an area for future research [5]. Choosing a statistic to represent each simulation is a key step in ABC. The statistic must sufficiently reduce the data so a reasonable comparison can be made (most of the time it is not computationally practical to compare the entire flow field at every location in space and time; comparing the entire flow field would be very high dimensionally and would result in the vast majority of simulated parameters being rejected). The statistic must also contain an identifiable signature left by the parameter of interest; if the parameter is unidentifiable or noninfluential, then the technique will not be able to discern the impact of changing the parameter from that measurement [107]. To that end, a variety of metrics were considered for this study.

First, the study sought to determine the mean inflow temperature parameter. One could reasonably assume based on the physics governing the simulation that the mean inflow temperature is likely tightly coupled with the mean temperature in the space above the hot jet. Since the flow

field is turbulent and nonlinear, the link between the inflow temperature and the flow field is most directly linked as close as possible to the inlet. Thus, it is desirable to have measurements of the mean temperature close to the jet surface. Next, the study sought to predict the the inflow temperature variation. Based on the physics, one could reasonably assume that this parameter is linked closely with the temperature standard deviation observed above the jet. As is the case for the first statistic, due to turbulent mixing and the nonlinear behavior of the flow field, a measurement close to the hot jet surface is desirable. To that end, the initial statistics focused on temperature measurements centered above the jet and as close above it as possible. Measurements were taken after the initial transient response had finished 100 seconds into the simulation. Measurements were recorded for the times 100 seconds to 120 seconds every 0.1 seconds.

4.1.3 Turbulent Buoyant Jet Simulation Results: Iteration 1

A set of simulations was performed using the computational domain described in Section 4.1.1. The jet inflow was prescribed a uniform velocity of 0.5 m/s. The mean temperature was assigned a value between 1400 K and 1800 K (inclusive of the bounds). Using a 10 K increment resulted in 41 different temperature values. A temperature of 1600 K represents the reference temperature. The temperature variability was assigned a value between 0 (no variation in temperature) and 0.2 (this means the standard deviation for the Gaussian distribution from which temperatures are drawn has a value of 20% of from the mean for that simulation), inclusive of the bounds. Using a 0.01 increment for the temperature variability resulted in 21 different parameter values. A temperature variability of 0.1 represents the reference value. Together, then, the reference parameters indicate that each cell of the jet will have a mean in time of 1600 K with standard deviation of 160 K. Each combination of temperature mean and temperature variability was computed for 861 unique simulations.

One should note that choosing the prior distribution of initial parameters can be a difficult task. In this case, the truth was known so the prior needed to simply extend above and below the known parameter values. When the truth is unknown, then the prior must be driven by knowledge

of the system – the more knowledge, the narrower the prior. For instance, if the unknown parameter is the inflow temperature, then the prior will be influenced by knowledge of what the fuel is and its associated adiabatic flame temperature (for an upper bound) and some idea of flame temperatures for very rich and very lean mixtures of burning that fuel (for a lower bound).

The ABC methodology was used to compare mean temperatures at 2 mm from the bottom of the domain (i.e. from the jet inflow) and centered width- and length-wise in the domain; this metric was used to predict the reference inlet mean temperature. Additionally, the temperature standard deviation was analyzed and compared at this location; this metric was used to predict the reference inlet temperature variability. Rejection distances of 20 K and 20 K were applied to the temperature mean and temperature standard deviation values, respectively.

Figure 4.2 shows a plot of the posterior for the first parameter, the jet inflow temperature mean. Note that the uniform distribution of initial temperatures morphs into a Gaussian distribution of accepted temperatures centered at 1603.5 K with standard deviation of 15.8 K. All of the temperatures that are very far away from the reference (e.g., 1400 K or 1800 K) are successfully rejected by the algorithm; this indicates that the initial prior was sufficiently wide. Similarly, Figure 4.3 shows a plot of the posterior for the second parameter, the jet inflow temperature variability. Note that, again, the uniform distribution of temperature variability morphs into a Gaussian distribution of accepted temperature variability centered at 0.102 with standard deviation of 0.012. The algorithm correctly rejected all of the temperature variability values that were far from the reference (e.g. 0.0 or 0.2); this indicates that the initial prior was sufficiently wide.

Other sets of measurements were considered to see if they would provide more information to the parameter estimation problem and yield better predictions. For another set of metrics, temperature measurements from other heights were considered. The temperature mean and temperature standard deviation at heights of 11.5 mm, 26.5 mm, and 40.5 mm all were obtained and analyzed in a similar manner to the original data at 2 mm. The temperatures at these additional heights produced posteriors located farther from the reference, with more spread and usually fewer accepted values. Thus, for a single measurement the one closest to the jet inflow appears to provide

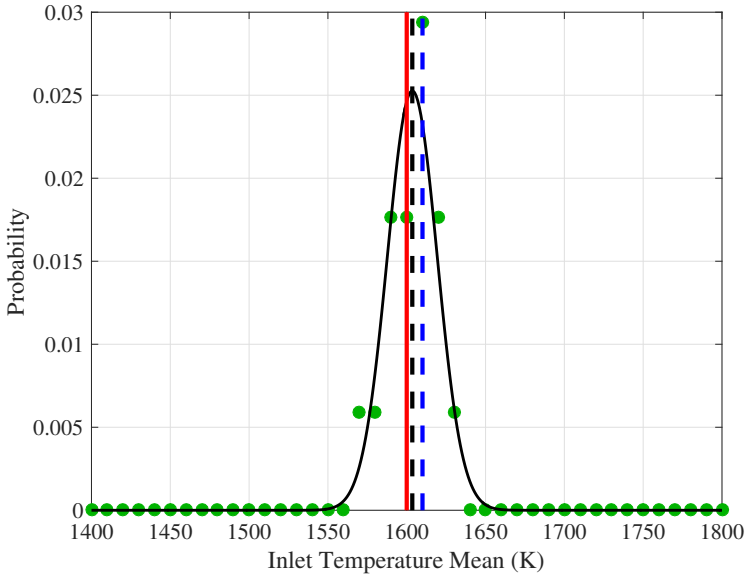


Figure 4.2: Probability distribution showing the marginalized posterior of the inlet temperature mean parameter. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1400 K and 1800 K. The inlet temperature variations were between 0.0 and 0.2. Rejection distances of 20 K and 20 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. Accepted mean temperature values shown by green circles. Reference temperature mean shown in red at 1600 K. The mode of accepted temperatures is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values.

the best information for predicting inflow conditions.

Finally, we considered the original measurement at the bottom location (2 mm) in addition to one of the other measurements. This set of two measurements also yielded posteriors centered farther from the reference, with more spread and usually fewer accepted values. The marginal posteriors are shown in Figures 4.4 and 4.5 for the case of using measurements at two heights. This confirms that the initial set of metrics, informed by the physics governing the flow situation, provides the best estimate of the parameters.

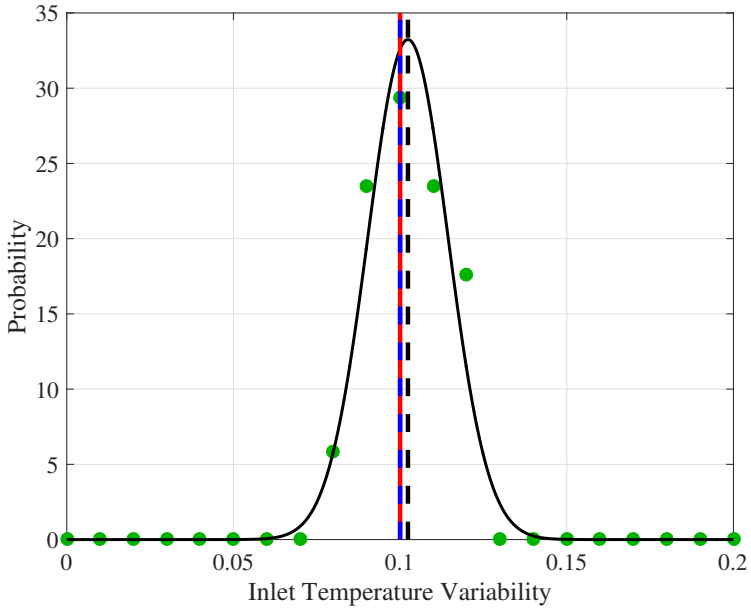


Figure 4.3: Probability distribution showing the marginalized posterior of the inlet temperature variation parameter. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1400 K and 1800 K. The inlet temperature variations were between 0.0 and 0.2. Rejection distances of 20 K and 20 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. Accepted temperature variation values shown by green circles. Reference temperature variation shown in red at 0.1. The mode of accepted temperature variabilities is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values.

4.1.4 Turbulent Buoyant Jet Simulation Results: Iteration 2

Based on the results from the first set of 3D RANS simulations, an additional set of simulations was executed using the same computational domain described in Section 4.1.1 but over a different parameter space. The jet inflow was prescribed the same uniform velocity of 0.5 m/s. The temperature parameters, however, were informed by the first iteration. Recognizing that the initial prior was very wide, this second iteration had a much narrower prior to provide a more refined estimate of the unknown parameters. The bounds for this second set of simulations were driven by the initial results; the prior from the initial set of simulations was truncated to exclude values that

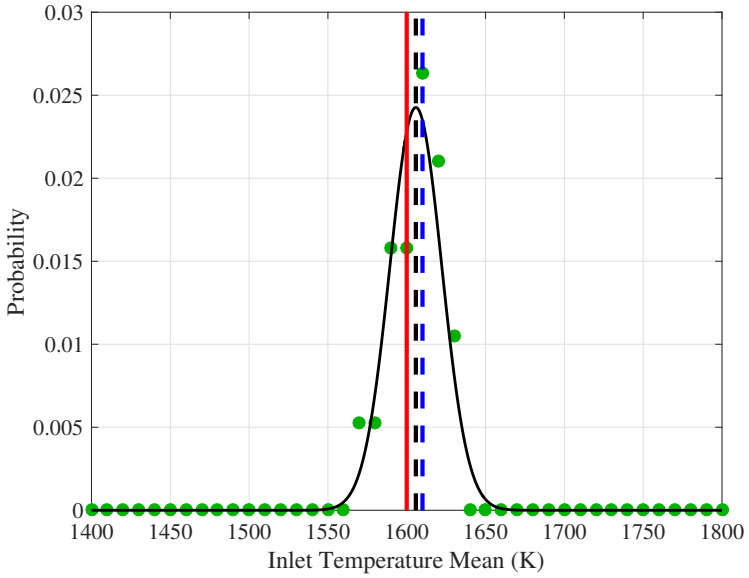


Figure 4.4: Probability distribution showing the marginalized posterior of the inlet temperature mean parameter. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1400 K and 1800 K. The inlet temperature variations were between 0.0 and 0.2. Rejection distances of 26 K and 26 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. *Additionally*, rejection distances of 39 K and 13 K were applied in ABC method to the mean temperature and temperature standard deviation measured farther above the jet (with height equal to 11.5 mm) and centered in the width and length directions. Accepted mean temperature values shown by green circles. Reference temperature mean shown in red at 1600 K. The mode of accepted temperatures is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values.

were determined through ABC to be very unlikely. The mean temperature was assigned a value between 1550 K and 1650 K (inclusive of the bounds). Using a 2.5 K increment (4 times more refined than iteration 1) resulted in 41 different temperature values. A temperature of 1600 K again represents the reference temperature. The temperature variability was assigned a value between 0.08 (up to 8% temperature variation from the mean for that simulation) and 0.12 (up to 12% temperature variation from the mean for that simulation), inclusive of the bounds. Using a 0.002 increment (5 times more refined than the first iteration) resulted in 21 different parameter values. A temperature variability of 0.1 represents the reference value. Each combination of temperature

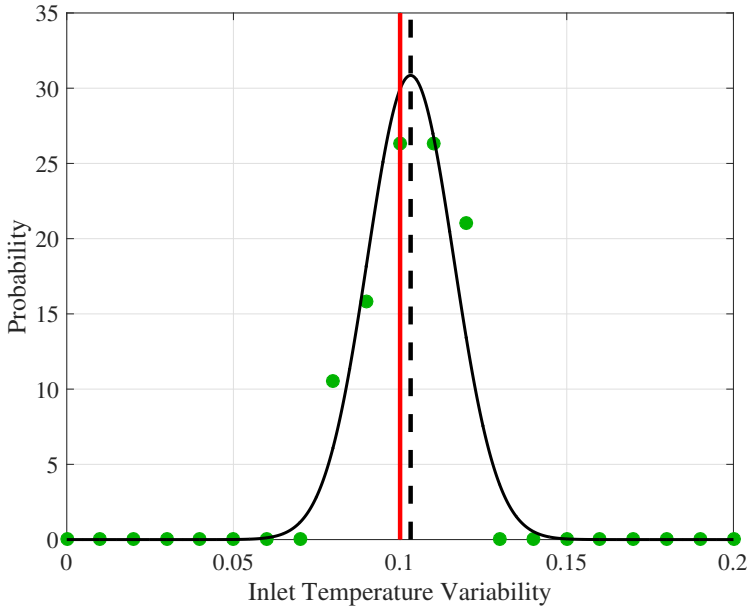


Figure 4.5: Probability distribution showing the marginalized posterior of the inlet temperature variation parameter. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1400 K and 1800 K. The inlet temperature variations were between 0.0 and 0.2. Rejection distances of 20 K and 20 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. *Additionally*, rejection distances of 39 K and 13 K were applied in ABC method to the mean temperature and temperature standard deviation measured farther above the jet (with height equal to 11.5 mm) and centered in the width and length directions. Accepted temperature variation values shown by green circles. Reference temperature variation shown in red at 0.1. The mode is not shown as there is no single mode. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values.

mean and temperature variability was thus once again computed for 861 unique simulations.

The ABC methodology was used to compare mean temperatures at 2 mm from the bottom of the domain (i.e., from the jet inflow) and centered width- and length-wise in the domain; this metric was used to predict the reference inlet mean temperature. Additionally, the temperature standard deviation was analyzed and compared at this location; this metric was used to predict the reference inlet temperature variability. Rejection distances of 12.5 K and 6.5 K were applied to the temperature mean and temperature standard deviation values, respectively. Figure 4.6 shows a plot

of the posterior for the first parameter, the jet inflow temperature mean. Note that the uniform distribution of initial temperatures morphs into a Gaussian distribution of accepted temperatures centered at 1588.5 K with standard deviation of 10.5 K. All of the temperatures that are far away from the reference (e.g., 1550 K or 1650 K) are successfully rejected by the algorithm; this indicates that the initial prior was sufficiently wide based on knowledge of the initial simulation. Similarly, Figure 4.7 shows a plot of the posterior for the second parameter, the jet inflow temperature variability. Note that, again, the uniform distribution of initial velocities morphs into a Gaussian distribution of accepted velocities centered at 0.103 with standard deviation of 0.006. The algorithm correctly rejected all of the temperature variability values that were far from the reference (e.g. 0.08 fluctuation or 0.12 fluctuation); this indicates that the initial prior was sufficiently wide.

The posterior shown in Figure 4.6 has a mean value farther from the reference value than the initial analysis with a wider prior. This is counter intuitive since the parameters used to generate it are much more tightly spaced. This apparent anomaly can be explained by Figure 4.8. This figure shows the prescribed inlet temperature versus the measured temperature at a location used for the above ABC analysis. Note that the mean temperature value observed for the reference case is skewed left compared to the other simulations run with the same inlet temperature. Thus, the mean temperature observed for the reference case appears, on average, more similar to cases executed with lower inlet mean temperatures resulting in the biased posterior of Figure 4.6. This result is not entirely surprising considering the nonlinear behavior of the Navier-Stokes equations; a small change in inlet temperature can result in nonlinear changes in the resulting flow field. Additionally, the stochastic nature of the jet can lead to periods with higher or lower mean temperatures than the reference value. This serves as a valuable warning about the nature of the data used as a reference; if it is biased, then the solution will also likely be biased. It is also important to put this bias in perspective, however, since the observed posterior is still close to the reference value and has less than 1% error indicating ABC still identified the correct parameter within the posterior.

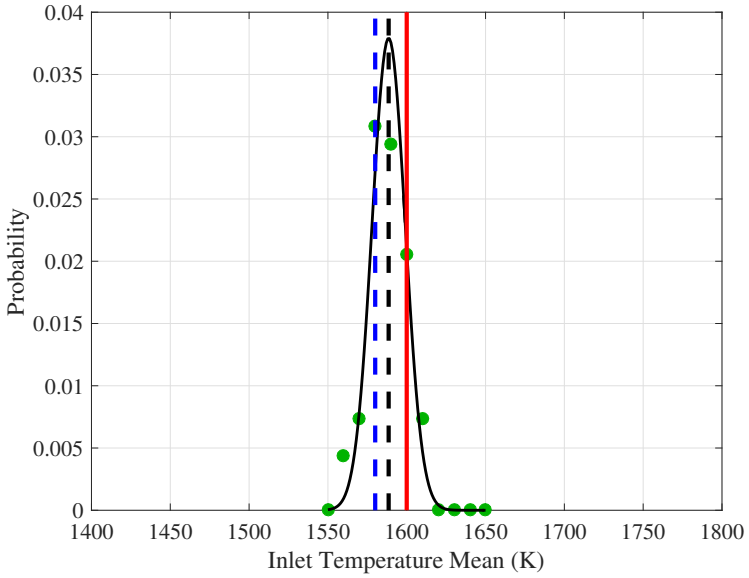


Figure 4.6: Probability distribution showing the marginalized posterior of the inlet temperature mean parameter for *Iteration 2*. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1550 K and 1650 K. The inlet temperature variations were between 0.08 and 0.12. Rejection distances of 12.5 K and 6.5 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. Accepted mean temperature values shown by green circles. Reference temperature mean shown in red at 1600 K. The mode of accepted temperatures is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values.

4.2 Approximate Bayesian Computation for a Turbulent Buoyant Jet with Rotating Cylinder

With the open case now analyzed (i.e. the burner alone with nothing above it), we next turn to another industrial application, that of a turbulent buoyant jet impinging on a rotating cylinder above it. Looking more closely at this configuration, ABC can be used to determine many unknown parameter values of interest. For an example observation, laser diagnostic methods [45, 134, 135] could be used to obtain line-of-sight average temperature or H_2O concentration measurements at various locations within the flow field. These measurements have rich temporal data that could be used to generate a discrete probability density function with associated relevant statistics (e.g.,

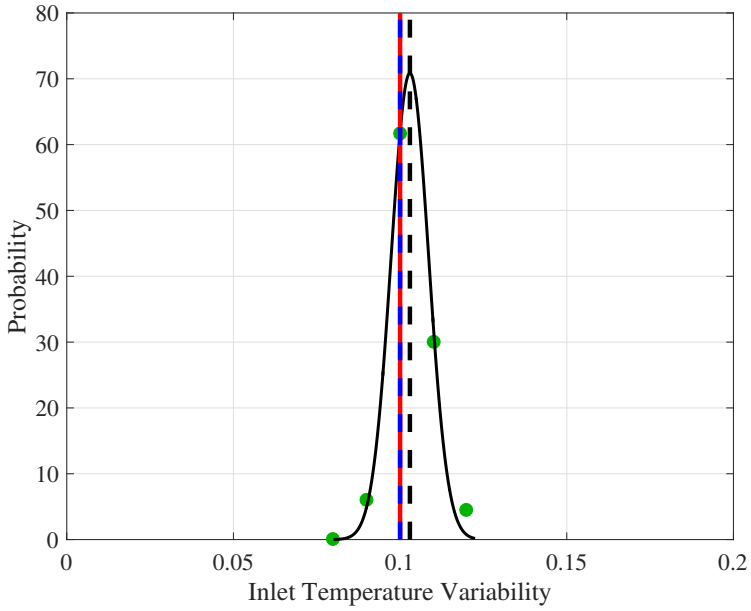


Figure 4.7: Probability distribution showing the marginalized posterior of the inlet temperature variation parameter for *Iteration 2*. A total of 861 simulations were run with unique combinations of inlet temperature mean and inlet temperature variability. The inlet temperature mean for each simulation was chosen to be between 1550 K and 1650 K. The inlet temperature variations were between 0.08 and 0.12. Rejection distances of 12.5 K and 6.5 K were applied in ABC method to the mean temperature and temperature standard deviation measured just above the jet (with height equal to 2 mm) and centered in the width and length directions. Accepted temperature variation values shown by green circles. Reference temperature variation shown in red at 0.1. The mode of accepted temperature variabilities is shown with the dashed blue line. The solid black line is a Gaussian curve with the mean of the accepted values (shown with dashed black line) and standard deviation of the accepted values.

time-averaged temperature and associated variance); this is $P(D)$ in Bayesian terms. Next, a computational model would simulate the experimental setup. The values for parameters of interest (e.g., boundary conditions including inlet temperature and velocity, cylinder temperature, absorptivity and rotational velocity, convection coefficient, species concentration, etc.) would then be chosen according to a prior distribution, $P(\theta)$. Each draw of parameters would produce a separate solution from a RANS simulation whose statistics would be analyzed and compared to the experimental data. As described previously, if the statistics agree according to a predefined threshold, the parameters are kept, or otherwise rejected. From many simulations, a posterior distribution will

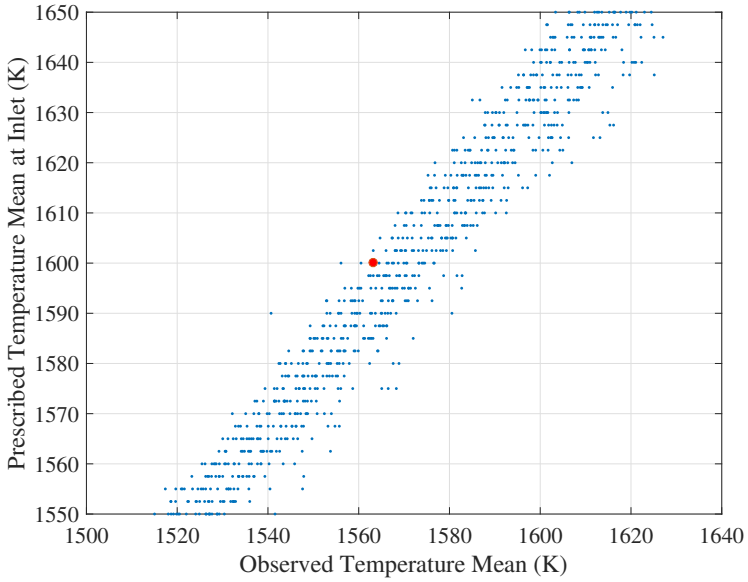


Figure 4.8: The plot's vertical axis shows the prescribed inlet mean temperature that is assigned and held constant for each simulation. The horizontal axis shows the observed temperature; this value is the time averaged value of the temperatures measured at the center of the inlet (width and length wise) at a height of 2 mm over the last 20 seconds of the simulation. The reference simulation, with a prescribed mean of 1600 K, is indicated with the red circle.

emerge indicating which parameters (θ) are probable given the data D; this is, again, the posterior, $P_{ABC}(\theta|D)$.

In this study, originally reported in [85], we have chosen to couple ABC with RANS simulations, instead of using a full Bayesian approach with approximate surrogate models. Moreover, observations from the case we wish to match come from a reference simulation using the same RANS model with known parameters; using the same model for the reference observations and comparison simulations provides the groundwork to estimate parameters for a complex flow simulation using ABC with limited reference data. This work is a precursor to the ultimate goal of combining RANS (or LES) simulations with experimental data to improve simulation accuracy.

4.2.1 Approximate Bayesian Computation Algorithm

In the present study, the ABC algorithm given in Section 2.2 was implemented to predict unknown parameter values in simulations of a rotating cylinder above a high-temperature turbulent buoyant jet. The first step in the analysis was to perform a reference RANS simulation using known inflow velocity and composition and cylinder rotational velocity. In the first study presented in Section 4.2.2, the inflow and cylinder rotational velocities comprised the parameters to be found. In the next study shown in Section 4.3, the inlet ratio of fuel and air served as the unknown parameter to see how well ABC could performing in determining species concentrations. These parameters served as reference parameters that would be determined by the ABC approach. The temperature and velocity at several points within the flow field were stored every 0.1 seconds for the first case, while the H_2O and temperature were measured for the second case in Section 4.3. Next, assuming the reference parameters (inflow velocity and cylinder velocity for the first study and inlet composition for the second study) were unknown, a range of parameters was chosen and many different simulations were performed. The jet and cylinder rotational velocities, along with inlet composition, simulated were chosen from a 'prior' estimation for what the true values might be. The basic idea underlying the ABC approach is that, based on system intuition, physical constraints, previous experience, etc., a researcher typically has a range of probable values that a parameter might take on; these upper and lower limits bound the cases for which simulations need to be run. The 'prior' distribution for this chapter was chosen based on reasonable assumptions found from experimental measurements of a similarly sized system. Once the reference case and a subsequent model case, with parameter values chosen from the prior distribution, are complete, the next step is to study the flow fields of each case. The statistics of each simulation are compared to statistics from the reference case. If the results are similar, the simulated jet and cylinder rotational velocities are stored as a possible candidate pair of parameter values. If the results are not similar, then that combination of velocities is discarded. Another simulation is then run with a new pair of parameters and the process repeats until a sufficiently high number of accepted parameter values

exists from which conclusions can be made about the results.

The ABC algorithm applied in this section is known as method D in Marjoram *et al.* [2] and is detailed in Section 2.2. Note that in the present study, the parameters, θ, that one seeks to find are a pair containing one inlet velocity and one cylinder rotational velocity for Section 4.2.2 and a fuel-to-air ratio for Section 4.3. The distributions of accepted θ values comprise the ABC posteriors, $P_{ABC}(\theta|D)$.

4.2.1.1 Selection of Rejection Distance, ε

In ABC, once the summary statistic, $\mathcal{S}$, is chosen and a distance calculated between each simulation and the reference case, the next task is to determine how small the 'rejection distance,' ε, should be to ensure that each result is sufficiently close enough to the reference case in order to be accepted. Traditionally this task is addressed by a variety of techniques, often relying on expert judgment to balance the spread in accepted parameter values with the number of accepted samples. In Section 4.2.2 we choose values for ε by balancing the precision in the solution (low standard deviation) against having sufficient confidence in the results (high number of accepted samples). Since the reference parameter values are known, this is a reasonable approach. Once the true value of the parameter is unknown, however, it becomes less clear how to determine how well this approach converges. Therefore, in Section 4.3 we introduce a novel approach of choosing an ε value that minimizes the confidence interval of the posterior found by sweeping across a wide range of ε values. The confidence interval is proportional to the standard deviation of accepted parameter values divided by the square root of the number of accepted values. The range of ε values must be wide enough so that the smallest value accepts very few of the proposed parameter values while the largest ε accepts nearly all of the proposed parameter values. For very small values of ε the majority of the cases are rejected resulting in a very small standard deviation of the posterior. As ε increases, the number of accepted samples increases initially reducing the confidence interval width. As ε continues to increase, however, the standard deviation starts to increase faster than can be compensated for by the larger sample size (recall the sample size is to the 1/2 power), so

the confidence interval width increases until all of the simulations are accepted at which point it remains constant. The minimum confidence interval width observed corresponds to a ε value that strikes a balance between having high precision in the posterior (i.e. low standard deviation) and high confidence in the posterior (i.e. a large number of accepted samples).

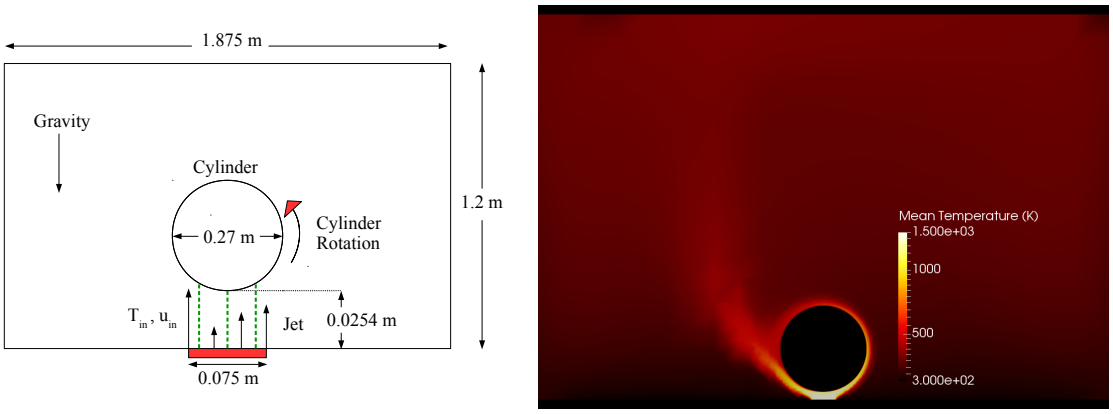
4.2.2 Determination of Jet and Cylinder Velocities

The present study uses ABC for two tasks: (*i*) determining jet and cylinder rotational velocities and (*ii*) determining the chemical composition of gases exiting the jet. The ABC approach in both cases is based on both test and reference data taken from RANS simulations. This section describes results for the first task and Section 4.3 describes results for the second task.

4.2.3 Simulation Setup

For the study of jet and cylinder rotational velocities, the buoyantPimpleFoam solver in the open-source computational fluid dynamics software OpenFOAM [119, 120, 132] was used to solve the compressible RANS equations in conjunction with the energy equation. The ideal gas equation of state was used to relate thermodynamic variables. Heat transfer mechanisms modeled in this simulation include convection from the cylinder surface, advection of temperature by hot gases, and radiation exchange between domain boundaries. The Menter SST k-ω two-equation eddy viscosity turbulence closure model [136] was chosen for its robustness and good performance at different length scales. The RANS equations were solved with second order accuracy in space and a blend of first and second order accuracy in time. Limiters on gradient and divergence schemes were implemented to aid convergence; under-relaxation was also used to assist solution convergence. Fluid viscosity and specific heat were assumed to be constant for the purpose of this study; the fluid in the domain is assumed to be dry air.

The RANS equations were solved on a 2D grid with a high temperature jet centered along the bottom of the domain. A schematic of the setup is shown in Figure 4.9(a) and an example temperature field is shown in Figure 4.9(b). The domain dimensions of 1.875 m $\times$ 1.2 m were



(a) Simulation domain

(b) Example temporally averaged temperature field

Figure 4.9: (a) Schematic showing the setup of the RANS simulation. Measurements taken along the three green vertical dashed lines. (b) Temperature field averaged over the last 15 s of the simulations with jet velocity 1 m/s and rotational velocity of 12.5 rad/s.

chosen to allow high temperature gases to exit the top of the domain primarily vertically, where the flow within the domain is comprised of contributions due to the jet inflow, velocity imparted by the rotating cylinder, and buoyant forces. The jet is centered at the bottom of the domain and is 0.075 m wide. In these simulations, the temperature and velocity of the jet inflow are assumed to be constant over the entire jet surface, where the jet temperature for all simulations in this section was held fixed at 1500 K. The jet is centered under a 10.5 inch diameter cylinder, with a 2.54 cm (1 inch) gap separating the jet and cylinder for the results in this section. Note that the gap is reduced to 1.27 cm (0.5 inch) in Section 4.3. The cylinder rotates counterclockwise about its center.

Fixed boundary conditions were imposed for the jet inlet and cylinder surface. The cylinder temperature and angular velocity were fixed, as were the jet inlet temperature and vertical velocity. All other domain boundaries were left open and allowed flow into or out of the domain. This is of particular interest when considering the bottom boundary, as fluid entrainment can become an important factor in fluid motion near the jet and cylinder. Pressure and temperature boundary conditions were used to ease relaxation to ambient conditions at domain boundaries. The RANS

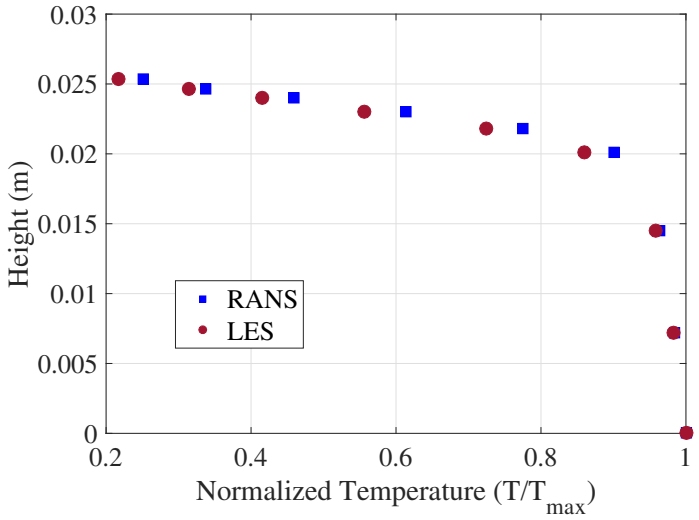


Figure 4.10: Plot shows height above the hot jet versus the temporal average of simulation temperature at the jet centerline with fixed rotational velocity for the cylinder and fixed jet inlet conditions. The red circles correspond to LES while the blue squares are from RANS data; both simulations have the same boundary conditions and were run using OpenFOAM version 2.2.1.

simulation compares favorably with a three-dimensional large-eddy simulation within the region of interest, as shown in Figure 4.10.

4.2.4 Approximate Bayesian Computation Setup

The ABC algorithm used to determine the jet and cylinder rotational velocities is the same as that described in Section 4.2.1. Each simulation was run with a specified and constant velocity at the jet inlet and for the cylinder. The resulting flow fields were then compared to a reference simulation. The reference simulation was an execution of the same code with known conditions corresponding to a jet velocity of 0.5 m/s and a cylinder angular velocity of 12.5 rad/s. Comparisons were made using statistics relevant to the flow field at hand, such as the mean temperature at a given location over the duration of the simulation, variance of the temperature, or a Hellinger distance to compare full temperature probability density functions.

For the ABC analysis, three different sets of simulations were run each with a different 'prior' distribution of jet inflow and cylinder rotational velocities (i.e. the span of parameters and

Table 4.1: ABC setup for baseline, refined, and further refined cases. This shows the range and increment for each parameter along with the total number of simulations in each case. Note that moving down in the table increases the refinement within the parameter space so that parameters simulated are closer together.

	Jet Inlet Velocity (m/s)		**Cylinder Rot. Vel (rad/s)**		**Number of**
Case	**Range**	**Increment**	**Range**	**Increment**	**Simulations**
Baseline	0.0–1.0	0.05	3–28	0.5	1071
Refined	0.3–0.7	0.005	7.5–17.5	0.1	8181
Further Refined	0.48–0.52	0.001	10–15	0.05	4141

increment of parameters are different for each case). A summary of the three cases is found in Table 4.1. The 'baseline' case sweeps across a wide set of parameters. Its results drive the 'refined' case's range of parameters while the 'refined' case's results inform the final 'further refined' case. These three cases show how the posterior distribution narrows and becomes smoother as the density of simulations close to the reference parameters increases.

As shown in Figure 4.9(b), a flow of hot gas enters the domain and then diffuses while flowing around the rotating cylinder. Measurements were taken 0.1 seconds apart in the gap between the cylinder and the jet. Measurement locations were at the jet centerline (i.e., the line extending vertically from the center of the jet) and two transverse locations 1 cm from either end of the jet (indicated by the green dashed lines in Figure 4.9(a)). Measurements were recorded at approximately 10 separate heights between the jet and cylinder for each of the three locations.

Figure 4.11 shows a sample temperature measurement recorded during a simulation with inlet velocity of 0 m/s and rotational velocity of 3 rad/s at a height of 4.1 mm above the jet inlet and centered in the domain. Note that the first 15 seconds of each simulation were ignored in order to eliminate the impact from transient startup effects. This period of time was determined through qualitative analysis of the wake behind the cylinder and by observing the temperature fluctuations within the gap between the cylinder and jet for bounding cases; the moving average shown in Figure 4.11 stabilizes after 15 seconds into the simulation.

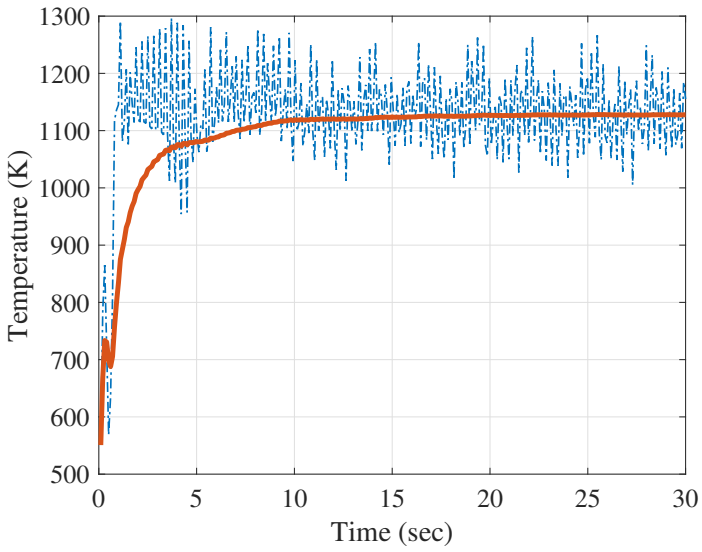


Figure 4.11: Temperature versus time along the jet centerline at a height of 4.1 mm above the jet. The dash-dot blue line indicates the instantaneous temperature, while the solid red line shows the moving average.

4.2.5 Selection of Simulation Statistics and Metrics

As will be shown in the following, the ABC method has been used to successfully identify the prescribed parameter values (0.5 m/s for inlet velocity and 12.5 rad/s cylinder rotational velocity). This success, however, required the selection of suitable statistics with which to analyze and determine the 'distance' from the reference data. The distance is calculated as the absolute value of the difference between the synthetic statistic and the reference statistic, namely $\delta(\mathcal{S}, \hat{\mathcal{S}}) = |\mathcal{S} - \hat{\mathcal{S}}|$. Here we have used three statistics:

(1) Time-averaged temperature at a single height close to the jet exit along the jet centerline. This is denoted Statistic 1.

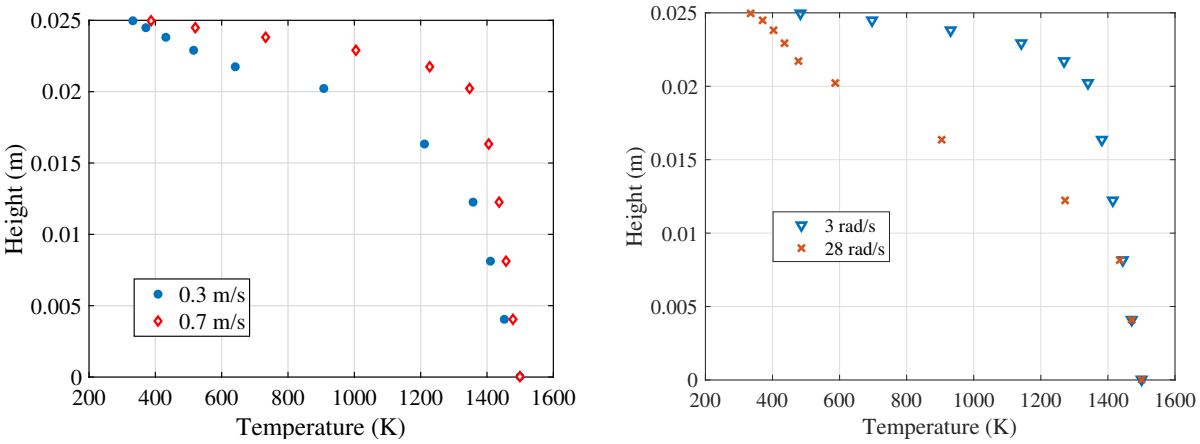
(2) Space- and time-averaged temperature over all available measurement heights along the jet centerline in the gap between the jet and cylinder. This is denoted Statistic 2.

(3) Time-averaged temperature at a single height close to the cylinder surface along the jet centerline. This is denoted Statistic 3.

The first statistic was specifically chosen to allow determination of the jet velocity using the ABC method. For the present geometry and setup, an increase in inlet velocity corresponds to an increase in measured temperature above the jet, as shown in Figure 4.12(a). This is due to the momentum imparted by the jet to the flow and its ability to carry high-temperature gases further into the domain. Thus, by comparing the mean temperature in the domain at a particular location, preferably close to the jet so that the influence of the rotating cylinder on the flow field will be minimized, one can readily predict a reference inlet velocity. In the present study, Statistic 1 was computed using measurements at a height of 4.1 mm above the jet.

In order to gain more insight into how the cylinder rotation impacts the flow field, more complicated metrics are required to be used in addition to Statistic 1 (which is always required to determine the jet inlet velocity). The temporal means of temperature at each measured height for the slowest rotational rate (3 rad/s) and the highest rotational rate (28 rad/s) are shown in Figure 4.12(b). Very near the jet exit, there is essentially no effect of rotation rate on the observed temperature and, thus, a single point measurement close to the jet (i.e., Statistic 1) is insufficient to constrain predictions of the cylinder velocity. Further above the jet, however, higher temperatures are observed when the cylinder is spinning more slowly. This occurs because, when the cylinder spins quickly, its boundary layer entrains more of the cold surrounding air and the wake separation moves closer to the hot jet, thereby moving more cool air into the gap between the jet and cylinder. Conversely, when the cylinder spins slowly, its boundary layer entrains less cold air, the wake separation moves further from the hot jet, and thus the temperature in the gap is dominated by the hot jet. With these insights, Statistic 2 was developed to compare each simulation to the reference case. This metric is calculated by taking the spatial and temporal mean of temperature over all available measurement heights along the jet centerline in the gap between the jet and cylinder. Statistic 2 was computed using measurements at heights of 4.1, 8.1, 12.3, 16.3, 20.2, 21.7, 22.9, 23.8, 24.5, and 25.0 mm above the jet.

It should be noted, however, that Statistic 2 introduces substantial storage requirements since a large amount of data is retained from the simulations. Additionally, the experimental burden to



(a) Jet velocity variation

(b) Cylinder velocity variation

Figure 4.12: Temporal average (over 15 s) of temperature as a function of height along the jet centerline. Panels show: (a) Variations in jet velocity, where blue circles correspond to a jet velocity of 0.3 m/s while red diamonds correspond to a jet velocity of 0.7 m/s, with a cylinder angular velocity of 12.5 rad/s in both cases; (b) Variations in cylinder velocity, where blue circles correspond to a cylinder angular velocity of 3 rad/s while red diamonds correspond to an angular velocity of 28 rad/s, with a jet velocity of 0.5 m/s in both cases.

obtain this many simultaneous temperature measurements would be immense. Therefore, a reduced set of measurements is desired that produces comparable posterior parameter distributions. In the present study, this set combines a measurement close to the jet along the jet centerline, at a height of 4.1 mm (Statistic 1 to determine inlet velocity), and one additional measurement closer to the cylinder, also centered and at a height of 25.0 mm above the jet exit. This is denoted Statistic 3.

For Section 4.2.6, pairs of jet and cylinder rotational velocities were retained or rejected based on the agreement between simulated statistics and reference statistics. All 'rejection distances' (ε) were chosen per guidance of Section 4.2.1.1 to reduce the number of accepted cases converging toward the reference solution, while simultaneously leaving a sufficiently high number of accepted cases to maintain confidence in the results. All results shown in Section 4.2.6 use metric 1 in combination with either metric 2 or metric 3.

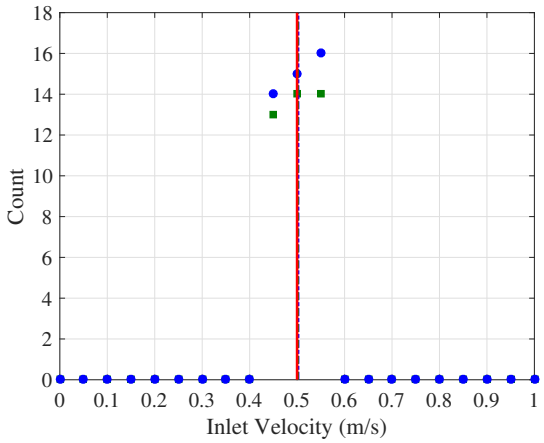
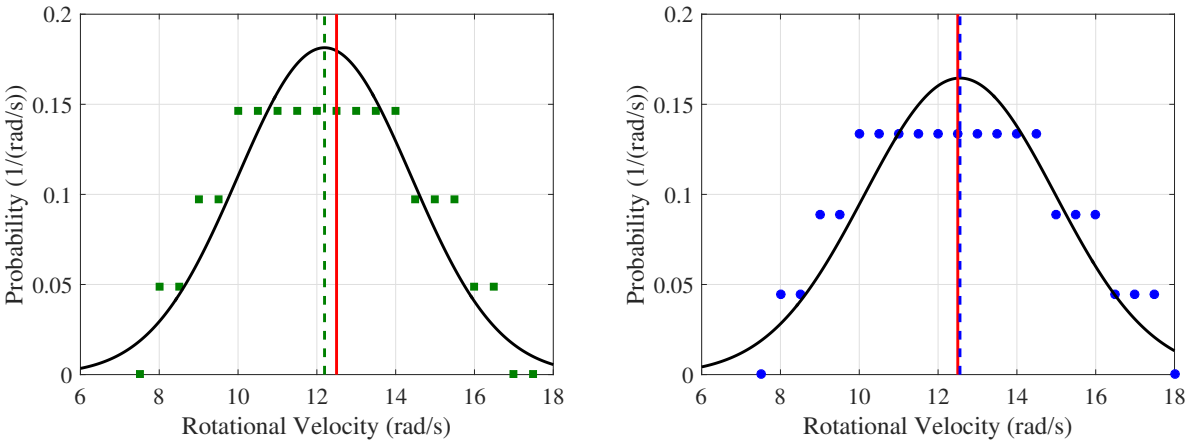


Figure 4.13: Number of accepted simulations versus jet inflow velocity for 'baseline' case. Here 1071 simulations are performed over uniform priors with widths 0.0–1.0 m/s and 3–28 rad/s. The reference velocity at 0.5 m/s is shown in red with the mean of accepted values using Statistic 1 combined with Statistic 2 (all measurement locations) shown in green at 0.501 m/s, while the mean of the accepted values using Statistics 1 and 3 (two measurement locations) is shown in blue at 0.502 m/s (both green and blue lines are obscured by the red line).

4.2.6 Results: Finding Inlet and Rotational Velocities

Figure 4.13 shows the number of 'accepted' simulations for different values of the jet inflow velocity. Although a uniform distribution of jet velocities is used in the simulations, spanning the range from 0.0 m/s to 1.0 m/s, the resulting histogram of accepted values shrinks to a much more narrow and uniform distribution centered on the reference value of 0.5 m/s. For Statistic 1, a rejection distance of 3.5 K was applied when comparing the temporal mean at one location in each simulation to the corresponding mean from the reference simulation. For Statistic 2, a rejection distance of 95 K was applied. Finally, for Statistic 3, the rejection distance was 125 K. Using these rejection distances, all jet velocities below 0.45 m/s and above 0.55 m/s have been rejected by the ABC algorithm. The mean of the accepted jet velocities is 0.501 m/s, giving an error of less than 1% compared to the reference value.

Similarly, Figure 4.14 shows probability density functions (pdfs) of 'accepted' simulations for different values of the cylinder rotational velocity, using Statistic 1 combined with either Statistic 2



(a) All vertical positions

(b) Two vertical positions

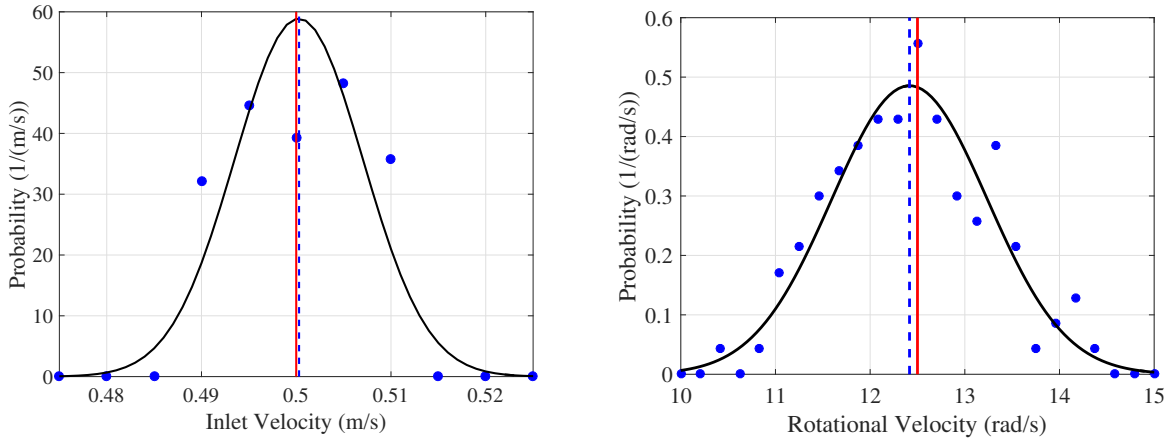
Figure 4.14: Probability density functions (pdfs) for the number of accepted simulations versus values of cylinder rotational velocity. Solid black lines show Gaussian curves with the same mean and standard deviation as the accepted values. Here in the 'baseline' case 1071 simulations are performed over uniform priors with widths 0.0–1.0 m/s and 3–28 rad/s. Reference rotational velocity shown in red at 12.5 m/s with the mean of the accepted values shown in green at 12.2 rad/s for case (a) and in blue at 12.6 for case (b).

or Statistic 3. Once again, the uniform distribution of initial cylinder rotational velocities morphs into much narrower distributions of accepted velocities centered around a mean of 12.2 rad/s for Statistic 2 and a mean of 12.6 rad/s for Statistic 3, giving errors of 2.4% and 0.8% compared to the reference value of 12.5 rad/s. This demonstrates that combined with Statistic 1, both Statistic 2 or Statistic 3 successfully identify the rotational velocity correctly. However, for the following studies, only Statistic 1 combined with Statistic 3 will be used due to decreased storage requirements.

Although the results in Figures 4.13 and 4.14 provide informative initial predictions of the distribution of likely parameter values for both the jet and cylinder velocities, the prior distribution of parameters were chosen to be wide enough to ensure the reference value would reside somewhere within its upper and lower bounds. As a result, the posterior parameter distributions are relatively wide and coarse. Thus, a subsequent set of simulations was executed with a narrower prior. The 'refined' width of the prior for this next set of simulations was determined based on the accepted range of parameters from the posteriors shown in Figures 4.13 and 4.14. Specifically, inlet velocities

ranged from 0.3 m/s to 0.7 m/s with a 0.005 m/s increment (10 times more resolved than original increment) while cylinder rotational velocities ranged from 7.5 rad/s to 17.5 rad/s with an interval of 0.1 rad/s (5 times more resolved than original increment). This parameter range and associated increment results in a prior distribution with 8,181 pairs of parameters (and consequently that many additional simulations). The rejection distances (ε) used were reduced compared to the initial study; this is permissible due to the high density of parameter values (i.e. with a smaller rejection distance one still maintains a sufficiently high number of accepted cases). The specific rejection distances used were 0.6 K for the temporal mean temperature comparison just above the jet inlet (Statistic 1) and 10 K for the temporal mean temperature comparison just below the cylinder (Statistic 3). Note for this 'refined' case, and the 'further refined' case, only two locations were used based on the success demonstrated using a combination of Statistic 1 and Statistic 3 for the 'baseline' case. Because of the increased density of parameter values and smaller rejection distances applied, narrower posterior distributions emerge indicating increased confidence; the posteriors are shown in Figure 4.15. Here the uniform distributions of initial jet inlet and cylinder rotational velocities morph into much narrower distributions of accepted velocities centered around means of 0.5003 m/s and 12.42 rad/s, giving errors of 0.06% and 0.64% compared to the reference values of 0.5 m/s and 12.5 rad/s, respectively.

To test how well the posterior distribution would converge, a final set of simulations was run with a 'further refined' prior distribution of parameters. Now, in an iterative fashion, the width of the prior distributions were chosen based on the refined posterior distributions provided in Figure 4.15. Specifically, inlet velocities ranged from 0.48 m/s to 0.52 m/s with a 0.001 m/s increment (50 times more resolved than original increment) while cylinder rotational velocities ranged from 10 rad/s to 15 rad/s with an interval of 0.05 rad/s (10 times more resolved than original increment). This parameter range and associated increment results in a prior distribution with 4,141 pairs of parameters (and consequently that many additional simulations). The rejection distances (ε) used were reduced compared to the initial study; this is permissible due to the high density of parameter values. Because of the increased density of parameter values, narrower posterior distributions

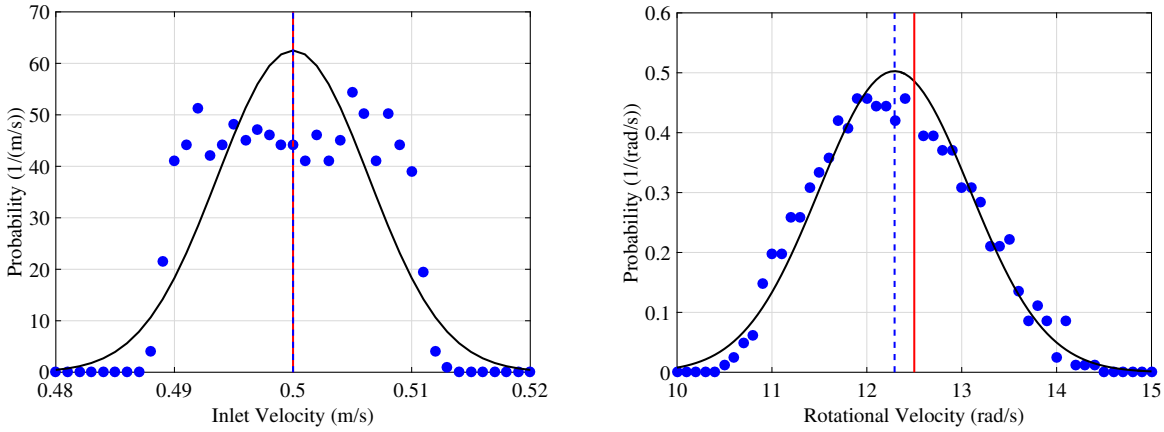


(a) Jet inlet velocity

(b) Cylinder rotational velocity

Figure 4.15: Probability density functions for the number of accepted simulations versus all simulated values of jet inlet velocity each with a pair of unique uniform velocity applied across the jet and cylinder rotational velocity; this is the 'refined' case. Here 8081 simulations are performed over uniform priors with widths of 0.3–0.7 m/s and 7.5–17.5 rad/s. Reference velocity shown in red at 0.5 m/s with the mean of the accepted values using only two measurement locations for metric 2 is shown in blue at 0.5003 m/s. Reference rotational velocity shown in red at 12.5 m/s with the mean of the accepted values shown in blue at 12.42 rad/s.

emerge containing many more accepted simulations indicating increased confidence; the posteriors are shown in Figure 4.16. Here the uniform distributions of initial jet inlet and cylinder rotational velocities morph into much narrower distributions of accepted velocities centered around means of 0.500001 m/s and 12.29 rad/s, giving errors of much less than 1% compared to reference inlet velocity of 0.5 m/s and 1.7% compared to the reference rotational velocity of 12.5 rad/s. Note that the mean accepted value for the inlet velocity is extremely close to the reference value while the estimated rotational velocity is actually worse than the two previous cases with less refined prior distributions of parameters. This is likely because the reference case produced statistics that are more representative of cases with slower rotational velocities; said differently, the statistics may be biased for the reference data causing bias in the resulting posterior, also. This is a natural occurrence due to the stochastic nature of the simulations and the fact that the summary statistics do not fully represent the underlying data. This provides a word of caution to users of ABC



(a) Jet inlet velocity

(b) Cylinder rotational velocity

Figure 4.16: Probability density functions for the number of accepted simulations versus all simulated values of jet inlet velocity each with a pair of unique uniform velocity applied across the jet and cylinder rotational velocity; this is the 'further refined' case. Here 4141 simulations are performed over uniform priors with widths of 0.48–0.52 m/s and 10–15 rad/s. Reference velocity shown in red at 0.5 m/s with the mean of the accepted values shown in blue at 0.500001 m/s. Reference rotational velocity shown in red at 12.5 m/s with the mean of the accepted values shown in green with the mean of the accepted values shown in blue at 12.29 rad/s.

that the resulting posterior, while "correct" in the sense that it causes the best match with the available observations, may not be centered on the true parameter given potential shortcomings of available observations. This is one reason why the distribution associated with the parameters is so important. For instance, the inlet velocity prediction is very narrow while the rotational speed is much wider indicating less certainty, as is warranted based on the discrepancy between the predicted and reference parameters.

A joint pdf provides an additional method to gain insight into the pairs of jet inflow velocity and cylinder rotational velocity that provide statistics most similar to the reference statistics. Figure 4.17 shows the joint pdf of both parameters' posterior values for the 'further refined' case. The peak of this joint pdf occurs at 0.5005 m/s and 12.45 rad/s which is consistent with the values shown in Figure 4.16.

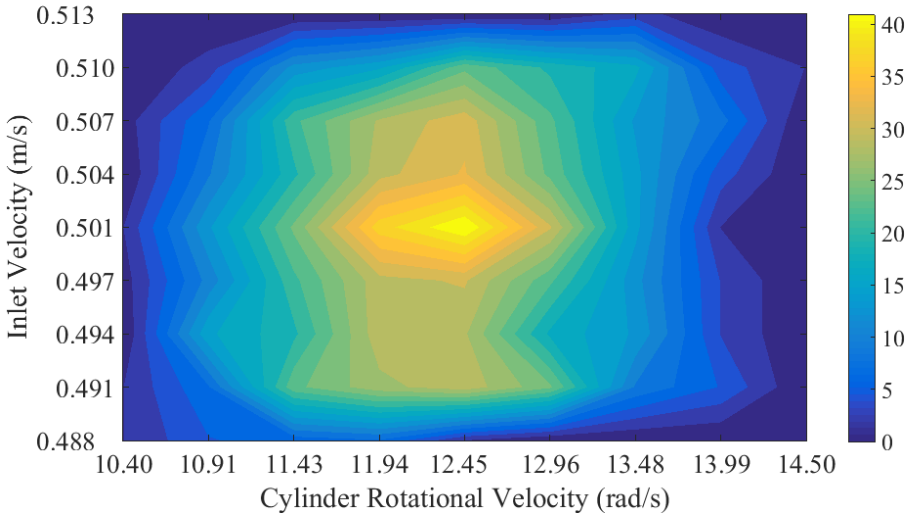


Figure 4.17: Joint probability distribution function (pdf) showing the accepted values for the 'further refined' case. Of the 4141 cases, 975 are accepted. The peak of this joint pdf occurs at 0.5005 m/s and 12.45 rad/s.

4.3 Determination of Jet Inflow Composition

The initial studies in this dissertation have focused on determining unknown boundary conditions related to temperature and velocity; these included the jet inflow velocity and temperature, as well as a moving wall boundary within the domain (i.e. the cylinder rotational velocity), . With these boundary conditions estimated using the ABC method, an additional iteration of ABC was performed to address another boundary condition of interest, namely the ratio of various gaseous species in the jet. A new simulation was created in OpenFOAM to allow species to be advected within the domain. A test case was performed wherein different mass fractions of species enter the domain. The species tracked in this simulation include N_2, O_2, CO_2, and H_2O; these are the dominant species resulting from the combustion of methane fuel (CH_4) with dry air. The quantity of each species was determined by altering the fuel-to-air ratio, ϕ, of a fictitious up-stream 'burner'; the reaction of methane fuel (CH_4) and oxidizer (dry air) is assumed to be complete before entering the domain leaving only products to enter the through the jet plus inert N_2 and any additional O_2 left over when there is insufficient fuel (i.e. for values of ϕ less than one). Values for ϕ ranged from

0.3 (much excess air) to 1.0 (sufficient fuel to react with all of the O_2 found in the air). These values of ϕ were chosen to ensure the adiabatic flame temperature for the reaction was above the 'burner' temperature (set to 1,000 K) and also to ensure all of the CH_4 would theoretically react in the burner before entering the domain. This is a very relevant assumption for industrial applications where the equivalence-ratio is usually set below 1 to ensure no fuel goes unburned. The final test for ABC presented in this chapter, then, is to determine the correct ϕ value for this fictitious 'burner' based on some measurement of either H_2O concentration or temperature at a location mid-way between the inlet and the cylinder. Note that either of these quantities, H_2O concentration or temperature, can be measured using existing techniques [45, 81, 134, 135, 137]. This application of ABC is relevant to experimental instances when H_2O concentrations or temperatures are recorded within a domain, but suspicion exists about the prescribed fuel-to-air ratio due to, for example, experimental procedure, fuel contents, or equipment calibration.

4.3.1 Simulation Setup

The open source computational fluid dynamics software OpenFOAM, version 4.1, was used to design and run RANS simulations for this study [119, 120]. The compressible RANS equations were solved in conjunction with the energy equation using the fireFoam solver [118]. The ideal gas equation was used to relate state variables. The fireFoam solver is very similar to buoyantPimpleFoam used in the Section 4.2.2, however it allows species tracking within the domain. The presence of multiple species, as opposed to only dry air, allowed the fluid viscosity and radiation properties to vary within the domain. These fluid properties change according to which species are present at each location as driven by the fluid dynamics governing the simulation. Heat transfer mechanisms that were modeled in this simulation include convection from the cylinder surface, advection of temperature by hot gases, and radiation exchange between domain boundaries and the fluid. The Menter Shear Stress Transport (SST) Scale-Adaptive Simulation (SAS) k-ω two-equation viscosity closure model [136, 138] was chosen for its robustness and good performance at different length scales. The RANS equations were solved with second order accuracy in space and time to obtain

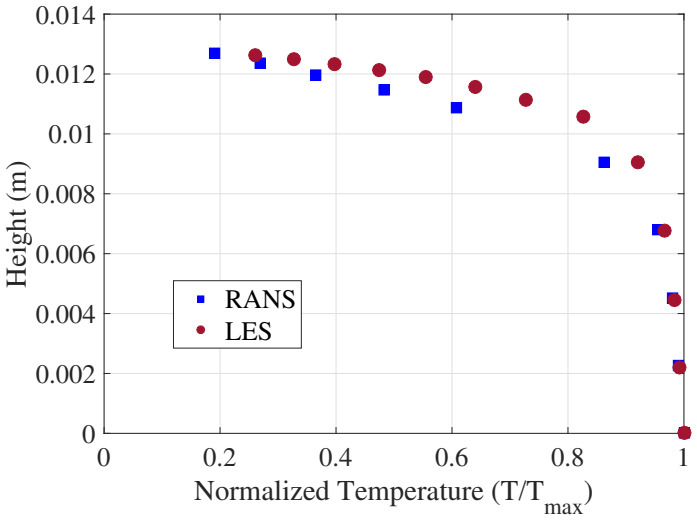


Figure 4.18: Plot shows height above the hot jet versus the temporal average of simulation temperature at the jet centerline with fixed rotational velocity for the cylinder and fixed jet inlet conditions. The red circles correspond to LES while the blue squares are from RANS data; both simulations have the same boundary conditions and were run using OpenFOAM version 4.1.

a solution. Limiters on velocity divergence were implemented to aid convergence; under relaxation also helped solution convergence. Fluid viscosity and specific heat were assumed to vary with temperature according to the Sutherland model [122] and JANAF tables [123], respectively. The RANS simulation compares favorably with a three-dimensional large-eddy simulation within the region of interest, as shown in Figure 4.18.

Within the domain, the description of Section 4.2.3 is quite similar to this final set of simulations however the gap between the inlet and the cylinder has been reduced to 1.27 cm (0.5 in.). Additionally, the inlet velocity is fixed at 0.3 m/s with a temperature of 1,000 K while the rotational velocity of the cylinder is fixed at 18 rad/s for all simulations. For these simulations, instead of assuming dry air filled the domain, the inlet was prescribed appropriate mass fractions of N_2, O_2, CO_2, and H_2O. The quantity of each species was determined based on the simulation's prescribed ϕ value and the associated reaction's equilibrium products as determined using NASA-built software Chemical Equilibrium with Applications (CEA) [139].

4.3.2 Approximate Bayesian Computation Setup

The ABC algorithm used to determine the jet inflow species composition is the same as that described in Section 4.2.1. Each simulation was run with a specified and constant species concentration at the jet inflow. The resulting flow fields were then compared to a reference simulation. The reference simulation was an execution of the same code with known conditions corresponding to a ϕ of approximately 0.65.

4.3.3 Selection of Simulation Statistics

Two metrics were considered in this study, namely H_2O concentration and temperature. The first metric emerges naturally based on its connection to the parameter being changed. This quantity is readily available computationally and experimentally making it an ideal candidate. The ABC procedure was also implemented using comparisons of H_2O concentration within the domain at a single spatial location between the jet inflow and the cylinder; this is directly tied to the species concentration at the boundary, though how it varies in the flow field is not readily known without performing a simulation. Temporal averages were used for all measurements over the last 10 seconds of 20 second simulations; here again we have ignored the transient behavior of the simulations to obtain steady-state statistics. H_2O mass fraction vertical profiles are shown in Figure 4.19; these profiles demonstrate how dramatically the H_2O profile and quantity can change within the domain for different values of ϕ. Of particular interest is how several of the profiles cross higher in the domain at a height of approximately 0.8 cm. Measurements were taken at the jet centerline and at a height of approximately 1.09 cm. Measuring at this height allows the complexities in the fluid domain to fully develop demonstrating how well ABC can perform with less than ideal measurements of H_2O concentration (note that at the jet inflow H_2O measurements are directly coupled to ϕ so predicting ϕ based on H_2O measurements at the jet inflow is trivial). The results of using H_2O concentration measurements to determine ϕ are presented in Section 4.3.4. Comparisons were also made using the temporally averaged temperature within the domain

at the same single spatial location. Compared to H_2O concentration measurements, temperature is not as sensitive to changes in the ϕ parameter. At the jet inflow, temperatures are nearly identical regardless of inlet ϕ; only at higher locations within the domain does ϕ have a noticeable impact on temperature. At the chosen measurement location of 1.09 cm above the burner, temperature provides a relevant observation by which to determine ϕ as shown in Section 4.3.5. This approach, i.e. using temperature measurements to predict inflow equivalence ratio, parallels that of Section 4.2.2 wherein we determined *velocity* boundary conditions using *temperature* measurements.

As described in Section 4.2.1.1, here ε is chosen to minimize the confidence interval half width. Figure 4.20(a) shows the confidence interval half width based on measurements of H_2O concentration versus a range of ε values. The confidence interval half width shown is calculated as the standard deviation of accepted parameter values divided by the square root of the number of accepted values multiplied by a scalar to obtain the 95^{th} percentile. Note that any other percentile could have been chosen resulting in the same suggested ε value to use, however the magnitude of the confidence interval half width would have differed. This method suggests using a rejection distance of 0.009 indicating how close the measured H_2O concentration needs to be compared with the reference H_2O concentration. This value of ε strikes a balance between having high precision in the posterior (i.e. low standard deviation) and high confidence in the posterior (i.e. a large number of accepted samples). Note that this value of ε is quite small relative to how well one might actually be able to experimentally measure H_2O concentration, in which case one might prefer to relax this constraint (i.e. increase ε). Increasing values of ε generally yield posteriors with similar mean values but with larger standard deviations. This is promising as it indicates that using less certain measurements would still yield reliable parameter estimates, albeit with less certainty.

4.3.4 Results: Finding ϕ Using H_2O Concentration Measurement

For this ABC test case, approximately 10,600 simulations were run each with a unique ϕ value randomly drawn from a uniform distribution between 0.3 and 1.0. Each simulation resulted in a unique temporally averaged H_2O concentration profile and temperature profile between the inlet

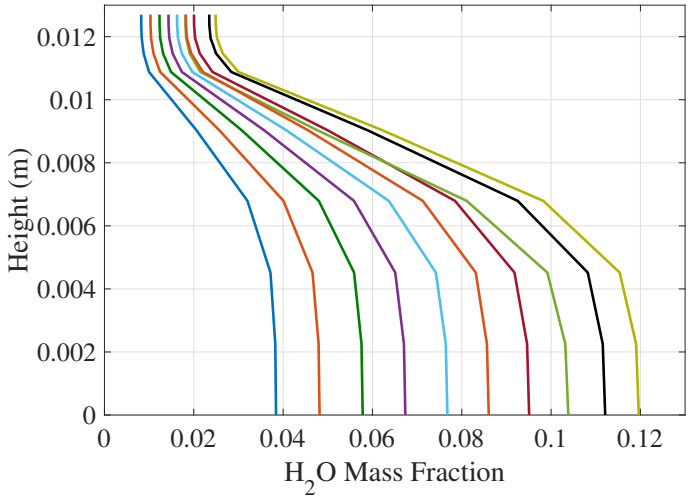
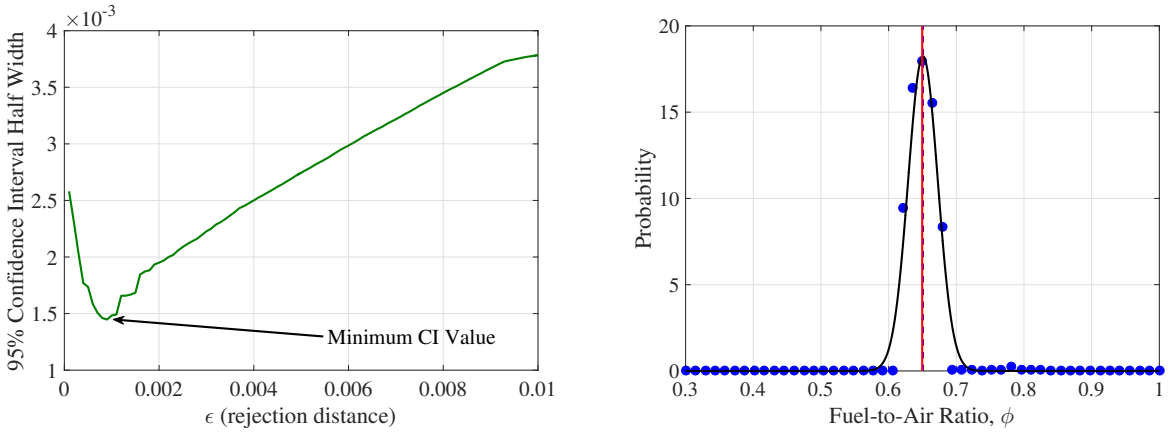


Figure 4.19: Plot shows height above the hot jet versus the temporal average of simulation H_2O mass fraction at the jet centerline with fixed rotational velocity for the cylinder and fixed jet inlet conditions. The ϕ values used to set the inflow conditions were evenly distributed from 0.3 to 1.0. Note that several of the profiles cross higher in the domain indicating a complex behavior exists. Thus, an analytical solution is not practical and simulations are necessary to understand the behavior of the fluid enabling accurate parameter estimation.

and the cylinder. Each simulation's results were compared to a reference case which was chosen to have a ϕ of approximately 0.65. Comparisons were made along the jet centerline at a height of 1.09 cm. This value of ϕ indicates there is excess air to ensure all of the fuel is adequately combusted. Figure 4.20(b) shows how well the measurement of H_2O concentration is able to provide an estimate of the correct ϕ value. While rejecting nearly all of the very high and low values, ABC gradually accepts more values close to the reference ϕ of $\sim$0.65 conforming closely to a Gaussian profile. The mean of the accepted ϕ values is 0.6507 with an error of $\sim$ 0.17%.

4.3.5 Results: Finding ϕ Using Temperature Measurement

Again using the same $\sim$10,600 simulations but now considering temperature measurements, we find an additional estimate of the appropriate ϕ value. The temporally averaged temperature within the domain is compared at a height of approximately 1.09 cm along the jet centerline. The rejection distance, ε, was again chosen by minimizing the confidence interval half width of

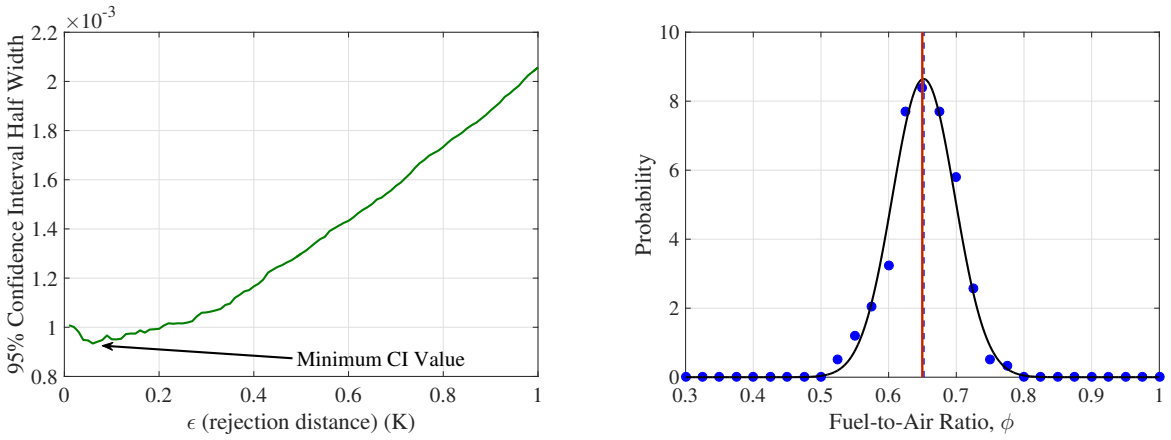


(a) Confidence interval

(b) Accepted simulations

Figure 4.20: (a) Confidence interval half width plotted against ε (rejection distance) values for H_2O *concentration* measurements. This value of ε strikes a balance between having high precision in the posterior (i.e. low standard deviation) and high confidence in the posterior (i.e. a large number of accepted samples). (b) Discrete PDF showing the number of accepted simulations versus values of inlet fuel-to-air ratio, ϕ, along with a Gaussian curve with the same mean and standard deviation as the accepted values. Here ∼10,600 simulations are performed over a uniform prior with ϕ 0.3–1.0. Statistic used is temporal mean H_2O *concentration* at one location in the domain. Reference ϕ shown in red at ∼0.65 with the mean of the accepted values shown with a dashed blue line.

the posterior generated by varying ε. The results of this ε sweep are shown in Figure 4.21(a). This method suggests using a rejection distance, ε, of 0.06 K indicating how close the measured temperature needs to be compared with the reference temperature. Note that this value of ε is also quite small relative to how well one might actually be able to experimentally measure temperature, in which case one might prefer to relax this constraint (i.e. increase ε). Figure 4.21(b) shows how well the temperature measurement is able to provide an estimate of the correct ϕ value. While rejecting nearly all of the very high and low values, ABC gradually accepts more values close to the reference ϕ of ∼0.65 conforming closely to a Gaussian profile. The mean of the accepted ϕ values is 0.6517 with an error of ∼ 0.32%. Comparing the results of using either temperature or H_2O measurements, one finds that both approaches perform extremely well and predict the inlet composition with less than 1% error. The temperature measurements perform slightly worse as expected since temperature in the domain is linked to concentrations at the inlet but less so than



(a) Confidence interval

(b) Accepted simulations

Figure 4.21: (a) Confidence interval half width plotted against ε (rejection distance) values for *temperature* measurements. This value of ε strikes a balance between having high precision in the posterior (i.e. low standard deviation) and high confidence in the posterior (i.e. a large number of accepted samples). (b) Discrete PDF showing the number of accepted simulations versus values of inlet fuel-to-air ratio, ϕ, along with a Gaussian curve with the same mean and standard deviation as the accepted values. Here ∼10,600 simulations are performed over a uniform prior with ϕ 0.3–1.0. Statistic used is temporal mean *temperature* at one location in the domain. Reference ϕ shown in red at ∼0.65 with the mean of the accepted values shown with a dashed blue line.

direct measurements of H_2O concentration.

4.4 Conclusions

As simulations and computational resources become more abundant, their potential is constrained by how well they can approximate real-world applications. In this light, data-driven engineering continues to be a rich field of research. The results shown in this chapter demonstrate that ABC is a powerful technique to estimate parameters of interest in complex RANS simulations, especially in many industrial applications. First, ABC successfully identified posterior distributions for two key parameters (inlet temperature mean and variability) driving a 3D turbulent buoyant jet problem. Through successive iterations the prior parameter space was effectively narrowed to provide additional insight regarding the parameter posterior. The narrower parameter space demonstrated how the stochastic nature of the unsteady 3D RANS simulations can cause the

statistics from a simulation to differ from expectation, thereby biasing the results. This chapter then presented a 2D case with a rotating cylinder above a turbulent, buoyant jet. In this case ABC successfully identified inlet velocity and the cylinder's rate of rotation. Next, ABC was able to identify proper inlet species composition. Both cases in this chapter demonstrate key statistics useful for identifying inlet parameters and provide valuable insights about where measurements should be taken and how many are necessary to provide information about those parameters. This study provides a basis of knowledge useful for moving into the final chapter of this dissertation which focuses on a novel application of ABC to combine a complex 3D LES case with reference observations coming from laser absorption experiments to predict several parameters of interest.

Chapter 5

Parameter Estimation in Turbulent Simulations using Approximate Bayesian Computation with Laser Spectroscopy

With ABC now established as an effective tool to perform parameter estimation for engineering systems, the next challenge is to apply it using reference data obtained through experimental measurements. [1] Specifically, the catalytic burner described in Chapter 4 will be measured using laser absorption spectroscopy methods. The measurements obtained include temporally resolved line-of-sight averaged values for temperature. Similar measurements are made in the numerical simulation to be compared against the experimental values. The ABC method compares the experimental measurements and each simulation to reject simulations that do not align well with the experiment and build a pdf of the likely parameters, while additionally providing insight about continued combustion within the domain.

5.1 Introduction for ABC using Laser Absorption Spectroscopy

In this chapter, we outline the ABC approach then use the method to predict unknown parameters for a turbulent, high-temperature buoyant jet above a catalytic burner. The unknown parameters of interest include an inlet boundary condition (velocity) as well as parameters used to model heat addition within the domain due to combustion. Experimental observations from a real-world system are obtained using laser absorption spectroscopy. High-fidelity three-dimensional (3D) large eddy simulations (LES) are used to model the corresponding flow field. ABC is used

[1] This chapter is based on work to be submitted to Experiments in Fluids [80].

to predict the inlet velocity boundary condition as well as heat addition parameters based on comparisons between the experimental observations and simulation results. The success of the ABC approach in the present tests suggests that ABC is a useful and versatile tool for engineering research focused on fluid and thermal systems.

In numerical simulations of practical fluids and combustion problems, there are often many parameters of interest – both known and unknown – that drive the solution. For instance, the inlet temperature and velocity of a buoyant jet determine the behavior of the resulting flow field, as well as the temperature distribution above the jet. The task of correctly assigning these parameters is crucial to building reliable computational models of real-world engineering systems. Additionally, experiments can be limited in the amount and type of data they collect so simulations can provide additional information to improve understanding of the system under investigation.

Often many of the parameters that are crucial to obtaining relevant, descriptive, and even predictive simulations, are unknown. This presents a problem when comparing simulation results to experimental results. Namely, are the differences observed in experiments and simulations due to modeling error (e.g. missing physics from the model) or due to not sufficiently characterizing all of the simulation boundary and initial conditions, transport properties, material properties, etc.?

The difficulty of determining unknown parameters can be approached by identifying distributions of probable parameter values, generating model data based on those parameters identified a priori, and then generating distributions for unknown parameter values using statistical inversions [4,6–8]. Various optimization methods have been used for parameter estimation (e.g., [16–20]), but these techniques focus on single values of unknown parameters, with limited measures of uncertainty when using potentially imperfect computational models and experimental data. Bayesian techniques naturally overcome this limitation by providing probability distributions of parameters. Several studies have investigated parameter estimation for engineering problems using fully Bayesian methods [9–12, 23], although such approaches require knowledge of a likelihood function. Calculating the likelihood function in these applications requires extensive observations or simplifying assumptions that have varying degrees of accuracy depending on the source of measurement

data and the complexity of the model. See [24] and sources therein for more details on parameter estimation using full-Bayesian approaches such as ensemble Kalman filtering (EnKF) and maximum likelihood (ML) estimation techniques.

In the present study, we use ABC to estimate parameters in a complex 3D computational simulation in order to obtain agreement with experimental measurements of a corresponding physical setup. The turbulent, high-temperature buoyant jet above a catalytic burner is the focus of this effort. This study focuses on a turbulent buoyant jet due to the geometrically simple yet physically complex nature of this compressible flow problem, which has many analogs in both engineering systems and environmental flows. While laser absorption spectroscopy used in this approach (see [81]) is able to obtain accurate temperature measurements above the burner surface, it provides little information about the inflow velocity. Additionally, temperature measurements suggest incomplete combustion within the catalyst but do not provide specific details as to all of the species exiting the burner and in what quantities they may be continuing to react above the burner surface. Therefore, ABC is used to obtain both estimates for the inflow velocity as well as heat source addition information for cases with continued combustion above the burner surface (indeed, for cases where no additional combustion is suspected a priori, ABC also predicts minimal heat addition above the burner). By performing ABC, we gain a much better understanding into the parameters (along with uncertainty) that create the best agreement between the simulation and the experimental observations; in addition to understanding the parameters, ABC also provides simulations with statistics that closely match the experimental observations.

By combining the strengths of experiments and simulations, the sparsity of experimental data and the unreliability of simulations can be overcome. A key challenge within engineering research is thus the use of sparse experimental information to improve simulation accuracy, and conversely, to use high-fidelity simulations to augment experimental observations.

5.1.1 ABC applied to LES Simulations and Laser Absorption Spectroscopy Observations

The ABC method serves as a bridge between experimental observations and computer simulations. It allows a set of unknown parameters in a simulation to be determined through comparison of simulation results with statistics from experimental or other reference observations. The ABC technique is described at length in Chapter 2. Considering more closely the particular application of a buoyant jet, ABC can be used, for example, to determine boundary conditions at the jet's surface. For example, laser spectroscopy methods [81,134,135] could be used to obtain line-of-sight absorption-weighted average temperature measurements at various locations within the flow field. These measurements will have temporal data that will be used to generate a time-averaged temperature vertical profile above the burner surface; this serves as a summary statistic $\mathcal{S}$ of the full temperature field, D. Next, a computational model will simulate the experimental setup to create full 3D temperature fields, $\hat{D}$, and will be sampled to obtain an equivalent vertical temperature profile, $\hat{\mathcal{S}}$. The values for parameters (θ) of interest (e.g., inlet velocity boundary condition and heat addition characteristics) will be chosen according to a prior distribution, $P(\theta)$. Each draw of parameters will produce a separate solution whose statistics will be analyzed and compared to the experimental data using a distance function, $\delta(\cdot,\cdot)$. As described previously, if the statistics agree according to a predefined threshold, i.e. $\delta(\cdot,\cdot) \leq \varepsilon$, the parameters will be kept, or otherwise rejected; the resulting distribution of parameters is the much desired posterior and is denoted $P(\theta|\delta(\mathcal{S}(\hat{D}),\mathcal{S}(D)) \leq \varepsilon)$.

The rejection ABC algorithm described above and implemented in this chapter, known as method D in Marjoram *et al.* [2], is summarized as follows:

(1) Generate parameter θ from the prior distribution $P(\theta)$.

(2) Simulate data $\hat{D}$ from model $\mathcal{M}$ with parameter θ, and compute the corresponding statistics $\hat{\mathcal{S}}$.

(3) Calculate the distance $\delta(\mathcal{S}(D),\mathcal{S}(\hat{D}))$ between $\mathcal{S}(D)$ and $\mathcal{S}(\hat{D})$.

(4) Accept θ if $\delta \leq \varepsilon$ (where ε is the rejection distance), and return to step 1.

Note that the generic parameter, θ, that one seeks to find could be the inlet mean velocity or a heat source strength in the domain when studying the turbulent buoyant jet.

In summary, ABC can be used to develop estimates for parameters given a set of observations. The method first samples a large set of parameters from the likely parameter space and runs a model forward using one of those parameters. Next, ABC compares the data from the simulation against the reference observations. If the simulation and observation are close, then ABC keeps the parameters; if the results are not close, then it rejects the parameters. This process continues until many parameters have been modeled and an adequate number have been accepted. These accepted parameters make up the desired posterior distribution.

5.2 Case Setup

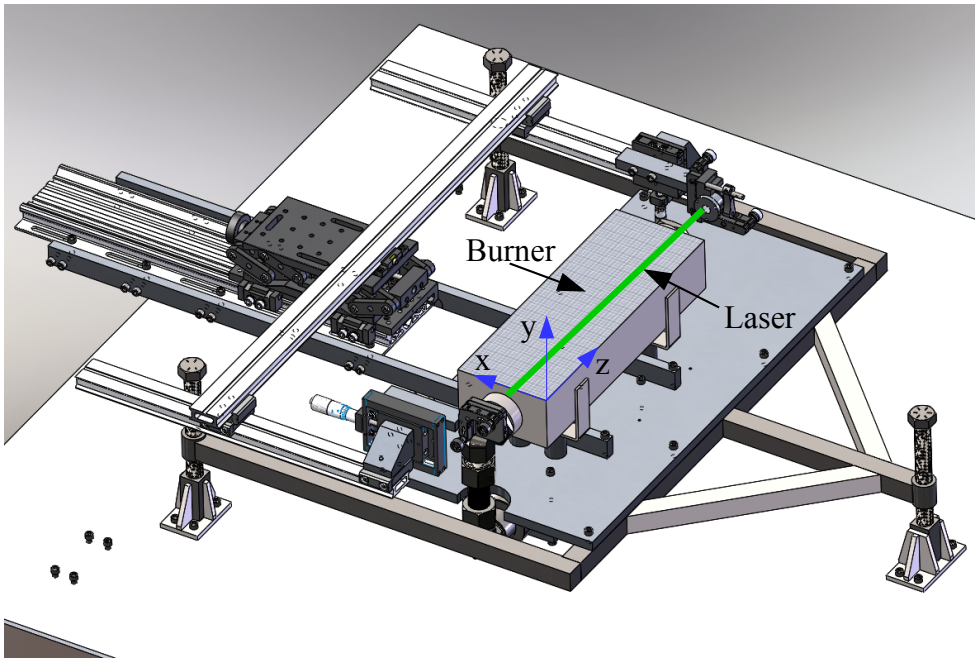
Motivated by an industrial burner configuration which produces a canonical type of flow, the system studied for this chapter is a heated turbulent buoyant jet created above a catalytic burner. The catalytic burner is 0.25 m wide by 0.075 m deep and fueled by methane and desiccant-dried air. An iron-chromium alloy catalyst bed is welded to an open-box metal frame. The catalyst is used to achieve combustion at lower temperatures thereby reducing pollutant emissions [140]. To premix the oxidizer and fuel, they pass through a bed of glass spheres within a cylindrical flame arrestor mounted beneath the combustor before entering the burner. Experimental measurements of the burner were made using laser spectroscopy, a technique which provides line-of-sight averaged, absorption-weighted temperature and species concentration data. As shown in Figure 5.1 (a) measurements are parallel to the long dimension of the combustor (termed the width in the remainder of the chapter and denoted by the z-axis). A CAD rendition of the burner is shown in Figure 5.1(a) while an image of the actual burner and laser optics is shown in Figure 5.1(b). A high-fidelity large eddy simulation (LES) case was developed to correctly model the relevant physics of the experimental setup while still being computationally affordable enough to execute

many (i.e. $\mathcal{O}(10^3)$) times.

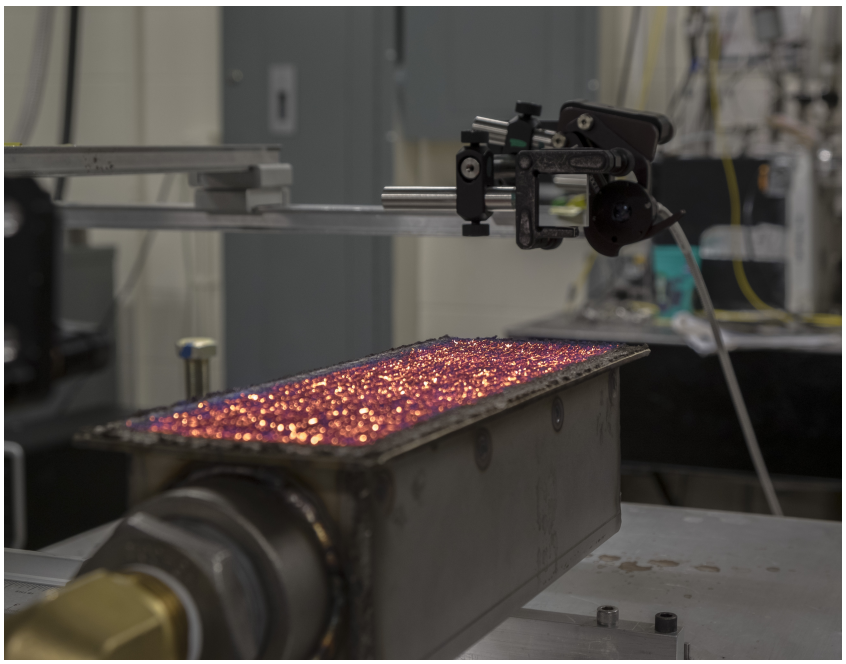
5.2.1 Experimental Approach: Laser Absorption Spectroscopy

An experimental apparatus was constructed to characterize the burner using tunable diode laser absorption spectroscopy (TDLAS). The catalytic combustor used in this work is the subject of previous efforts [82, 141]. It is a scaled down version of one that could be used for industrial processing; thus it is desirable to have a sensor that can meet the requirements of an industrial environment to aid in system monitoring and control. TDLAS provides a non-intrusive, absolute (quantitative), time-resolved (i.e. millisecond time resolution), robust, and portable sensor [142,143] to measure temperature at different locations in the buoyant jet. Other laser-based techniques such as Raman or Rayleigh scattering and laser induced fluorescence can also provide quantitative spatial and temporal information [144,145], however, such measurements often require large optical setups that are not ideal for industrial applications. The optical configuration for the combustor characterization is shown in Figure 5.1. A CAD rendition of the burner is shown in Figure 5.1(a) while an image of the actual burner is shown in Figure 5.1(b).

Wavelength modulation spectroscopy (WMS), a specialized form of TDLAS, is utilized in this effort. WMS also appears in studies of temperature and species mole fraction in a variety of combustion environments, from shock tubes to atmospheric flames [137, 146–153]. In WMS, the amount of light absorbed by a substance as a function of wavelength is quantified. Light is absorbed if the wavelength is resonant with a quantum transition of a target molecular species. The absorption by a particular quantum transition is directly and quantitatively related to the macroscale thermodynamic properties of the target gas such as temperature, pressure and species concentration. In TDLAS, the wavelength of a diode laser is tuned across one or two transitions of the target species. WMS is a form of TDLAS that utilizes additional high frequency signals and signal isolation to reduce noise for realistic systems. To conduct WMS, a fast modulation is applied to the injection current of a tunable diode laser, which results in rapid wavelength and intensity variations of the laser light. A lock-in amplifier isolates the series of harmonics



(a) CAD Rendition of Catalytic Burner



(b) Physical Configuration of Burner

Figure 5.1: Experimental setup for laser absorption spectroscopy above catalytic burner

in the signal measured at the detector that occur due to the fast intensity modulation and the effect of absorption. The second harmonic is especially useful because it is distinctly sensitive to absorption. Normalizing the second harmonic signal, denoted $2f$, by the first harmonic ($1f$) creates a ratio (i.e., $2f/1f$) that is sensitive to absorption but insensitive to intensity fluctuations (i.e. this ratio makes the sensor robust against harsh environments), and can deliver calibration-free absolute measurements [134]. While the specific shape of an absorption transition depends on the other non-absorbing gases present in the system (e.g., CO_2, CO, etc. in combustion systems), the integrated absorbance is independent of collision partner, and therefore does not require assumptions about, or measurements of, the other gases in the system. Beer's law, shown in Eqn. 5.1, simulates the transmitted intensity of the laser through the absorbing gases. Beer's law is given as

$$I_t(t) = I_0(t)exp(-\alpha(\lambda(t))), \tag{5.1}$$

where $I_t(t)$ is the intensity of the transmitted modulated light after it passes through the absorbing gases, $I_0(t)$ is the initial (un-attenuated) intensity, $\lambda(t)$ is the wavelength of the laser as a function of time, and $\alpha(\lambda)$ is the absorbance of the probed molecular species as a function of wavelength. The initial intensity, I_0, is measured on a photodetector for each laser, while the absorbance, $\alpha(\lambda)$, is simulated using a Voigt profile. Beer's law combines all three quantities [i.e., $I_0(t)$, $\lambda(t)$ and $\alpha(\lambda)$] to simulate I_t.

The integrated absorbance can be calculated through a fitting routine by scanning the laser wavelength across an absorption transition [137, 154, 155]. The fitting routine used in the current analysis is similar to the one described in Refs. [137, 154, 155], which simulates the laser tuning over the entire absorption feature as the fast modulation is applied. Using the measured WMS $2f/21$ as a baseline, the simulated $2f/1f$ signal is adjusted until the two match closely. To do so, the Voigt line-shape profile parameters (line-center, integrated area, Doppler width, and Lorentz width) are adjusted in a nonlinear fitting code until the profiles match. The spectral database HITRAN 2012 and the validated line parameters from [156] are used to create initial estimates for the Voigt parameters. The integrated areas that emerge from the fitted Voigt profiles for both of

these two different measured H_2O transitions permit the calculation of temperature.

Two-line thermometry calculates temperature using the integrated absorbance calculated from the WMS fit to two H_2O transitions. Integrated absorbance, A, can be calculated from

$$A = \int \alpha d\lambda = S(T)P\chi_{abs}L = \int_0^L S(T,x)P\chi_{abs}(x)dx, \tag{5.2}$$

where $S(T,x)$ is the line-strength of the probed transition at temperature T and location x, P is the total pressure (assumed constant for the entire pathlength), $\chi_{abs}(x)$ is the species mole fraction of the absorbing gas at location x, and L is the total pathlength of the laser beam across the gas.

Taking the ratio of the integrated areas (A_1 and A_2) of two absorption features yields a quantity (R) that only depends on the line-strengths, which are given as functions of temperature by

$$R = A_1/A_2 = f(T). \tag{5.3}$$

For more information on two-line thermometry see [157].

The ratio of the two absorption transitions is directly related to temperature while each is independently related to species mole fraction. Based on a ratio of the signal absorbed at the two different wavelengths (and hence for two different quantum transitions), a temperature can be calculated [82].

The current effort uses two NEL Inc. distributed feedback (DFB) diode lasers centered on H_2O absorption transitions at 1391.7 nm and 1468.9 nm. These absorption transitions have two primary advantages that make them ideal for temperature measurements in this system. First, they have a large difference in the lower state energy (1045 cm^{-1} and 3319 cm^{-1}, respectively); this optimizes the temperature sensitivity at elevated temperatures. Second, Goldenstein and Hanson [156] validated these line-strength parameters providing ideal conditions for fitting the Voigt profile. The lasers are combined and simultaneously passed across the long dimension of the burner (i.e. the lasers initially travel above the burner for 0.25 m parallel to the z-axis). To increase signal-to-noise ratio, the laser is reflected twice, causing the laser to pass over the burner three times creating an effective path length of 0.75 m. The sensor has been previously validated

within a controlled system with known conditions [82]. The temperature measurements were found to be accurate within 3%. Above this temperature, uncertainty values are extrapolated using the nonlinear relationship between absorption and temperature resulting from eqn. 5.3.

5.2.1.1 Experimental Campaign

With the experimental technique established and setup built, an experimental campaign was undertaken to characterize the heated buoyant jet above the burner using a vertical temperature profile while the burner operates at several different conditions. A vertical profile yields valuable information about the behavior of the flow field. It also can provide information about whether continued combustion is likely occurring above the burner surface. Thus, a y-shaped optical stage is mounted on a scissor jack that enables vertical translation of the laser beam. Laser measurements were taken at heights of 0.5 mm to 10.5 mm in 2 mm increments, then 10.5 mm to 150.5 mm in 10 mm increments above the burner surface all centered depth-wise across the burner (i.e. 37.5 mm from the edge of the combustor); this results in temperature measurements at 20 vertical positions. The tighter measurement spacing close to the burner provides additional information in the region most likely to contain sustained combustion. The burner is fueled with methane and run at two power fluxes, each with a unique equivalence ratio. The equivalence ratio, ϕ, is a parameter used to describe the fuel-to-air ratio in combustion by dividing the actual fuel-to-air ratio by the stoichiometric fuel-to-air ratio ($\phi < 1$ is fuel-lean, $\phi = 1$ is stoichiometric, and $\phi > 1$ is fuel-rich). Alicat mass flow controllers regulate the flow rates of air and methane setting the incoming power flux and equivalence ratio. The first power flux is set to 16 W/cm^2 with an equivalence ratio of $\phi = 0.80$; these conditions define 'Case 1'. Operating the burner at this pair of conditions is thought to result in complete combustion within the catalyst such that all of the reactions are essentially finished by the time the hot products exit the burner surface. This assumption was verified by probing various equivalence ratios and powers and determining a condition where OH, a radical that commonly appears during methane-air combustion, was not detected. Additionally, temperature consistently decreases for this case indicating no significant

Table 5.1: Experimental conditions

Case	Power Flux (W/cm^2)	ϕ
1	16	0.80
2	27	0.85
3	27	1.00

heat addition above the burner (i.e. no significant combustion above the burner). An additional case (referred to herein as 'Case 2') was investigated with inlet power significantly higher than Case 1 but with a similar fuel-to-air ratio; its conditions are as follows: power flux of 27 W/cm^2 and equivalence ratio set to $\phi = 0.85$. Finally, an inlet condition was probed experimentally with conditions that suggest continued reactions occurred above the burner surface; specifically, the temperature increased vertically above the burner surface and a high temperature was maintained farther into the domain. This case, which shall be referred to as 'Case 3', had inlet conditions as follows: power flux of 27 W/cm^2 and equivalence ratio set to $\phi = 1.00$. Note that this power flux is the same as Case 2, but the fuel-to-air ratio is significantly higher than Cases 1 and 2. The additional fuel entering the combustor in Case 3 causes additional chemiluminescence above the catalyst surface which would likely correspond with continued combustion. Experimental cases are summarized in Table 5.1. A fume hood is located several feet above the burner to evacuate exhaust gases while having negligible impact on the fluid and thermodynamics observed close to the burner surface. A CAD rendition of the burner is shown in Fig. 5.1(a) while an image of the actual burner is shown in Fig. 5.1(b). Calibration data, along with additional setup information and data processing techniques, can be found in [82]. Ten minutes of data were obtained at each height (starting sequentially from the bottom measurement) and for each set of burner conditions. Based on an Allan deviation analysis, outputs were averaged with 0.25 s intervals. Time-averaged vertical temperature profiles from the three experimental conditions are shown in Fig. 5.2.

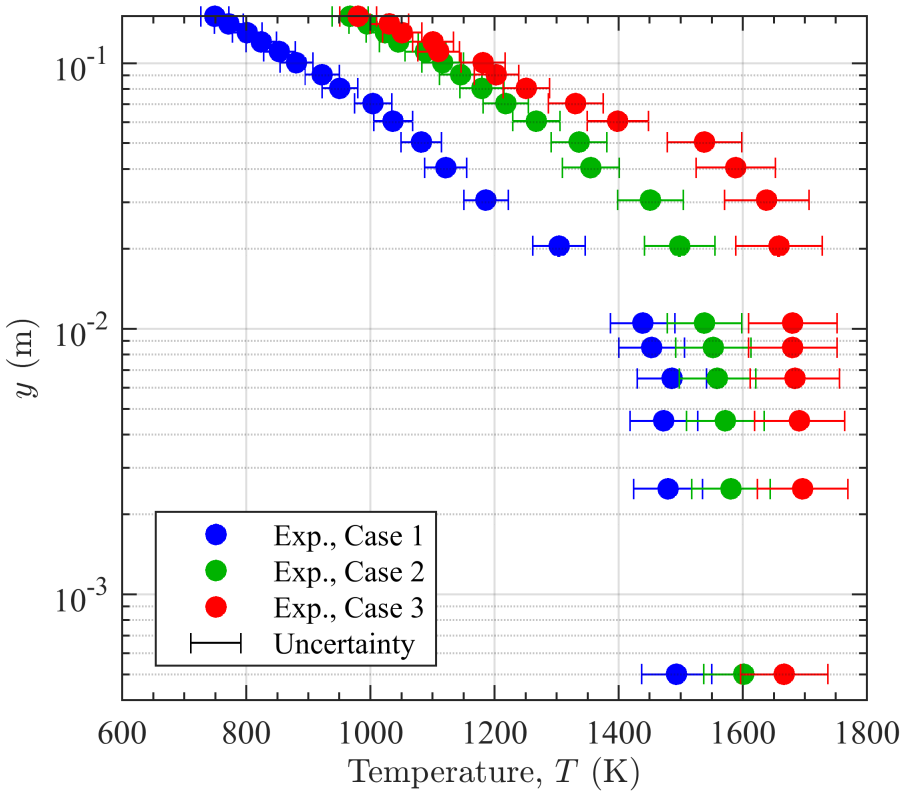


Figure 5.2: Vertical temperature profiles corresponding to three experimental cases. Experimental uncertainty shown based on instrument validation

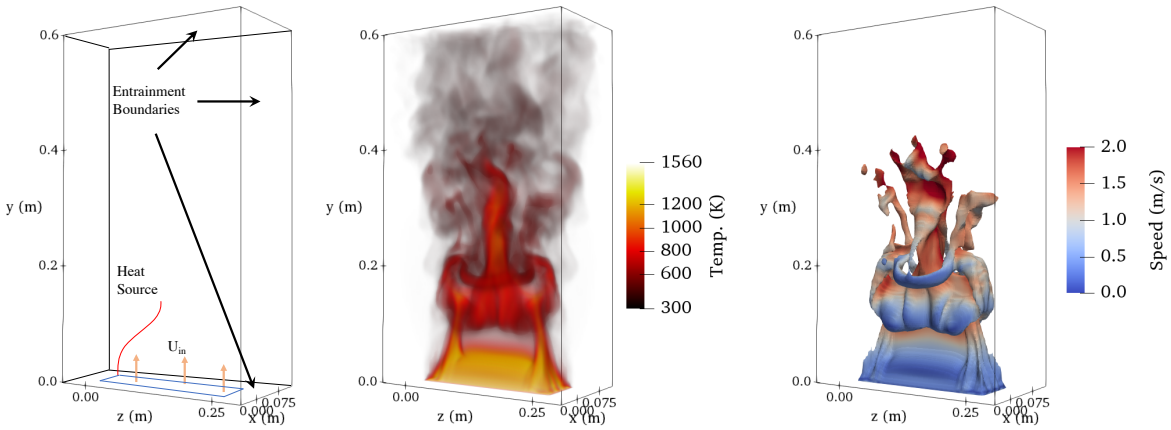
5.2.2 Computational Approach: Large Eddy Simulation (LES)

OpenFOAM, version 4.x [119, 120], was used to design and run a three-dimensional (3D) LES computation for this study using the FireFOAM solver [118]. A non-reacting heated, buoyant jet with a rectangular exit matching the size of the catalytic burner was simulated. The low-pass filtered compressible Navier-Stokes equations were solved with second-order accuracy in space and time in conjunction with mass, enthalpy, and species conservation equations using FireFOAM. The resulting large eddy simulation (LES) equations were closed using a dynamic one-equation eddy sub-grid scale model. Pressure-velocity coupling was accomplished using the PIMPLE algorithm, which combines the pressure-implicit split-operator and the semi-implicit method for pressure-linked equations. Heat transfer mechanisms modeled in this simulation include conduc-

tion, advection of temperature by species within the domain, and radiative losses approximated with the discrete ordinate method. The LES equations were solved on a 3D grid with approximately 2.4 million hexahedral control volumes stretched vertically to enhance resolution near the burner, resulting in a maximum vertical resolution of approximately 0.8 mm. The grid was uniform in horizontal directions. The domain is shown in Fig. 5.3 (a). Domain dimensions (0.35 m wide by 0.175 m deep by 1.2 m tall) were chosen to allow the bottom-driven jet to exit the domain primarily vertically due to the vertical inflow velocity and buoyant forces. The jet inflow was centered at the bottom of the domain and is 0.25 m wide and is 0.075 m across. Time stepping was adaptive and controlled by a maximum CFL condition set to 0.4. After a sufficiently long period to allow the flow field to fully develop, mean statistics for temperature were computed over a duration of 1 s. Samples obtained using this time period were found to be within approximately 1% of samples obtained from a 20-second window while allowing many more simulations to be run with different inlet parameters. Example temperature and speed fields from a simulation are shown in Fig. 5.3 (b) and (c), respectively.

Inlet conditions were fixed at the temperature measured at the combustor surface. Complete combustion was assumed and mass fractions of N_2, O_2, H_2O and CO_2 were fixed accordingly as specified by the NASA software Chemical Equilibrium Analysis (CEA) [139]; an equivalence ratio was specified as $\phi = 0.80$ for Case 1, $\phi = 0.85$ for Case 2, and $\phi = 1.0$ for Case 3 in order to match experimental conditions. Although cases with continued combustion would not have equilibrium conditions at the burner surface, the true composition of the species is unknown and hence this serves an anchor point; since combustion is not modeled in the CFD simulation the exact composition of the species at the inlet has limited impact on the conclusions of the analysis. Additionally, post processing CFD simulations, as described in Section 5.2.5, does not depend on the absolute quantity of H_2O present in to domain, but rather the distribution of the H_2O, which is assumed to enter the domain uniformly in space and time.

To account for the possibility of continued combustion within the domain heat addition was modeled above the combustor. This was accomplished by adding a source term to the enthalpy



(a) Simulation domain (b) 3D temperature rendering (c) 3D speed rendering

Figure 5.3: (a) Schematic showing the setup of the 3D LES domain. (b) Temperature field (volume rendering within an isosurface at 500 K and additional volume rendering at higher locations) a single snapshot in time of a simulation with inlet mean temperature of 1515 K and inlet speed of 0.38 m/s with no additional heat added above the burner surface. (c) Speed field corresponding to temperature shown in part (b)

equation in OpenFOAM. The heat was added using a half-Gaussian distribution with the peak occurring at the burner surface then trailing off in intensity as height increases above the burner (in the y-direction) as depicted in Figure 5.3 (a). Heat was applied only in the 0.075 m by 0.25 m region directly above the burner; at x and z-locations beyond the edge of the burner the heat addition is set to zero. The half-Gaussian distribution was parameterized using the total amount of power added (also referred to as the "heat source strength") and the height above the burner corresponding to three standard deviations (also referred to as the "heat source height"). At y-locations above the heat source height the heat addition is set to zero. This approach to heat addition most closely resembles continued premixed combustion above the burner surface since the most intense combustion occurs at the surface (where there would be sufficient mixed fuel and oxidizer to react) then decreases as the flow moves away from the burner surface as combustion reactants are consumed.

5.2.3 ABC Statistic and Metric Selection

The ABC method applied is given in Section 5.1.1. In this ABC algorithm each set of parameters is chosen from a uniform grid of values where one parameter is held fixed while the other parameters are varied over a predefined set of values. Then, the first parameter is changed slightly and again paired with the same group of values for the other parameters. Choosing a statistic to represent each simulation is a key step in ABC. The statistic must significantly reduce the data so a reasonable comparison can be made (usually it is not computationally practical to compare the entire flow field at every location in space and time, so instead a summary statistic is utilized). The statistic must also contain an identifiable signature left by the parameter of interest; if the parameter is unidentifiable or noninfluential, then the technique will not be able to discern the impact of changing the parameter from that measurement [107]. To that end, a variety of statistics were considered for this study.

First, the study sought to determine the mean inflow velocity parameter. One could reasonably assume based on the physics governing the simulation that the mean inflow velocity is likely coupled with the mean temperature in the space above the hot jet. This was demonstrated further by Christopher *et al.* [79] wherein they showed that temperature measurements at single locations within approximately 0.5 to 1.5 burner diameters above the burner could be used to infer inflow velocity parameters. Here we have built upon that knowledge by using mean temperature values at multiple locations simultaneously. This statistic is referred to as the vertical temperature profile. Additional statistics of temperature were considered, including the full PDF, but due to the short time duration of the simulations the higher order statistics are likely less converged. Additionally, the mean temperature profile is considered by many as the benchmark for matching experimental observations. Temperature values were sampled from within the 3D domain according to the experimental setup described in section 5.2.1.1.

Once the appropriate statistic is obtained (i.e. a vertical profile of temperatures), the next step is to determine the appropriate metric by which to compare the observation data and the

simulation data. Here L1 and L2 norms, along with RMSE, were investigated. Another approach is to integrate the difference in temperature values along the height. This would mean measurements closer in space (i.e. those nearest to the burner) would be weighted less than those spread out farther (i.e. those highest up in the temperature profile). Posterior distributions were fairly insensitive to the statistic chosen. Thus, we concluded using the RMSE provides a robust and common metric; additionally, Chai and Draxler [158] show that RMSE is appropriate in many cases for comparing model to observational data, especially when model error is assumed to be Gaussian, as is likely the case here. We also choose to weight each measurement equally because, although the bottom measurements are very close together, this spacing is intentional since matching the lower portion of the profile is key to helping determine whether continued combustion is observed above the burner surface.

Lastly, once the statistic and metric are chosen, one must decide how closely each simulation must match each simulation in order to be deemed acceptable. The rejection distance, ε, used for ABC is a threshold value; any simulations with a distance compared to the observations larger than ε are rejected. Thus, it is important to carefully chose ε in order to obtain reasonable results. A value of ε that is too large will return all of the simulations and produce a biased posterior distribution identical to the prior distribution; thus, no additional information is gained. Conversely, an ε that is too small rejects all but a single case resulting in a posterior with very low confidence (i.e. high Monte Carlo error). Marin *et al.* [59] and Lintusaari *et al.* [60] argue that a smaller ε is generally preferred, but the size is largely determined by the available computational resources. Based on high computational costs associated with performing 3D LES simulations with sub 1 mm resolution, a limited number ($\mathcal{O}(10^3)$) of simulations can be performed for this study. Thus, using the approach of previous efforts (e.g. [2, 58–60, 104]), posteriors corresponding to various ε values are displayed showing convergence as epsilon is reduced.

5.2.4 ABC Prior Distribution Selection

Once experimental observations for the three separate cases were available, we implemented the ABC procedure to identify the parameters, along with inherent uncertainty, that provide the best agreement between simulations and the experimental results. The first step in this procedure is to identify an appropriate prior distribution of parameters. The prior distribution needs to be wide enough to encompass realistic values for the parameters, but narrow enough to keep the procedure computationally affordable. The first decision was to use a fixed temperature for all simulations for a given case. This decision was made to provide computational savings as each parameter one seeks adds approximately one order of magnitude more simulations (i.e. the number of simulations, N, needed to identify n parameters goes as $N \sim 10^n$). More importantly, the temperature measurements at the lowest height are believed to represent the exit temperature of the burner well since they should be relatively uniform and the measurement is extremely close to the burner surface; in fact, the laser skims the burner surface. Thus, since there are approximately three bottom temperatures observed (Case 1 is at approximately 1500 K, Case 2 is at approximately 1600 K and Case 3 is at approximately 1685 K at the inlet) there are three temperatures required for all simulations. Next, the inlet velocity is a primary quantity of interest in these simulations as it determines how much momentum enters the domain and hence how high the hot temperature gases are vertically advected once in the domain. Knowledge of the likely velocities to simulate comes from a mass balance based on measured fuel and air rates entering the combustor, along with a known exit temperature, pressure and area. The next parameters relate to heat addition above the burner and are used to determine whether additional combustion above the burner surface is likely present in each case. The two parameters needed to characterize the heat addition are the heat source strength (i.e. the quantity of power added in Watts) and the height of the addition, i.e. how spread out the half-Gaussian distribution is. The lower limit for heat addition is zero heat addition indicating no additional combustion; the upper limit on heat addition is determined from an energy balance recognizing how much power is going into the system from the fuel and

how much energy it would take to heat the gases to the measured inlet temperature. Next, limits were decided upon for height of heat addition. This parameter indicates where the half-Gaussian distribution is three standard deviations tall and hence the majority of the energy has been inputted. A lower limit was chosen to confine the heat addition very close to the burner. Based on previous experimental measurements showing increases in temperature within the vertical profile above the catalytic burner, an upper limit was chosen. With these three parameters identified (velocity, heat source strength, and height of half-Gaussian heat addition), a uniform grid was established and equally divided providing a structured grid of parameters to simulate. The uniform grid was chosen to be the least biased within its domain of support. Admittedly, outside of this domain parameters are not considered indicating confidence that the true parameters lie within the parameter bounds. The grid of parameters was chosen over other viable options (i.e. random draws, Latin hypercube sampling, etc.) due to the limited number of simulations to be computed; this method effectively covers the entire parameter space.

5.2.5 Comparison of Computational and Experimental Results

In order to accurately compare the computational simulation results to the experimental observations, great care was taken to ensure the measured temperatures were equivalent. Though a spatial average of the cells through which the laser passes would be the simplest option, this ignores the nonlinear effect of temperature on absorption and the proportional effect of the water mole fraction. Thus, a more sophisticated approach is used to calculate a line-of-sight (LOS) averaged absorption-weighted temperature from the computationally obtained 3D temperature fields. Specifically, Eqn. 5.2 is used to calculate the integrated absorbance that would occur for each of the two lasers passing through the CFD domain. Note that pressure is assumed to be constant for the LOS and hence cancels out when taking the ratio of Eqn. 5.3. To accomplish this conversion, the line strength (i.e. how much each laser is absorbed at a given temperature) corresponding to each temperature is calculated based on the technique of [45]. The partition function used in this equation is determined from [159]. The line strength is integrated, for each laser independently,

across the entire LOS while weighting each value by the normalized water mole fraction at the location corresponding to each temperature. The ratio of integrated line-strength values obtained for each laser is used to calculate an LOS averaged absorption-weighted temperature as given in Eqn. 5.3 at a given height. These temperatures can then be averaged in time to get a mean temporal temperature value (for a given height). As discussed in Section 5.2.3, these temporally averaged temperatures at all specified heights for a given simulation form the statistic (i.e. the vertical temperature profile) to be compared for each simulation in the ABC method. Temperatures were recorded within the computational domain at the same centered location (i.e. centered in x-dimension 3.75 cm from burner edge) across the burner width (parallel to z-axis; see Figure 5.1 (a)) for each location above the burner corresponding to the heights used for the experimental measurements (see Section 5.2.1.1 for more details). For each snapshot in time, all temperatures and H_2O mole fractions at a given height and for a given LOS are used to calculate an LOS absorption-weighted temperature.

In addition to the LOS averaging and absorption weighting, the simulation data must also be adjusted to account for uncertainty due to the measurement procedure. To account for this, white noise was added to the simulation data. A random number between -3% and $+3\%$ was drawn, then this amount of uncertainty was added to the entire profile of temperature means; this assumes any bias observed for an experiment is correlated at all heights due to being part of one continuous data set. This procedure was repeated one thousand times for each simulation. Note that the variation from one profile to another is generally larger due to changing parameters versus adding this uncertainty.

5.3 Results

Here we identify the most likely parameters needed to create agreement between the simulation results and the experimental observations for each of the three cases. Marginal and joint posterior distributions for simulation parameters show regions with likely parameter choices. These results indicate probable inlet velocities for each case as well as identify when combustion is or is

not likely to be occurring above the burner surface. Temperature profiles and fields for the simulations with parameters drawn from the modal values of each posterior distribution are also shown in order to illustrate the unique behavior exhibited by each of the three cases.

5.3.1 Posterior Distributions

ABC provides a distribution of the parameters most likely to generate statistics close to the experimental observations. By collecting all of the parameters corresponding to simulations with the smallest distance when compared to the experimental observation, one can construct posterior distributions. In the current study posterior distributions indicate which velocity and heat addition parameters are most likely to produce temperature profiles matching the experimental profiles. Another outcome from ABC is realized if the model is assumed to have small error compared to the experiment in terms of how it represents relevant physics, geometry, etc. If this is true, then the parameters obtained provide information about the experimental setup itself that may be unfeasible to obtain otherwise; potential reasons why experimental data might not obtainable include limited system access, existing historical measurements that can not be readily reproduced, limitations of current experimental techniques, etc. [4]. In this particular application, the inlet velocity and heat source addition are not feasible to measure and must be inferred. The velocity of the system is extremely difficult to probe experimentally because of the high temperatures and complex combustion within, and potentially above, the catalyst. These conditions would melt many velocity measurement instruments and make seeding the flow difficult, as would be needed for particle image velocimetry (PIV). Any potential combustion occurring above the catalyst is also extremely difficult to predict or quantify due to the complex nature of the reaction mechanisms within the catalyst, combined with how it behaves across a variety of power fluxes and ϕ values. For an additional example of information provided by ABC to predict a missing temperature value in the vertical profile, see Section 5.3.3.

Next we review the posterior distributions for each case individually. For Case 1, which has a power influx of 16 W/cm^2 and $\phi = 0.80$, the inlet speed estimation does well as shown in the top

row of Fig. 5.4. The posterior distribution mode of 0.3 m/s, as given in Table 5.2, is slightly lower than the predicted inlet speed region shown in red in Fig. 5.4. The range of inlet velocities for the inflow speed accounts for uncertainty in quantities used to predict the inflow speed including volumetric flow rates provided by flow meters, atmospheric pressure, and temperature. The lower secondary peak at 0.13 m/s corresponds to cases with more heat addition; this shows that additional heat addition can compensate for lower velocities; however, too much heat addition concentrated within the domain combined with low inlet velocities can lead to increases in temperature in the simulations that exceed the increase in temperature observed experimentally. Furthermore, this case was probed experimentally and shows no signs of additional combustion (undetectable amounts of OH and no appreciable increase in temperature above the surface as seen in Fig. 5.2); thus the peak seen at 0.3 m/s corresponding to 0 W heat addition appears to be the most probable condition corresponding to the experimental measurements. Turning next to the height of heat addition, the posteriors appear to have little information gain compared to the prior. This is seen in the top row of Fig. 5.5; for the lowest amount of heat addition (0 W) the heat height is spread out equally among all possible options. This is because once the strength is set to zero, the height is irrelevant because any height chosen results in no heat addition, and thus all heights are equivalent.

For Case 2, which has a power influx of 27 W/cm^2 and $\phi = 0.85$, the inlet speed estimation does extremely well as shown in the middle row of Fig. 5.4. The mode of 0.63 m/s, as given in Table 5.2, is extremely close to the mean speed of 0.65 m/s predicted experimentally. This case was probed experimentally and shows no signs of additional combustion (low quantities of OH and no appreciable increase in temperature above the surface as seen in Fig. 5.2); this matches the posterior which appears to converge to 0 W of heat addition above the burner. Similar to the description given for Case 1, with a heat addition of 0 W the resulting height is arbitrary since all will produce identical results (no heat addition); this trend is observed in Fig. 5.5. The primary difference observed in this case, then, compared to Case 1, is that additional power flux into the system caused an increase in mass flow rate which resulted in an increase in velocity at the boundary. This is captured very well by the inlet speed posteriors of Fig. 5.4. Additionally, no

combustion appears to occur above the burner surface and this is also corroborated by the posterior distributions.

For Case 3, which has a power influx of 27 W/cm^2 and $\phi = 1.00$, conditions are substantially different. The power flux in Case 3 is significantly higher than that of Case 1 indicating a significant increase in the amount of reactions taking place in the burner. Compared to Case 2, the amount of air accompanying the fuel is decreased substantially providing fewer opportunities for completely mixed combustion to occur. Thus, in Case 3 it is more likely that a portion of the combustion might not complete within the combustor causing some of the reactants to exit the burner surface. Looking at the posteriors presented in the bottom row of Fig. 5.4, this prediction appears to be correct as heat addition above zero Watts is most likely (the posterior peaks at approximately 750 W). The posterior estimate for heat addition is very interesting in that it shows additional heat is usually needed in order to match the experimental profile. This makes sense as the profile maintains a high temperature, even increasing in temperature as one moves away from the burner. This behavior is very indicative of continued combustion above the burner and is supported by the ABC parameter predictions. The height of the heat addition does not seem to have strong identifiability based on the relatively flat posteriors (compared to the other parameters), but does exhibit a slight preference toward more stretched out heat addition (peaking around 0.0625 m). For heat addition sizes larger than this, the posterior starts to drop off. To fully characterize this behavior, additional simulations were completed with heat addition occurring up to 0.1 m (which is 0.03 m above the original cutoff). These more stretched out heat addition heights resulted in lower densities within the posterior indicating that the identified peak of 0.0625 m is well within the bounds of the prior distribution. Looking at the joint posteriors, an informative trends stand out. As heat source strength increases, the heat source height also increases indicating that as more power is added it must be spread out more. The inlet speed estimate is very close to the speed predicted from the experimental mass balance approach; the speed is slightly lower than Case 2 due to the reduced air entering the system. Looking at the joint posteriors in the bottom row of Fig. 5.5 it appears that lower velocities require very high amounts of heat addition in order to

Table 5.2: Experimental conditions and parameters from posterior mode

Case	Pwr Flx (W/cm^2)	ϕ	Pred. In Spd (m/s)	Inlet Spd (m/s)	Ht Str (W)	Heat Ht. (m)
1	16	0.80	0.37	0.30	0	N/A
2	27	0.85	0.65	0.63	0	N/A
3	38	1.00	0.62	0.47	750	0.06

produce profiles that match the experimental vertical temperature profile. While this is possible, it's more likely that higher velocities would be observed within the combustor corresponding to higher velocities in the simulation of approximately 0.6 m/s. These velocities are within the support of the posteriors for Case 3 as seen by the overlap in the red velocity prediction region and the posteriors.

Looking at all three cases together, it is clear that ABC consistently predicts both reasonable inflow velocities as well as heat addition. These parameter predictions result in temperature profiles from simulations that well match experimental observations as will be seen next.

5.3.2 Temperature Profile and Field Comparisons

The temperature profiles for simulations corresponding to the mode of the posterior distributions are shown in Fig. 5.6. These profiles correspond to the parameters listed in Table 5.2. Note that while these parameters are the best estimates in order of providing profiles that consistently match observations, different parameters may produce better singular profiles but these parameters would have little statistical relevance.

The temperature fields corresponding to simulations with parameters at the mode of the posterior distributions compared to the experimental cases (as listed in Table 5.2) are found in Fig. 5.7. These fields illustrate the unique behavior seen in each of the three cases. The low velocity corresponding to Case 1 appears to substantially change the flow dynamics seen in Case 1 compared to the other two cases.

5.3.3 Observing System Experiment

As a final test of this procedure to understand its predictive capabilities, an Observing System Experiment (OSE) (see, for example, [160,161]) was conducted. In this test, ABC was performed in the same fashion as above. However, one of the experimental measurements was excluded from the reference data. After comparing all of the simulations to the reference data (with temperatures at a particular height excluded from the comparison), a set of simulations is identified through ABC as having the smallest distance from the experimental observation. These simulations provide an estimate of the temperature at the chosen height above the burner surface; this temperature estimate can be compared to the experimental data (which in this case is available but was not used to generate the list of best simulations) and provides insight into how well the chosen simulations can predict missing experimental data. The data for Case 1 is used as an example. Since there appears to be a marked change in the temperature gradient at approximately 1.05 cm above the burner surface, this temperature is removed from the data set and the ABC procedure is executed as in Section 5.1.1 with ε set to accept 5% of the simulations. The ABC analysis proves extremely informative and useful for this temperature inference. The original set of all simulations (i.e. the prior) has a mean temperature of 1660 K ($\sigma = 196K$) at 1.05 cm; then, after ABC reduces the candidate simulations to the posterior (which includes only 5% of the original set of simulations) the resulting mean temperature is 1456 K ($\sigma = 52K$), which is within approximately 1% of the experimentally measured temperature of 1439 K. Note that including the experimental measurement at 1.05 cm above the surface improves the ABC prediction of this temperature to 1453 K. This quick demonstration indicates ABC is extremely effective at providing predictive information within the domain where the simulation and reference data match, even when there is no corresponding reference data to inform the inference. In future efforts this approach could be used to design experiments with the fewest necessary temperature measurements to still provide similar parameter estimations.

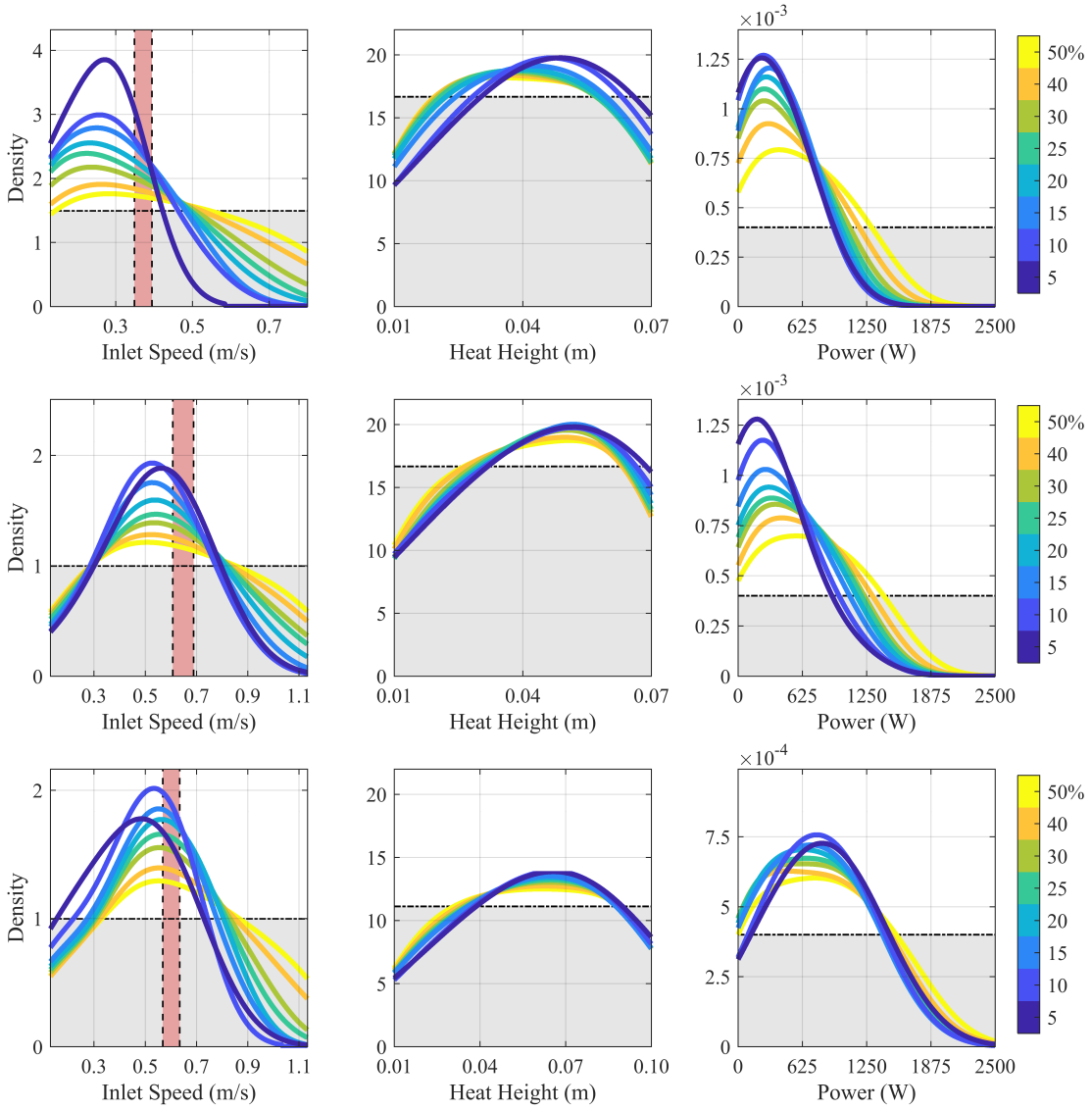


Figure 5.4: Posterior distributions for cases 1, 2 and 3 (top, middle and bottom row, respectively) for inlet speed, height height, and power (left, middle and right columns, respectively). Line colors correspond to percentage of accepted simulations as determined by ε parameter. Inlet speed posteriors show range of predicted velocities in the experimental results

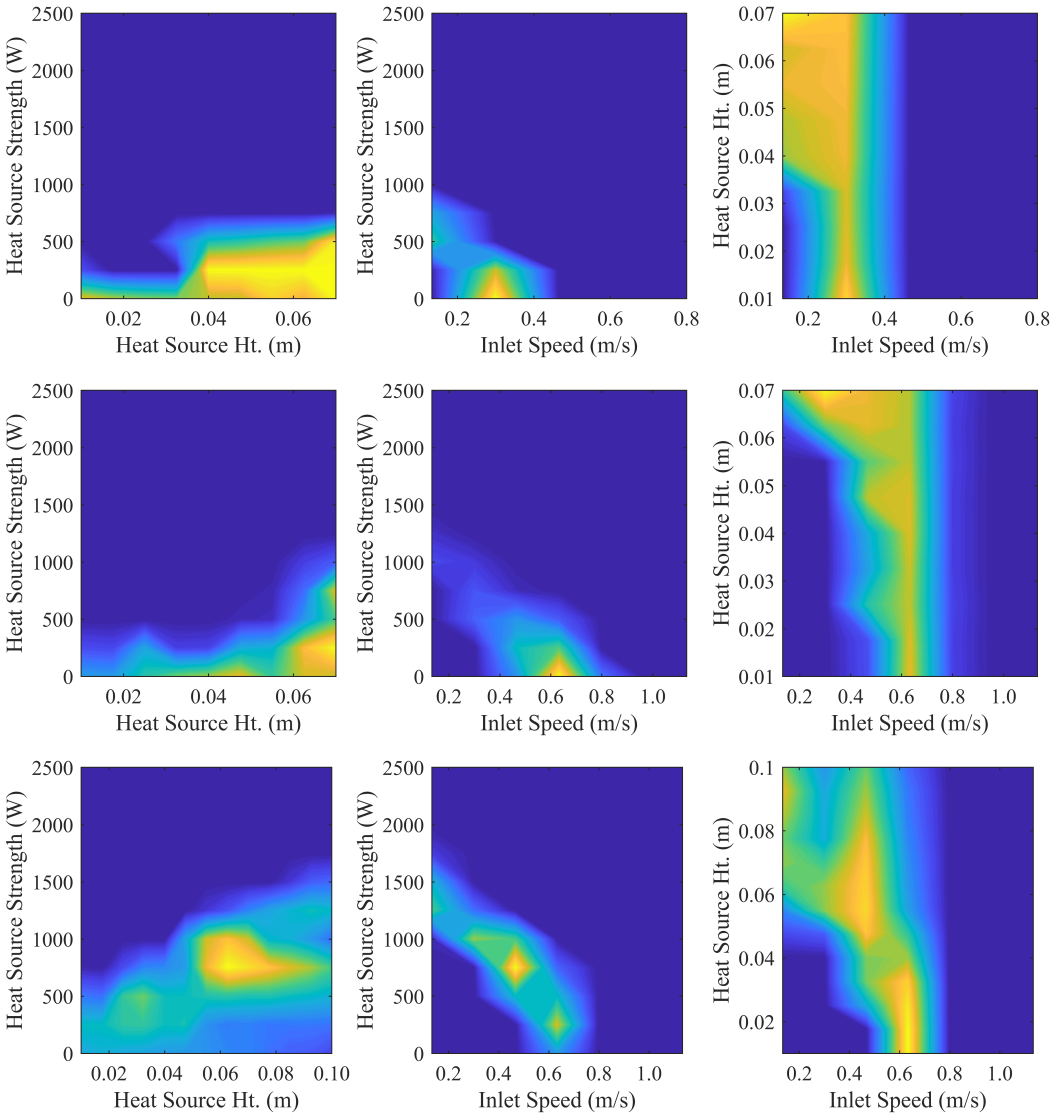


Figure 5.5: Marginal posterior distributions for cases 1, 2 and 3 (top, middle and bottom row, respectively) for two parameters taken when ε is set to accept 20% of simulations. Colors represent density of posterior with yellow being highest density and blue being lowest density

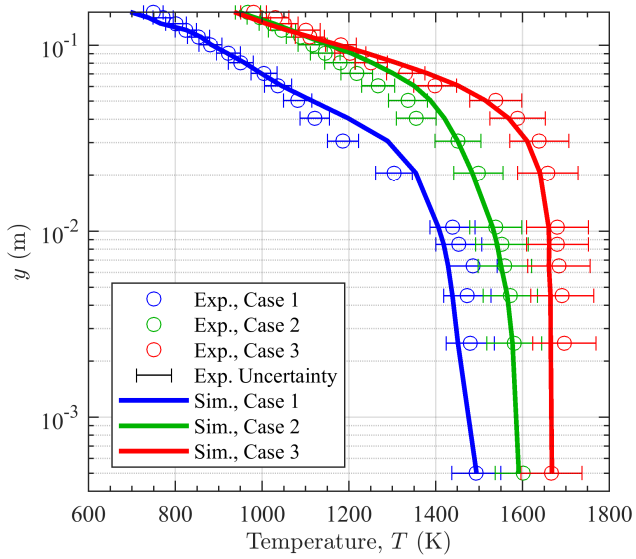
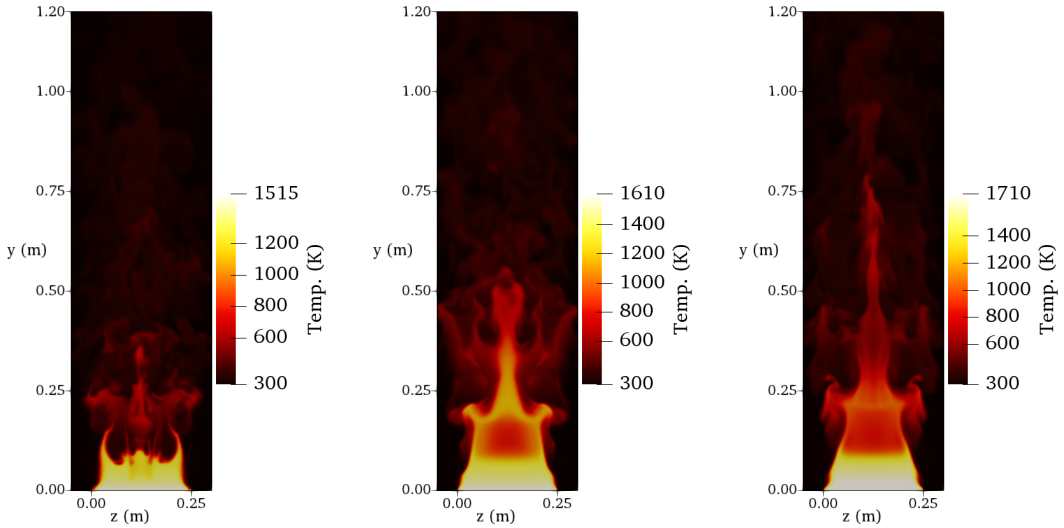


Figure 5.6: Temperature profiles for experimental data are shown with empty circles. Error bars denote experimental uncertainty. Solid lines denote simulation data. Data colored according to experimental case as determined by power influx



(a) Case 1 (b) Case 2 (c) Case 3

Figure 5.7: 2D slice of temperature fields (taken in the middle of the burner in the x-direction) corresponding to three simulations, each with with parameters taken from the mode of the marginal posterior distributions in Figure 5.4 and listed in Table 5.2 for each of the three cases.

5.4 Conclusions

The present study demonstrates that ABC is a powerful technique to estimate parameters of interest in engineering applications, in particular those of complex thermal-fluid flows. ABC successfully identifies a posterior distribution for three key parameters (inlet velocity, as well as heat source strength and height) driving a 3D turbulent buoyant jet problem while the observations come from a laser absorption spectroscopy experiment. This chapter demonstrates key statistics useful for identifying inlet and model parameters and provides valuable insights about where measurements should be taken to provide information about those parameters. Based on the parameter posteriors, we gain additional insight into this case by identifying if combustion is likely occurring and in what quantity. Additionally, using an OSE, we are able to successfully predict a missing temperature from the experimental data.

This study leaves several questions to be answered in future efforts. First, how well will this procedure work on more complicated 3D geometries? Additionally, how will this procedure extend to physical setups containing known quantities of combustion (e.g. ribbon burner applications as shown in [86]) where simulations include reaction mechanisms? Additionally, what turbulence model parameters could be adjusted to better close the LES equations for this particular application?

Chapter 6

Conclusions and Future Research

The tools built in this dissertation demonstrate the efficacy of using ABC to estimate parameters for computational engineering models. Specifically, the ABC method has been successfully applied to a 0D Navier-Stokes (logistics) equation, a 2D high-temperature turbulent buoyant jet using LES and DNS, a 2D RANS high-temperature buoyant jet with a rotating cylinder above it, a 2D LES forced-helium plume, and a 3D RANS high-temperature buoyant jet; all of these cases used data from simulations to characterize and develop ABC as a tool in this class of thermal-fluid engineering problems with complex flows. As a final application, ABC was used with a 3D LES high-temperature buoyant jet using laser absorption spectroscopy. This final case provided useful information enabling the simulation to match the experiment better than seen previously, and also provided additional insight about the real-world system. Each application is a novel use of ABC to find parameters in flows of complex thermal-fluid systems. Parameters successfully estimated include: a parameter analogous to the Reynolds number for the 0D Navier Stokes eqn., inlet mean temperature and variability, inlet mean velocity, inlet composition, inlet velocity frequency and amplitude, rate of rotation for a roller placed above the jet, and heat addition size and strength above the jet to model heat addition resulting from continued combustion. Estimating each of these parameters has required identifying useful statistics within the flow domain that are connected to the parameters of interest and that reduce the dimensionality of the problem substantially. Specific conclusions related to the ABC technique and each case follow.

6.1 Conclusions

- Several comments can be made about the conclusions to be drawn concerning the statistics used throughout the dissertation. Generally summary statistics have been required to reduce the dimensionality of the data from thousands or millions of locations in space and time, to a handful. Despite this drastic reduction, the summary statistics have usually enabled unknown parameters to be identified. Statistics taken with the same variable as the unknown parameter (i.e. velocity measurements to predict velocity parameter) generally performed best, though alternate approaches (temperature measurements to predict velocity parameters) also proved to be effective. Also, for cases when the statistic and parameter variable matched, observations closest to the boundary of interest performed best. When the variable was not the same, then space may be required between the boundary and measurement to allow the information to propagate from one field to the other.

- The rejection distance, ε, is an important component of ABC that determines the interpretation of results. Several strategies for selecting ε are proposed, however all still rely on user judgment to ensure a balance exists between Monte Carlo error from having too few simulations accepted and an unbiased simulation from too closely resembling the prior distribution, while also keeping in mind computational affordability.

- Though only applied to the 0D Navier-Stokes equations, the Markov chain Monte Carlo (MCMC) ABC approached proved very useful at concentrating effort in regions of the parameter space with higher likelihood of producing acceptable simulation results. This provides helpful guidance for choosing parameters specific to the MCMC approach such as which transition kernel to use and how should it be specified.

- ABC using RANS simulations and reference data performed exceptionally well due to the ability of the simulations to converge very well and their often monotonic change to variations in input parameters. This enabled very slight changes in input parameters to

yield appreciable differences in observations.

- Cases relying on LES reference data performed well, though required an ensemble average for the reference data. This served as a "best case scenario" (or OSSE) to demonstrate the capability of ABC to predict parameters in a very complex and dynamic flow field.
- Cases relying on DNS reference data performed well, though required an ensemble average for the reference data which can be expensive to compute. This case demonstrated the impact resulting from model bias; if bias is small then the parameters can still be well predicted. For larger bias, however, parameter estimations become much less certain. Experimental reference data provides the most practically useful but challenging opportunity for parameter estimation using ABC. The temperature profiles obtained from ABC allow better agreement between simulations and experiments than has been seen previously (for the catalytic burner shown in this dissertation). Additionally, the parameters resulting from ABC provide previously unknown information regarding the system, namely the most likely inlet velocity and the amount (or lack) of combustion above the catalytic burner.

6.2 Future Work

Several directions exist for future work and are presented next. These possibilities show a handful of the bountiful future research opportunities for researchers using approximate Bayesian computation (ABC).

- One of the most pressing future directions is to implement a more efficient method whereby fewer simulations are rejected by careful sampling within the parameter domain space to select more parameters with favorable statistics compared to the reference observations. Two techniques that are adaptable to this need are sequential Monte Carlo [162] and ABC using MCMC [2].
- More efficient ABC algorithms would allow more expensive simulations since fewer would be required. This would enable future research opportunities to model combustion within

the domain using reaction mechanisms. These results could then utilize an additional observation from laser absorption spectroscopy, namely species concentration.

- Gains in efficiency would enable more complex parameter estimation problems such as predicting more parameters simultaneously (e.g. spatially and temporally varying temperature and velocity).

- Turbulence parameters would ideally be found during the ABC process to ensure that the turbulence modeling is optimal for the given application (see [5] for initial work in this direction).

- ABC could be used to determine the optimal placement and required quantity of measurements to obtain converged posterior estimates of parameters. This would inform future experimental campaigns to ensure the requisite number of observations are obtained while avoiding superfluous experiments.

- Parameters predicted through ABC could drive higher fidelity simulations that would then provide additional information about the flow field not available from lower fidelity simulations or through experiments.

Bibliography

[1] D. J. Thomson. Spectrum estimation and harmonic analysis. Proceedings of the IEEE, 70(9):1055–1096, 1982.

[2] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences, 100(26):15324–15328, 2003.

[3] B. M. Cetegen, Y. Dong, and M. C. Soteriou. Experiments on stability and oscillatory behavior of planar buoyant plumes. Physics of Fluids, 10(7):1658–1665, 1998.

[4] W. L. Oberkampf and C. J. Roy. Verification and validation in scientific computing. Cambridge University Press, 2010.

[5] O. A. Doronina, J. D. Christopher, C. A. Z. Towery, P. E. Hamlington, and W. J. A. Dahm. Autonomic Closure for Turbulent Flows Using Approximate Bayesian Computation. AIAA Paper, AIAA-2018-0594, 2018.

[6] W. L. Oberkampf and T. Trucano. Validation methodology in computational fluid dynamics. In Fluids 2000 Conference and Exhibit, page 2549, 2000.

[7] W. L. Oberkampf and T. G. Trucano. Verification and validation in computational fluid dynamics. Progress in Aerospace Sciences, 38(3):209–272, 2002.

[8] W. L. Oberkampf, T. G. Trucano, and C. Hirsch. Verification, validation, and predictive capability in computational engineering and physics. Applied Mechanics Reviews, 57(5):345–384, 2004.

[9] D. C. Estumano, F. C. Hamilton, M. J. Colaço, A. J. K. Leiroz, H. R. B. Orlande, R. N. Carvalho, and G. S. Dulikravich. Bayesian estimate of mass fraction of burned fuel in internal combustion engines using pressure measurements. Engineering Optimization IV, pages 997–1003, 2015.

[10] S. Mosbach, A. Braumann, P. L. W. Man, C. A. Kastner, G. P. E. Brownbridge, and M. Kraft. Iterative improvement of Bayesian parameter estimates for an engine model by means of experimental design. Combustion and Flame, 159(3):1303–1313, 2012.

[11] P. G. Constantine, Q. Wang, A. Doostan, and G. Iaccarino. A surrogate accelerated Bayesian inverse analysis of the HyShot II flight data. AIAA Paper, AIAA-2011-2037, 2011.

[12] I. Vrbik, R. Deardon, Z. Feng, A. Gardner, and J. Braun. Using individual-level models for infectious disease spread to model spatio-temporal combustion dynamics. Bayesian Analysis, 7(3):615–638, 2012.

[13] A. H. Elsheikh, I. Hoteit, and M. F. Wheeler. Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. Computer Methods in Applied Mechanics and Engineering, 269:515–537, 2014.

[14] S. Yeşilyurt and A. Patera. Surrogates for numerical simulations; optimization of eddy-promoter heat exchangers. Computer Methods in Applied Mechanics and Engineering, 121(1):231–257, 1995.

[15] L. Zeng, L. Shi, D. Zhang, and L. Wu. A sparse grid based Bayesian method for contaminant source identification. Advances in Water Resources, 37:1–9, 2012.

[16] A. G. Salinger, R. P. Pawlowski, J. N. Shadid, and B. G. van Bloemen Waanders. Computational analysis and optimization of a chemical vapor deposition reactor with large-scale computing. Industrial & Engineering Chemistry Research, 43(16):4612–4623, 2004.

[17] W. Jahn, G. Rein, and J. L. Torero. Forecasting fire dynamics using inverse computational fluid dynamics and tangent linearisation. Advances in Engineering Software, 47(1):114–126, 2012.

[18] H. Madsen. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in Water Resources, 26(2):205–216, 2003.

[19] S. Wang and X. Xu. Parameter estimation of internal thermal mass of building dynamic models using genetic algorithm. Energy Conversion and Management, 47(13-14):1927–1941, 2006.

[20] E. Pemha and E. Ngo Nyobe. Genetic algorithm approach and experimental confirmation of a laser-based diagnostic technique for the local thermal turbulence in a hot wind tunnel jet. Progress In Electromagnetics Research, 28:325–350, 2011.

[21] H. Kato, A. Yoshizawa, G. Ueno, and S. Obayashi. A data assimilation methodology for reconstructing turbulent flows around aircraft. Journal of Computational Physics, 283:559–581, 2015.

[22] X. Gao, Y. Wang, N. Overton, M. Zupanski, and X. Tu. Data-assimilated computational fluid dynamics modeling of convection-diffusion-reaction problems. Journal of Computational Science, 21:38–59, 2017.

[23] J. Sousa, C. García-Sánchez, and C. Gorlé. Improving urban flow predictions through data assimilation. Building and Environment, 132:282–290, 2018.

[24] J. R. Stroud, M. Katzfuss, and C. K. Wikle. A bayesian adaptive ensemble kalman filter for sequential state and parameter estimation. Monthly Weather Review, 146(1):373–386, 2018.

[25] 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. Journal of Computational Physics, 324:115–136, 2016.

[26] W. N. Edeling, M. Schmelzer, R. P. Dwight, and P. Cinnella. Bayesian predictions of reynolds-averaged navier–stokes uncertainties using maximum a posteriori estimates. AIAA Journal, 56(5):2018–2029, 2018.

[27] A. P. Singh and K. Duraisamy. Using field inversion to quantify functional errors in turbulence closures. Physics of Fluids, 28(4):045110, 2016.

[28] M. Khalil and H. N. Najm. Probabilistic inference of reaction rate parameters from summary statistics. Combustion Theory and Modelling, pages 1–31, 2018.

[29] J. Tu, G-H. Yeoh, and C. Liu. Computational fluid dynamics: a practical approach. Butterworth-Heinemann, 2018.

[30] C. Tropea, A. L. Yarin, and J. F. Foss. Springer handbook of experimental fluid mechanics, volume 1. Springer Science & Business Media, 2007.

[31] S. R. Turns. An introduction to combustion, volume 499. McGraw-hill New York, 1996.

[32] R. D. Moser, J. Kim, and N. N. Mansour. Direct numerical simulation of turbulent channel flow up to re τ= 590. Physics of fluids, 11(4):943–945, 1999.

[33] P. Moin and K. Mahesh. Direct numerical simulation: a tool in turbulence research. Annual review of fluid mechanics, 30(1):539–578, 1998.

[34] P. E. Hamlington, A. Y. Poludnenko, and E. S. Oran. Interactions between turbulence and flames in premixed reacting flows. Physics of Fluids, 23(12), 2011.

[35] P. E. Hamlington, D. Krasnov, T. Boeck, and J. Schumacher. Statistics of the energy dissipation rate and local enstrophy in turbulent channel flow. Physica D, 241:169–177, 2012.

[36] C-H. Moeng. A large-eddy-simulation model for the study of planetary boundary-layer turbulence. Journal of the Atmospheric Sciences, 41(13):2052–2062, 1984.

[37] P. Sagaut. Large eddy simulation for incompressible flows: an introduction. Springer Science & Business Media, 2006.

[38] P. E. Hamlington, L. P. Van Roekel, B. Fox-Kemper, K. Julien, and G. P. Chini. Langmuir–submesoscale interactions: descriptive analysis of multiscale frontal spindown simulations. Journal of Physical Oceanography, 44(9):2249–2272, 2014.

[39] C. H. Moeng and P. P. Sullivan. Large-eddy simulation. Encyclopedia of Atmospheric Sciences, 2:232–240, 2015.

[40] J. R. Edwards, J. A. Boles, and R. A. Baurle. Large-eddy/reynolds-averaged navier–stokes simulation of a supersonic reacting wall jet. Combustion and Flame, 159(3):1127–1138, 2012.

[41] R. Pecnik, V. E. Terrapon, F. Ham, G. Iaccarino, and H. Pitsch. Reynolds-averaged navier-stokes simulations of the hyshot ii scramjet. AIAA journal, 50(8):1717–1732, 2012.

[42] P. E. Hamlington and W. J. A. Dahm. Computational validation of new reynolds stress closure for nonequilibrium effects in turbulent flows. AIAA Paper, 1323:2009, 2009.

[43] National Research Council et al. Getting up to speed: The future of supercomputing. National Academies Press, 2005.

[44] P. Machamer. The stanford encyclopedia of philosophy, Dec 2014.

[45] R. K. Hanson, R. M. Spearrin, and C. S. Goldenstein. Spectroscopy and optical diagnostics for gases. Springer, 2015.

[46] E. Brookner. Tracking and Kalman Filtering Made Easy. Wiley Online Library, 1998.

[47] F. Rabier. Overview of global data assimilation developments in numerical weather-prediction centres. Quarterly Journal of the Royal Meteorological Society, 131(613):3215–3233, 2005.

[48] I. M. Navon. Data assimilation for numerical weather prediction: a review. In Data assimilation for atmospheric, oceanic and hydrologic applications, pages 21–65. Springer, 2009.

[49] J. Anderson. Welcome to the data assimilation research testbed - dart, 2016.

[50] G. Evensen. Data assimilation: the ensemble Kalman filter. Springer Science & Business Media, 2009.

[51] PHILIPPE Courtier, J-N Thépaut, and Anthony Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. Quarterly Journal of the Royal Meteorological Society, 120(519):1367–1387, 1994.

[52] Fuqing Zhang, Meng Zhang, and James A Hansen. Coupling ensemble kalman filter with four-dimensional variational data assimilation. Advances in Atmospheric Sciences, 26(1):1–8, 2009.

[53] S. G. Benjamin, D. Dévényi, S. S. Weygandt, K. J. Brundage, J. M. Brown, G. A. Grell, D. Kim, B. E. Schwartz, T. G. Smirnova, and T. L. Smith. An hourly assimilation-forecast cycle: The RUC. Monthly Weather Review, 132(2):495–518, 2004.

[54] J. Mandel, J. D. Beezley, J. L. Coen, and M. Kim. Data assimilation for wildland fires. Control Systems, IEEE, 29(3):47–65, 2009.

[55] Xinfeng Gao, Yijun Wang, Nathaniel Overton, Ian May, and Xuemin Tu. Data assimilated computational fluid dynamics algorithm for combustion. In 54th AIAA Aerospace Sciences Meeting, page 1810, 2016.

[56] J. R. Edwards, C. H. Patton, H. Mirgolbabaei, T. J. Wignall, and T. Echekki. 4D Data Assimilation for Large Eddy Simulation of High Speed Turbulent Combustion. AIAA Paper, AIAA-2015-3836, 2015.

[57] 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. Journal of Computational Physics, 324:115–136, 2016.

[58] M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz. Approximate Bayesian Computation. PLoS Computational Biology, 9(1):e1002803, 2013.

[59] J. M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate Bayesian computational methods. Statistics and Computing, 22(6):1167–1180, 2012.

[60] J. Lintusaari, M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander. Fundamentals and recent developments in approximate bayesian computation. Systematic biology, 66(1):e66–e82, 2017.

[61] Jan Snyman. Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms, volume 97. Springer Science & Business Media, 2005.

[62] J. Farrell and A. Pontau. Co-optimization of fuels & engines, 2016.

[63] A. O'Hagan. Bayesian analysis of computer code outputs: a tutorial. Reliability Engineering & System Safety, 91(10):1290–1300, 2006.

[64] Stefano Conti, John Paul Gosling, Jeremy E Oakley, and Anthony O'Hagan. Gaussian process emulation of dynamic computer codes. Biometrika, 96(3):663–676, 2009.

[65] S. Bhattacharya. A simulation approach to Bayesian emulation of complex dynamic computer models. Bayesian Analysis, 2(4):783–815, 2007.

[66] M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate bayesian computation in population genetics. Genetics, 162(4):2025–2035, 2002.

[67] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface, 6(31):187–202, 2009.

[68] K. Csilléry, M. G. B. Blum, O. E. Gaggiotti, and O. François. Approximate Bayesian computation (ABC) in practice. Trends in Ecology & Evolution, 25(7):410–418, 2010.

[69] M. A. Beaumont. Approximate bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution and Systematics, 41(379-406):1, 2010.

[70] N. J. R. Fagundes, N. Ray, M. A. Beaumont, S. Neuenschwander, F. M. Salzano, S. L. Bonatto, and L. Excoffier. Statistical evaluation of alternative models of human evolution. Proceedings of the National Academy of Sciences, 104(45):17614–17619, 2007.

[71] J. A. Vrugt and M. Sadegh. Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. Water Resources Research, 49(7):4335–4345, 2013.

[72] D. J. Nott, L. Marshall, and J. Brown. Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? Water Resources Research, 48(12), 2012.

[73] M. Sadegh and J. A. Vrugt. Bridging the gap between GLUE and formal statistical approaches: Approximate Bayesian computation. Hydrology and Earth System Sciences, 17(12), 2013.

[74] B. Olson. Stochastic weather generation with approximate bayesian computation. Master's thesis, University of Colorado at Boulder, 2016.

[75] B. Olson and W. Kleiber. Approximate Bayesian computation methods for daily spatiotemporal precipitation occurrence simulation. Water Resources Research, pages 1–21, 2017.

[76] D. D. Lucas, A. Gowardhan, P. Cameron-Smith, and R. L. Baskett. Impact of meteorological inflow uncertainty on tracer transport and source estimation in urban atmospheres. Atmospheric Environment, 143:120–132, 2016.

[77] A. B. Abdessalem, N. Dervilis, D. Wagg, and K. Worden. Model selection and parameter estimation in structural dynamics using approximate bayesian computation. Mechanical Systems and Signal Processing, 99:306–325, 2018.

[78] M. K. Vakilzadeh, J. L. Beck, and T. Abrahamsson. Using approximate bayesian computation by subset simulation for efficient posterior assessment of dynamic state-space model classes. SIAM Journal on Scientific Computing, 40(1):B168–B195, 2018.

[79] J. D. Christopher, N. T. Wimer, C. Lapoint, T. R. S. Hayden, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter estimation for complex thermal-fluid flows using approximate bayesian computation. Physical Review Fluids, 2018.

[80] J. D. Christopher, D. J. Petrykowski, T. R. S. Hayden, C. Lapointe, N. T. Wimer, S. P. Nigam, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter estimation in turbulent simulations using approximate bayesian computation with laser spectroscopy. Experiments in Fluids, 2018.

[81] T. R. S. Hayden, D. J. Petrykowski, A. Sanchez, S. P. Nigam, C. Lapointe, J.D. Christopher, N.T. Wimer, A. Upadhye, M. Strobel, P.E. Hamlington, and G.B. Rieker. Characterization of oh, h 2 o, and temperature profiles in industrial flame treatment systems interacting with polymer films. Proceedings of the Combustion Institute, 2018.

[82] T. R. S. Hayden, N. T. Wimer, C. Lapoint, J. D. Christopher, S. P. Nigam, A. Upadhye, M. Strobel, P. E. Hamlington, and G. B. Rieker. Characterization of the buoyant jet above a catalytic combustor uinsg wavelength modulations spectroscopy. International Journal of Heat and Mass Transfer, 2018.

[83] J. D. Christopher, D. J. Petrykowski, T. R. S. Hayden, C. Lapointe, N. T. Wimer, S. P. Nigam, I. Grooms, P. E. Hamlington, and G. B. Rieker. Parameter estimation using wavelength modulation spectroscopy temperature measurements and approximate bayesian computation. In Optics and Photonics for Energy and the Environment. Optical Society of America, 2018.

[84] J. D. Christopher, N. T. Wimer, T. R. S. Hayden, C. Lapointe, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter Estimation for a Turbulent Buoyant Jet Using Approximate Bayesian Computation. AIAA Paper, AIAA-2017-0531, 2017.

[85] J. D. Christopher, C. Lapointe, N. T. Wimer, T. R. S. Hayden, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter Estimation for a Turbulent Buoyant Jet with Rotating Cylinder Using Approximate Bayesian Computation. AIAA Paper, AIAA-2017-3629, 2017.

[86] T. R. S. Hayden, N. T. Wimer, C. Lapointe, J. D. Christopher, A. S. Makowiecki, P. E. Hamlington, and G. B. Rieker. Characterization of the output from a catalytic combustor using wavelength modulation spectroscopy. In 47th AIAA Fluid Dynamics Conference, page 4424, 2017.

[87] T. Hayden, N. Wimer, C. Lapointe, J. Christopher, P.E. Hamlington, and G. B. Rieker. Characterization of a jet above a catalytic combustor using wavelength modulation spectroscopy. Proceedings of the 10th U.S. National Meeting on Combustion, 2017.

[88] T. R. S. Hayden, N. T. Wimer, C. LaPointe, J. D. Christopher, S. Nigam, P. E. Hamlington, and G. B. Rieker. Wavelength modulation spectroscopy of OH radical in an industrial flame. In Optics and Photonics for Energy and the Environment, pages ETh2A–4. Optical Society of America, 2017.

[89] C. Lapointe, J. D. Christopher, N. T. Wimer, T. R. S. Hayden, G. B. Rieker, and P. E. Hamlington. Optimization for internal turbulent compressible flows using adjoints. In 23rd AIAA Computational Fluid Dynamics Conference, page 4115, 2017.

[90] N. T. Wimer, C. Lapointe, T. R. S. Hayden, J. D. Christopher, A. Y. Poludnenko, G. B. Rieker, and P. E. Hamlington. Near-and far-field properties of high-temperature turbulent buoyant jets. In 47th AIAA Fluid Dynamics Conference, page 4423, 2017.

[91] O. A. Doronina, C. A. Z. Towery, J. D. Christopher, I. Grooms, and P. E. Hamlington. Turbulence Model Development Using Markov Chain Monte Carlo Approximate Bayesian Computation. AIAA Paper, AIAA, 2019.

[92] J. D. Christopher, N. T. Wimer, C. Lapointe, T. R. S. Hayden, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter estimation for a pulsating turbulent buoyant jet using approximate Bayesian computation. Bulletin of the American Physical Society, 2017.

[93] J. D. Christopher, N. T. Wimer, T. R. S. Hayden, C. Lapointe, I. Grooms, G. B. Rieker, and P. E. Hamlington. Parameter Estimation for a Turbulent Buoyant Jet Using Approximate Bayesian Computation. Bulletin of the American Physical Society, 61, 2016.

[94] O. Doronina, J. D. Christopher, P. E. Hamlington, and W. J. A. Dahm. Autonomic closure for turbulent flows using approximate Bayesian computation. Bulletin of the American Physical Society, 2017.

[95] S. P. Nigam, C. Lapointe, J. D. Christopher, N. T. Wimer, T. R. S. Hayden, G. B. Rieker, and P. E. Hamlington. Flame structure and dynamics for an array of premixed methane-air jets. Bulletin of the American Physical Society, 2017.

[96] N. Wimer, C. Lapointe, T. R. S. Hayden, J. D. Christopher, G. B. Rieker, and P. E. Hamlington. Effects of exit variability on near-field statistics for turbulent buoyant jets. In Bulletin of the American Physical Society, 2016.

[97] C. Lapointe, N. T. Wimer, T. R. S. Hayden, J. D. Christopher, G. B. Rieker, and P. E. Hamlington. Scaling analysis of temperature variability between a rotating cylinder and a turbulent buoyant jet. In Bulletin of the American Physical Society, 2016.

[98] F. Hartig, J. M. Calabrese, B. Reineking, T. Wiegand, and A. Huth. Statistical inference for stochastic simulation models–theory and application. Ecology Letters, 14(8):816–827, 2011.

[99] D. B. Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. The Annals of Statistics, 12(4):1151–1172, 1984.

[100] S. Tavaré, D. J. Balding, R. C. Griffiths, and P. Donnelly. Inferring coalescence times from DNA sequence data. Genetics, 145(2):505–518, 1997.

[101] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Molecular Biology and Evolution, 16(12):1791–1798, 1999.

[102] Oliver Ratmann, Christophe Andrieu, Carsten Wiuf, and Sylvia Richardson. Model criticism based on likelihood-free inference, with an application to protein network evolution. Proceedings of the National Academy of Sciences, 106(26):10576–10581, 2009.

[103] Michael GB Blum and Viet Chi Tran. Hiv with contact tracing: a case study in approximate bayesian computation. Biostatistics, 11(4):644–660, 2010.

[104] J. Rohmer, M. Rousseau, A. Lemoine, R. Pedreros, J. Lambert, and A. Benki. Source characterisation by mixing long-running tsunami wave numerical simulations and historical observations within a metamodel-aided ABC setting. Stochastic Environmental Research and Risk Assessment, pages 1–18.

[105] Gareth W Peters, Scott A Sisson, and Yanan Fan. Likelihood-free bayesian inference for α-stable models. Computational Statistics & Data Analysis, 56(11):3743–3756, 2012.

[106] Brandon M Turner and Trisha Van Zandt. A tutorial on approximate bayesian computation. Journal of Mathematical Psychology, 56(2):69–85, 2012.

[107] R. C. Smith. Uncertainty quantification: theory, implementation, and applications, volume 12. SIAM, 2013.

[108] W. Neiswanger, C. Wang, and E. Xing. Asymptotically exact, embarrassingly parallel mcmc. arXiv preprint arXiv:1311.4780, 2013.

[109] L. Murray. Distributed markov chain monte carlo. In Proceedings of Neural Information Processing Systems workshop on learning on cores, clusters and clouds, volume 11, 2010.

[110] J. Corander, M. Gyllenberg, and T. Koski. Bayesian model learning based on a parallel mcmc strategy. Statistics and computing, 16(4):355–362, 2006.

[111] S. C. Crow and F. H. Champagne. Orderly structure in jet turbulence. Journal of Fluid Mechanics, 48(3):547–591, 1971.

[112] J. A. Lovett and S. R. Turns. Experiments on axisymmetrically pulsed turbulent jet flames. AIAA Journal, 28(1):38–46, 1990.

[113] R. B. Farrington and S. D. Claunch. Infrared imaging of large-amplitude, low-frequency disturbances on a planar jet. AIAA journal, 32(2):317–323, 1994.

[114] S. Marzouk, H. Mhiri, S. El Golli, G. Le Palec, and P. Bournot. Numerical study of momentum and heat transfer in a pulsed plane laminar jet. International Journal of Heat and Mass Transfer, 46(22):4319–4334, 2003.

[115] W. Kriaa, H. B. Cheikh, H. Mhiri, G. Le Palec, and P. Bournot. Numerical study of free pulsed jet flow with variable density. Energy Conversion and Management, 49(5):1141–1155, 2008.

[116] Charles P Arnold Jr and Clifford H Dey. Observing-systems simulation experiments: Past, present, and future. Bulletin of the American Meteorological Society, 67(6):687–695, 1986.

[117] Ming Xue, Mingjing Tong, and Kelvin K Droegemeier. An osse framework based on the ensemble square root kalman filter for evaluating the impact of data from radar networks on thunderstorm analysis and forecasting. Journal of Atmospheric and Oceanic Technology, 23(1):46–66, 2006.

[118] Y. Wang, P. Chatterjee, and J. L. de Ris. Large eddy simulation of fire plumes. Proceedings of the Combustion Institute, 33(2):2473–2480, 2011.

[119] C. Greenshields. OpenFOAM | The OpenFOAM Foundation, March 2016.

[120] H. G. Weller, G. Tabor, H. Jasak, and C. Fureby. A tensorial approach to computational continuum mechanics using object-oriented techniques. Computers in Physics, 12(6):620–631, 1998.

[121] A. Yoshizawa. Statistical theory for compressible turbulent shear flows, with the application to subgrid modeling. Physics of Fluids, 29(7):2152–2164, 1986.

[122] W. Sutherland. LXXV. A dynamical theory of diffusion for non-electrolytes and the molecular mass of albumin. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 9(54):781–785, 1905.

[123] D. R. Stull and H. Prophet. JANAF thermochemical tables. Technical report, DTIC Document, 1971.

[124] L. Le Cam and G. L. Yang. Asymptotics in statistics: Some basic concepts. Springer Science & Business Media, 2012.

[125] D. J. C. MacKay. Information theory, inference and learning algorithms. Cambridge university press, 2003.

[126] D. B. Percival and A. T. Walden. Spectral analysis for physical applications. Cambridge University Press, 1993.

[127] W. Van Drongelen. Signal processing for neuroscientists: an introduction to the analysis of physiological signals. Elsevier, 2006.

[128] D. W. Scott. On optimal and data-based histograms. Biometrika, 66(3):605–610, 1979.

[129] M. G. B. Blum. Regression approaches for approximate bayesian computation. arXiv preprint arXiv:1707.01254, 2017.

[130] M. G. B. Blum and O. François. Non-linear regression models for approximate bayesian computation. Statistics and Computing, 20(1):63–73, Jan 2010.

[131] R. P. Dwight and Z.H. Han. Efficient uncertainty quantification using gradient-enhanced kriging. AIAA Paper, AIAA-2009-2276, 2009.

[132] OpenFOAM/OpenFOAM-2.2.x.

[133] B. E. Launder and D. B. Spalding. The numerical computation of turbulent flows. Computer methods in applied mechanics and engineering, 3(2):269–289, 1974.

[134] G. B. Rieker, J. B. Jeffries, and R. K. Hanson. Calibration-free wavelength-modulation spectroscopy for measurements of gas temperature and concentration in harsh environments. Applied Optics, 48(29):5546–5560, 2009.

[135] G. B. Rieker. Wavelength-modulation spectroscopy for measurements of gas temperature and concentration in harsh environments. PhD thesis, Stanford University, 2009.

[136] F. Menter and T. Esch. Elements of industrial heat transfer prediction. 16th Brazilian Cong. of Mech. Eng.(Uberlandia, Brazil), 2001.

[137] T. R. S. Hayden and G. B. Rieker. Large amplitude wavelength modulation spectroscopy for sensitive measurements of broad absorbers. Optics Express, 24(24):27910–27921, 2016.

[138] Y. Egorov and F. Menter. Development and application of SST-SAS turbulence model in the DESIDER project. Advances in Hybrid RANS-LES Modelling, pages 261–270, 2008.

[139] J. M. Bonnie and G. Sanford. Computer program for calculation of complex chemical equilibrium compositions and applications. User's Manual and Program Discription, 1996.

[140] S. A. Cottilard. Catalytic combustion. Nova Science Publ., 2011.

[141] T. R. S. Hayden, N. Malarich, D. J. Petrykowski, S. P. Nigam, J. D. Christopher, C. Lapointe, N. T. Wimer, P.E. Hamlington, and G. B. Rieker. Oh radical measurements in combustion environments using wavelength modulation spectroscopy and dual frequency comb spectroscopy near 1491 nm. Applied Physics B, 2018.

[142] M. G. Allen. Diode laser absorption sensors for gas-dynamic and combustion flows. Measurement Science and technology, 9(4):545, 1998.

[143] M. Lackner. Tunable diode laser absorption spectroscopy (tdlas) in the process industries–a review. Reviews in Chemical Engineering, 23(2):65–147, 2007.

[144] J. W. Daily. Laser induced fluorescence spectroscopy in flames. Progress in energy and combustion science, 23(2):133–199, 1997.

[145] C. S. Mcenally, L. D. Pfefferle, A. M. Schaffer, M. B. Long, R. K. Mohammed, M. D. Smooke, and M. B. Colkei. Characterization of a coflowing methane/air non-premixed flame with computer modeling, rayleigh-raman imaging, and on-line mass spectrometry. Proceedings of the Combustion Institute, 28(2):2063–2070, 2000.

[146] H Li, A Farooq, JB Jeffries, and RK Hanson. Near-infrared diode laser absorption sensor for rapid measurements of temperature and water vapor in a shock tube. Applied Physics B, 89(2-3):407–416, 2007.

[147] M. A. Bolshov, Y. A. Kuritsyn, and Y. V. Romanovskii. Tunable diode laser spectroscopy as a technique for combustion diagnostics. Spectrochimica Acta Part B: Atomic Spectroscopy, 106:45–66, 2015.

[148] T. Cai, T. Tan, G. Wang, W. Chen, and X. Gao. Gas temperature measurements using wavelength modulation spectroscopy at 1.39 μm. Optica Applicata, 39(1), 2009.

[149] V. Ebert, T. Fernholz, C. Giesemann, H. Pitz, H. Teichert, J. Wolfrum, and H. Jaritz. Simultaneous diode-laser-based in situ detection of multiple species and temperature in a gas-fired power plant. Proceedings of the Combustion Institute, 28(1):423–430, 2000.

[150] R. K. Hanson and D. F. Davidson. Recent advances in laser absorption and shock tube methods for studies of combustion chemistry. Progress in Energy and Combustion Science, 44:103–114, 2014.

[151] Q. Huang, F. Wang, H. Zhang, J. Yan, M. Ni, and K. Cen. In-situ co measurement of gas and oil combustion flame using near infrared tunable diode laser with direct and modulated absorption signals. Optics Communications, 306:99–105, 2013.

[152] Z. Qu, R. Ghorbani, D. Valiev, and F. M. Schmidt. Calibration-free scanned wavelength modulation spectroscopy–application to h 2 o and temperature sensing in flames. Optics express, 23(12):16492–16499, 2015.

[153] S. Wagner, B. T. Fisher, J. W. Fleming, and V. Ebert. Tdlas-based in situ measurement of absolute acetylene concentrations in laminar 2d diffusion flames. Proceedings of the Combustion Institute, 32(1):839–846, 2009.

[154] K. Sun, X. Chao, R. Sur, C. S. Goldenstein, J. B. Jeffries, and R. K. Hanson. Analysis of calibration-free wavelength-scanned wavelength modulation spectroscopy for practical gas sensing using tunable diode lasers. Measurement Science and Technology, 24(12):125203, 2013.

[155] C. S. Goldenstein, C. L. Strand, I. A. Schultz, K. Sun, J. B. Jeffries, and R. K. Hanson. Fitting of calibration-free scanned-wavelength-modulation spectroscopy spectra for determination of gas properties and absorption lineshapes. Applied optics, 53(3):356–367, 2014.

[156] C. S. Goldenstein and R. K. Hanson. Diode-laser measurements of linestrength and temperature-dependent lineshape parameters for h2o transitions near 1.4 μm using voigt, rautian, galatry, and speed-dependent voigt profiles. Journal of Quantitative Spectroscopy and Radiative Transfer, 152:127–139, 2015.

[157] R. K. Hanson and P. K. Falcone. Temperature measurement technique for high-temperature gases using a tunable diode laser. Applied Optics, 17(16):2477–2480, 1978.

[158] T. Chai and R. R. Draxler. Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. Geoscientific model development, 7(3):1247–1250, 2014.

[159] L. S. Rothman, I. E. Gordon, Y. Babikov, A. Barbe, D. C. Benner, P. F. Bernath, M. Birk, L. Bizzocchi, V. Boudon, and L. R. Brown. The hitran2012 molecular spectroscopic database. Journal of Quantitative Spectroscopy and Radiative Transfer, 130:4–50, 2013.

[160] M. Yamaguchi, T. Iriguchi, T. Nakazawa, and C-C. Wu. An observing system experiment for typhoon conson (2004) using a singular vector method and dotstar data. Monthly Weather Review, 137(9):2801–2816, 2009.

[161] F. Bouttier and G. Kelly. Observing-system experiments in the ecmwf 4d-var data assimilation system. Quarterly Journal of the Royal Meteorological Society, 127(574):1469–1488, 2001.

[162] S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences, 104(6):1760–1765, 2007.