Parameter Estimation for a Turbulent Buoyant Jet with Rotating Cylinder Using Approximate Bayesian Computation

Jason D. Christopher*, Caelan Lapointe*, Torrey R. S. Hayden*,
Nicholas T. Wimer*, Aniruddha Upadhye†, Melvyn Branch‡,
Mark Strobel§, Ian Grooms¶, Gregory B. Rieker‖, and Peter E. Hamlington‖

Assigning accurate boundary, initial, and geometric conditions, as well as fluid properties, in numerical simulations of real-world applications continues to be a major challenge in computational fluid dynamics research. However, recent advances have provided new and increasingly sophisticated methods for accurately determining these conditions and properties. This paper focuses on a data-driven parameter estimation technique called Approximate Bayesian Computation (ABC), which allows numerical simulation parameters to be determined from experimental or other "truth" data. In this proof-of-concept study, the ABC approach is demonstrated for a rotating cylinder above a high-temperature turbulent buoyant jet, which is an engineering problem of relevance to the commercial aerospace industry. Here the "truth" observations come from a two-dimensional Reynolds-averaged Navier Stokes (RANS) simulation that serves as a known test case against which other simulations are compared. In particular, the "truth" case has known jet inflow and cylinder rotational velocities, and we show that the ABC approach is able to correctly identify the true values of these parameters. This success indicates that ABC can be extended to real-world engineering systems and that, in the future, ABC will allow experimental observations to accurately drive the selection of boundary, initial, and geometric conditions, as well as fluid properties, in numerical simulations.

Nomenclature

θ	Parameter of interest
D	Data (or observation)
$p(D)$	Probability distribution function of observation
$p(D\vert\theta)$	Probability of the observed data given the parameter, i.e., the Likelihood
$p(\theta\vert D)$	Probability of the parameters given the observed data, i.e., the Posterior
$p(\theta)$	Probability distribution function of the parameter, i.e., the Prior
S	Statistics of the data, D
$\delta(\cdot,\cdot)$	Distance function
ε	Rejection distance threshold

I. Introduction

Once a system of governing equations and a numerical approach have been specified, the accuracy with which numerical simulations are able to model real-world flow and combustion problems is largely determined

*PhD Student, Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309
†Senior Research Engineer, Corporate Research Process Laboratory, 3M Corporation, St. Paul, MN 55144
‡Professor Emeritus, Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309
§Corporate Scientist and Group Leader, Corporate Research Process Laboratory, 3M Corporation, St. Paul, MN 55144
¶Assistant Professor, Department of Applied Mathematics, University of Colorado, Boulder, CO 80309
‖Assistant Professor, Department of Mechanical Engineering, University of Colorado, Boulder, CO 80309

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 problem and the 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 to improve simulation accuracy, Approximate Bayesian Computation (ABC) is used here to reliably estimate boundary conditions in a simulation of a rotating cylinder above a turbulent buoyant jet. This particular configuration is relevant to flame treatments of polymer films and other materials used in the commercial aerospace industry. The simulations performed here are two-dimensional (2D) Reynolds-averaged Navier-Stokes (RANS) simulations, and this work is one of the first attempts to apply ABC for a complex engineering-relevant fluid dynamics problem using a RANS approach. This study builds upon work originally developed by Christopher *et al.* [1, 2], where ABC was used to determine unknown parameters for a zero-dimensional fluids problem and an open turbulent buoyant jet.

Previously the ABC approach has been primarily applied to biological challenges,[3–7] although its use has been expanding to other application areas such as geophysical sciences.[8–12] A significant amount of research has attempted to use Bayesian methods[13–16] to estimate parameters for engineering applications, but all such attempts have required the calculation of a complex likelihood function. In order to obtain this likelihood function, researchers typically make assumptions that limit the usefulness of their model. A major simplification used by many Bayesian practitioners involves substituting a surrogate model[14,15,17,18] in place of a more robust simulation.

In the present study, we have chosen to keep a robust and complete RANS solution instead of a more approximate surrogate model. Moreover, observations from the case we wish to match come from a "truth" simulation using the same RANS model with known parameters; by using the same model for the truth observations and comparison simulations provides the groundwork to estimate parameters for a complex flow simulation using ABC with limited truth data.

II. Background

Computational simulations provide a wealth of data about how real-world engineering problems behave in a variety of settings. Due to the ease of changing parameters, numerical methods easily transition from standard operational conditions to extreme bounding cases. Through their results, engineers and scientists typically try to accomplish two primary tasks: (*i*) characterization and analysis, and (*ii*) design and optimization. Each of these tasks presents unique challenges, yet together they can yield valuable insights into performance and design for a variety of applications, including industrial processing, jet and rocket engines, internal combustion engines, and power generation. In the present study, we focus on characterization and analysis, with a specific emphasis on combining RANS simulations with "truth" RANS simulations, and eventually, combining RANS simulations with experimental data to improve the accuracy of the simulations. For more details concerning design and optimization of complex flows, see, for example, King *et al.* [19].

II.A. Characterization and Analysis

The two primary methods for characterization and analysis of real-world systems are numerical simulations and experiments. In general, theoretical or analytical approaches are not feasible for most real-world problems due to the complexity of real systems (e.g., geometry, chemical kinetics, turbulent flow fields, highly-coupled physics, etc.).

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, geometry, and conditions. On the other end of the spectrum are extremely expensive, but much more accurate, computer models built from three-dimensional (3D) CAD-generated geometries and involving the solution of physics-based coupled partial differential equations. Such computational fluid dynamics (CFD) approaches include direct numerical simulations (DNS), large eddy simulations (LES), and RANS simulations. In the present study, we focus on relatively fast RANS simulations in order to obtain flow fields comparable to more costly LES results (for more details on the simulations and their comparison to LES results see Section IV.A).

Despite the high computational cost of many simulations, they are still often considered to be "cheaper" than experiments, particularly in terms of financial cost and time. 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.

In contrast, experimental methods can answer a variety of engineering analysis questions and typically cannot avoid involving all of the relevant physics. Experimental measurements can be made extremely accurate based on calibrations against known conditions. However, compared to simulations, experimental measurements typically only provide a sparse picture of a system. For example, only a limited number of measurement locations and times, and a limited number of parameters. can be probed (e.g., only the temperature or species concentrations at a few locations, instead of all locations). Moreover, experiments are generally more expensive because they require costly measurement equipment in addition to the expenses of designing and building the actual system to be measured. Constructing and configuring the measurement equipment as well as the system to be measured takes a considerable amount of time, in addition to the duration of the experiments themselves. Lastly, experiments can be inherently more dangerous than computer simulations, prohibiting tests of extreme conditions (especially in combustion).

By combining the strengths of simulations and experiments, the unreliability of simulations and the sparsity of experimental data can be circumvented. A key challenge within engineering research is thus the use of sparse experimental information to improve simulation accuracy.

II.B. Approximate Bayesian Computation (ABC)

The ABC method allows one to determine unknown parameters in a simulation by comparing simulation results with the full data or statistics from experimental, or "truth", observations. For additional details regarding the choice to use ABC over other data-driven techniques see [1].

The ABC technique[20–22] is based on Bayes' theorem, $P(\theta|D) = P(D|\theta) * P(\theta)/P(D)$, and solves for the probability of the parameters, θ, given the observations or data (D) available through experimentation or, alternatively, another implementation of a model which may be more refined (i.e., the goal is to solve for $P(\theta|D)$, also known as the posterior). Note that the data, D, can also be represented by its statistics, $\mathcal{S}$ (and for the approximate data, $\hat{D}$, the statistics are $\hat{\mathcal{S}}$). The probability density function of the unknown parameter, independent of any 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, though is often taken out of the equation since it can be treated as a normalization constant. Lastly, the probability of the data occurring given a particular parameter, $P(D|\theta)$, is known as the likelihood function; the likelihood function is usually very difficult to obtain either due to it being mathematically intractable or computationally unfeasible. This is because, for a given set of parameters in a stochastic system, it is not possible to know how the solution will progress and thus explicitly determine what measurements you would expect to obtain. 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 is most relevant when there is no available likelihood function; instead of requiring the exact likelihood function $P(D|\theta)$, the ABC method instead computes many approximate likelihood values (denoted $P(\hat{D}|\theta)$). To eventually generate these approximate likelihood values, the method first samples values of the parameter 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 what prior to use is based, as the name implies, on prior information a practitioner has about what possible values a parameter might take. The choice of what prior to use should not significantly impact the final posterior distribution, however it may increase the computational cost if it is very dissimilar to the posterior since many simulated parameters will be rejected. The method next takes a value of θ (i.e., a parameter) and runs the model forward to come up with model data (approximate value of $P(D|\theta)$, namely $P(\hat{D}|\theta)$). ABC then compares each approximate realization ($P(\hat{D}|\theta)$) with the experimental data, $P(D)$, and either keeps the parameters used to produce that approximate likelihood, or rejects the parameters depending on a predefined distance function and associated threshold. The distance function is simply a method to compare a simulation's results to an observation. The distance function could, for instance, compare the mean of the observations to the mean of the approximate likelihood.

When performing the comparison, the metric used must accurately represent the data; if this is true, it means the statistics are sufficient to describe the data and are therefore useful when comparing the approximate likelihood and the observation. Choosing an appropriate distance function along with the appropriate statistic to analyze the flow field is a complicated task because many statistics will not contain the desired information about the flow field. Thus, choosing an appropriate and informative statistic is a key task to successfully implementing ABC algorithms. Once a comparison is made between each simulation and the "truth", 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, $P(\theta|D)$, is known as the posterior and is indeed the very item one seeks to find in the parameter estimation problem.

Looking more closely at the particular application of a buoyant jet with a rotating cylinder above it, ABC can be used to determine unknown parameter values. For an example observation, laser diagnostic methods[23,24] could be used to obtain line-of-sight average temperature measurements at various locations within the flow field. These measurements would have very rich temporal data that could be used to generate a discrete probability density function with associated relevant statistics (e.g., 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, 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 (i.e., an approximate likelihood, $P(\hat{D}|\theta)$) 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. Note that a Markov chain Monte Carlo (MCMC) approach can be utilized to increase parameter acceptance rates. 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, $P(\theta|D)$.

In summary, ABC is a powerful tool to develop estimates for parameters given a set of 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 accepted parameters make up the desired posterior distribution.

III. Approximate Bayesian Computation for a Turbulent Buoyant Jet with Rotating Cylinder

An ABC algorithm was implemented to predict unknown parameter values in a simulation of a rotating cylinder above a high-temperature turbulent buoyant jet. The first step in the analysis was to simulate a RANS solution using known inflow and cylinder rotational velocities. These parameters served as true parameters that subsequent simulations would try to determine. The temperature and velocity at several points within the flow field were stored every 0.1 seconds. Next, assuming the "true" parameters (inflow velocity and cylinder velocity) were unknown, a range of parameters was chosen and a simulation was run. The inlet velocities and rotational velocities 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 set of parameters needs to be large enough to ensure it contains the "true" solution, however the wider it gets the more simulations that need to be run, thereby significantly increasing the simulation cost. Once the "truth" 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 the statistics of the "truth" case. If the results are similar, the simulated inlet velocity and cylinder rotational velocity 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 to draw conclusions about the results.

This ABC algorithm, known as method D in Marjoram *et al.*,[21] is summarized as follows:

1. Generate parameter θ from the prior distribution $P(\theta)$.
2. Simulate $\hat{D}$ 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 the generic parameter, θ, that one seeks to find is a pair containing one inlet velocity and one cylinder rotational velocity for this particular study.

IV. Approximate Bayesian Computation using Reynolds-averaged Navier-Stokes Simulation

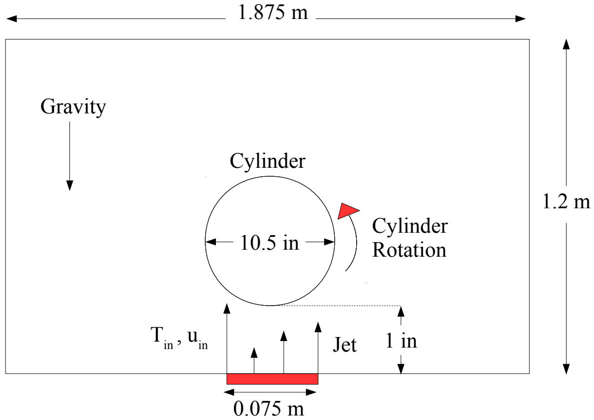
With the ABC method described in Section III, a simulation was performed in 2D using the RANS equations. The open source computational fluid dynamics software OpenFOAM, version 2.2.x, was used to design and run RANS simulations for this study.[25,26] The compressible RANS equations were solved in conjunction with the energy equation. 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. The Menter SST k-ω two-equation viscosity closure model[27] 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 to obtain a solution. Limiters on velocity divergence and Laplacian schemes were implemented to aid convergence; under relaxation also helped solution convergence. Fluid viscosity and specific heat were assumed to be constant for the purpose of this study. The RANS simulation compares favorably to a three-dimensional large-eddy simulation.

IV.A. Simulation Setup

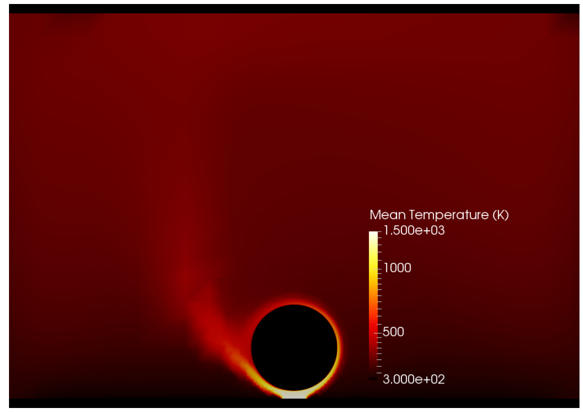
The RANS equations were solved on a 2D grid with a high temperature jet inflow centered along the bottom of the domain. A schematic of the setup is shown in Figure 1(a) and an example temperature field is shown in Figure 1(b).

The domain dimensions were chosen to allow the bottom-driven jet to exit the domain primarily vertically due to the vertical inflow velocity, velocity imparted by the rotating cylinder, and buoyant forces. The jet inflow is centered at the bottom of the domain and is 0.075 m wide. Each node of the jet inflow is prescribed a temperature and velocity. Note that the inlet temperature and velocity are fixed spatially and temporally, but could be allowed to vary in each dimension according to a prescribed distribution. The jet temperature for all simulations shown is held fixed at 1500 K. The jet is centered under a 10.5 inch diameter cylinder, with a 1 inch gap separating the jet and cylinder. The cylinder rotates counterclockwise about its center. The cylinder rotational velocity varies with each simulation to values between 3.0 rad/s and 28.0 rad/s, and is held fixed for the duration of each simulation.

Fixed temperature and velocity boundary conditions were imposed for the jet inlet and cylinder surface. The two were treated as walls; the cylinder temperature and angular velocity were fixed, as were the inlet temperature and vertical velocity. All other domain boundaries were left open and allowed fluid flow in or



(a) Simulation domain



(b) Example temporally averaged temperature field

Figure 1. (a) Schematic showing the setup of the RANS domain. (b) Temperature field averaged over the last 15 seconds of the simulations with inlet velocity 1 m/s and rotational velocity of 12.5 rad/s.

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 inlet and cylinder. Pressure and temperature boundary conditions were used to ease relaxation to ambient conditions at domain boundaries.

The ABC method applied is given in Section III; this is the basic ABC algorithm with no Markov-chain method. The MCMC-ABC algorithm is an area of future research. A simulation is run with a specified and constant velocity at the inlet and for the cylinder. The resulting flow field is then compared to a "truth" simulation. The "truth" simulation is an execution of the same code with known conditions. The comparison is 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 the full temperature probability distribution function.

IV.B. Turbulent Convection Simulation Results

A simulation was run using a domain width of 1.875 meters and height of 1.2 meters. The jet inflow was prescribed a uniform temperature and velocity. The vertical component of the inflow velocity was assigned between 0.0 m/s and 1.0 m/s, with a velocity of 0.5 m/s representing the "true" velocity. The cylinder rotational velocity was assigned a value between between 3.0 rad/s and 28.0 rad/s, with a cylinder angular velocity of 12.5 rad/s representing the "true" velocity. The flow of the hot gas enters the domain through the jet inflow and then diffuses into the domain 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 center of the jet, as well as 1 cm toward the center from either side of the jet. Measurements were recorded at approximately 10 separate heights between the jet and cylinder for each of the three locations.

Figure 2 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 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 cumulative mean is shown in the figure and clearly stabilizes after 15 seconds into the simulation.

The ABC methodology was used to successfully identify the "true" prescribed parameter values (0.5 m/s for inlet velocity, and 12.5 rad/s cylinder rotational velocity). A key aspect of ABC involves choosing a suitable statistic with which to analyze the flow. This often involves studying the flow field and observing changes in its behavior based on changes to the parameters. For this geometry and setup, an increase in inlet velocity corresponds to an increase in measured temperature above the jet. 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

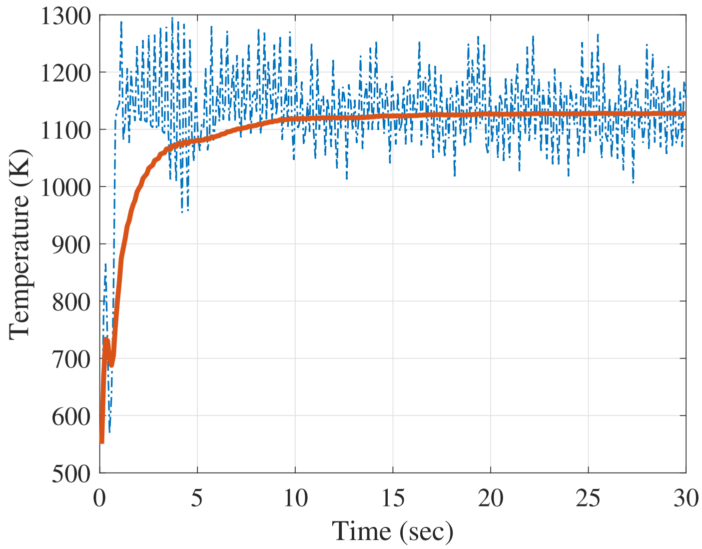


Figure 2. Temperature versus time at a height of 4.1 mm above the jet surface, centered in the domain. The dash-dot blue line indicates the instantaneous temperature, while the solid red line shows the cumulative mean.

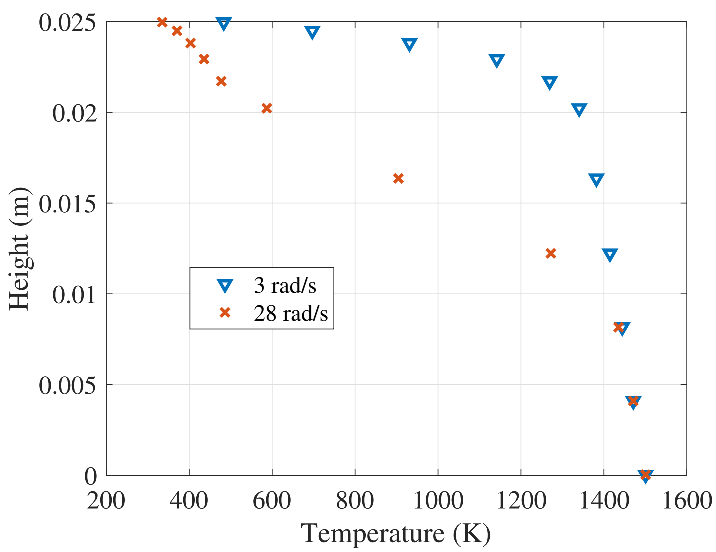


Figure 3. Plot shows height above the hot jet versus the temporal average (for last 15 seconds) of simulation temperature at the center location width-wise in domain with inlet velocity equal to 0.5 m/s. The blue triangles correspond to a cylinder rotational velocity of 3 rad/s while the red "x" marker corresponds to a rotational velocity of 28 rad/s. This plot shows the temperatures are significantly higher for the case when the cylinder is rotating more slowly (blue triangles).

jet so that the influence of the rotating cylinder on the flow field will be minimized, one can readily predict a "true" inlet velocity. However, this statistic provides little information regarding the rotational velocity of the cylinder.

In order to gain more insight into how the cylinder rotation impacts the flow field, the mean temporal temperature was observed at each measured height location for the slowest rotational rate (3 rad/s) and the highest rotational rate (28 rad/s). These results are shown in Figure 3. One can see that more of the locations have high temperatures when the cylinder is spinning more slowly, however when the cylinder spins quickly lower temperatures are observed throughout the measured domain. This trend is observed 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 of the cooler air into the gap. Conversely, when it spins slowly its boundary layer entrains less cold air, the wake separation moves further from the hot jet,

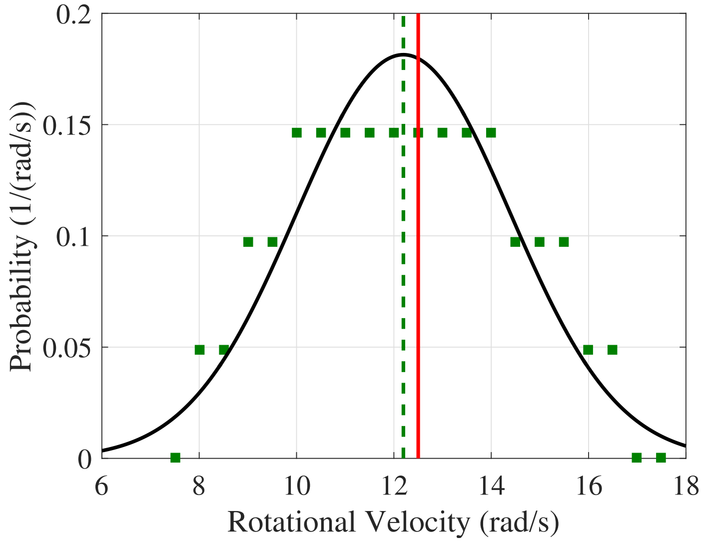


Figure 4. Discrete Probability Density Function (PDF) showing the number of accepted simulations versus values of cylinder rotational velocity along with a Gaussian curve with the same mean and standard deviation as the accepted values. One-thousand seventy one simulations were run with unique uniform temperatures applied across the jet and unique cylinder rotational velocities. The ABC method rejected rotational velocities that were too high and too low compared to the "true" value by comparing the mean temperature in the domain at one location and across *all vertical positions* for measurements taken at the center width-wise location. "True" rotational velocity shown in red at 12.5 m/s with the mean of the accepted values shown in teal at 12.2 rad/s.

and thus the gap's temperature is dominated by the hot jet.

With this information now available, a second statistic was developed to compare each simulation to the "truth" case. The second metric is found by first taking the temporal mean of the temperature observed in the gap for a given location. Then, this value is averaged with the other temporal means found at different heights (for the same location). Thus, the second metric quantifies the spatial and temporal mean temperature observed within the gap. For the first metric, a rejection distance of 3.5 K was applied when comparing the temporal mean at one location in each simulation to the truth simulation. The measurement location for the first metric is at 4.1 mm above the jet surface, centered in the domain. For the second metric comparing the temporal and spatial mean temperatures, a rejection distance of 95 K was applied. This second metric compared temperatures taken at locations centered in the domain between the jet and the cylinder and 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. These rejection distances were chosen to reduce the number of accepted cases converging toward the "true" solution, while simultaneously leaving a sufficiently high number of accepted cases to maintain confidence in the results.

Figure 5 shows the resulting histogram of the jet inflow temperature. Note that the uniform distribution of initial velocities spanning the full range of 0.0 m/s to 1.0 m/s shrinks into a much more narrow and accurate uniform distribution centered close to the "true" value of 0.5 m/s. Moreover, all velocities below 0.45 m/s and above 0.55 m/s have been rejected by the ABC algorithm. The mean of the accepted inlet velocities is 0.501 m/s, an error of less than 1%. Similarly, Figure 4 shows a histogram of the cylinder rotational velocity. Note that, again, the nearly uniform distribution of initial velocities morphs into a much narrower distribution of accepted velocities centered around the mean at 12.2 rad/s, an error of 2.4% compared to the "true" value of 12.5 rad/s.

IV.C. Turbulent Convection Simulation Results with Fewer Measurements

The experimental burden to obtain this many simultaneous temperature measurements would be immense. Therefore, a reduced set of measurements is desired that produces comparable posterior parameter distributions. Based on the initial metrics used, a reduced set of measurements was found. This set combines a measurement close to the jet, centered, at a height of 4.1 mm and one additional measurement closer to the cylinder, also centered and at a height of 22.9 mm. This pair of measurements captures much of the information about the flow field, resulting in the histograms seen in Figures 5 and 6. The rejection distances

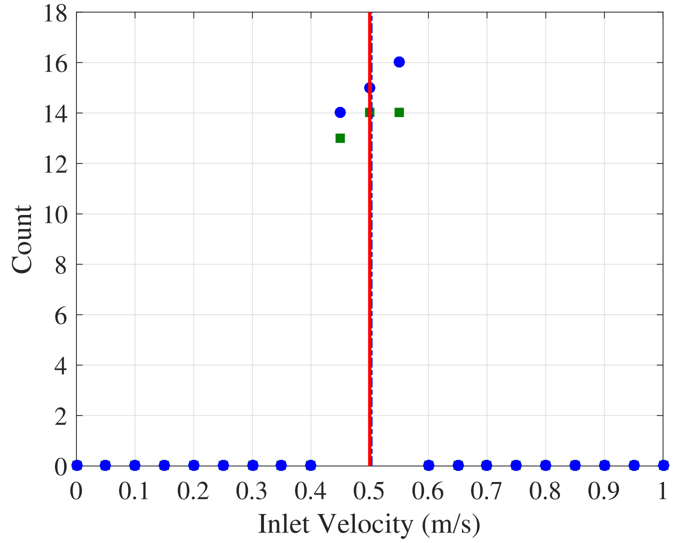


Figure 5. Scatter plot showing the number of accepted simulations versus all simulated values of jet inlet velocity. One-thousand seventy one simulations were run with unique uniform temperatures applied across the jet and unique cylinder rotational velocities. The ABC method rejected velocities that were too high and too low compared to the "true" value by comparing the temporal mean temperature in the domain at one location (metric 1) and by comparing the temporal and spatial means across vertical positions for measurements taken at the center width-wise location (metric 2). "True" velocity shown in red at 0.5 m/s with the mean of the accepted values using all measurements for metric 2 shown in green at 0.501 m/s, while the mean of the accepted values using only two measurement locations for metric 2 is shown in blue at 0.502 m/s.

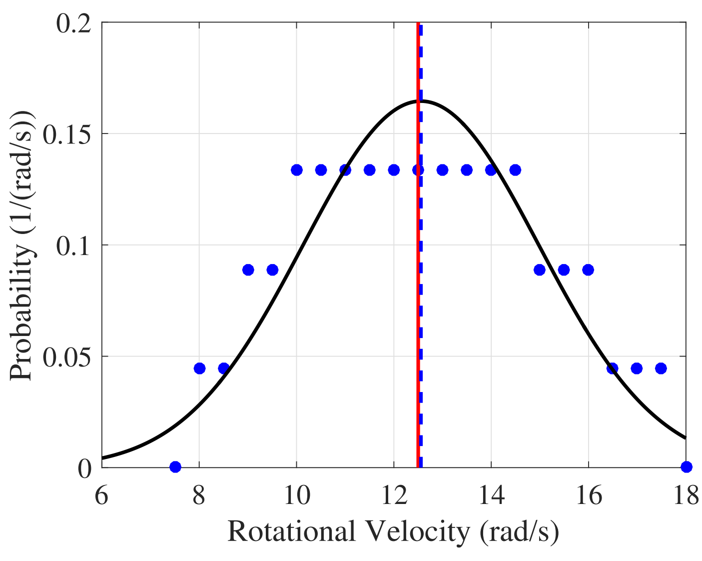


Figure 6. Discrete PDF showing the number of accepted simulations versus values of cylinder rotational velocity along with a Gaussian curve with the same mean and standard deviation as the accepted values. One-thousand seventy one simulations were run with unique uniform temperatures applied across the jet and unique cylinder rotational velocities. The ABC method rejected rotational velocities that were too high and too low compared to the "true" value by comparing the mean temperature in the domain at one location and across *two vertical positions* for measurements taken at the center width-wise location. "True" 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 teal at 12.6 rad/s.

used were 3.5 K for the first metric (temporal mean at 8.1 mm) and 125 K for the second metric (temporal and spatial mean of both measurements). These results are very promising because they show that spatially sparse measurements can provide useful information about the flow field when predicting flow parameters.

V. 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 herein demonstrate that ABC is a powerful technique to estimate parameters of interest in complex RANS simulations, while leaving several questions to be answered in our future efforts. First, how well will this procedure work on more complicated 3-dimensional geometries? Next, will this procedure extend to "true" statistics obtained experimentally instead of through an additional iteration of the same simulation? Additionally, what conclusions can be drawn about the appropriate statistics to use for ABC in engineering applications? Future work will provide answers to these questions and explore further methods to accomplish parameter estimation.

Acknowledgments

N.T.W., C.L., and T.R.S.H. acknowledge gift support from the 3M Company. J.D.C. and G.B.R. were supported, in part, by NSF Award No. CBET 1454496. P.E.H. was supported, in part, by AFOSR Award No. FA9550-14-1-0273. This work utilized the Janus supercomputer, which is supported by the National Science Foundation (award number CNS-0821794), the University of Colorado Boulder, the University of Colorado Denver, and the National Center for Atmospheric Research (NCAR). The Janus supercomputer is operated by the University of Colorado Boulder.

References

[1] Jason D Christopher, Nicholas T Wimer, Torrey R S Hayden, Caelan Lapointe, Ian Grooms, Gregory B Rieker, and Peter E Hamlington. Parameter estimation for a turbulent buoyant jet using approximate bayesian computation. In *55th AIAA Aerospace Sciences Meeting*, 2017.

[2] Jason D Christopher, Nicholas T Wimer, Torrey RS Hayden, Caelan Lapointe, Ian Grooms, Gregory B Rieker, and Peter E Hamlington. Parameter estimation for a turbulent buoyant jet using approximate bayesian computation. *Bulletin of the American Physical Society*, 61, 2016.

[3] Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.

[4] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael PH 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.

[5] Katalin Csilléry, Michael GB Blum, Oscar E Gaggiotti, and Olivier François. Approximate bayesian computation (abc) in practice. *Trends in ecology & evolution*, 25(7):410–418, 2010.

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

[7] Nelson J. R. Fagundes, Nicolas Ray, Mark Beaumont, Samuel Neuenschwander, Francisco M. Salzano, Sandro L. Bonatto, and Laurent Excoffier. Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences*, 104(45):17614–17619, 2007.

[8] Jasper A Vrugt and Mojtaba Sadegh. Toward diagnostic model calibration and evaluation: Approximate bayesian computation. *Water Resources Research*, 49(7):4335–4345, 2013.

[9] David J Nott, Lucy Marshall, and Jason Brown. Generalized likelihood uncertainty estimation (glue) and approximate bayesian computation: What's the connection? *Water Resources Research*, 48(12), 2012.

[10] M Sadegh and JA Vrugt. Bridging the gap between glue and formal statistical approaches: approximate bayesian computation. *Hydrology and Earth System Sciences*, 17(12), 2013.

[11] Branden Olson. *Stochastic Weather Generation with Approximate Bayesian Computation*. PhD thesis, University of Colorado at Boulder, 2016.

[12] Donald D Lucas, Akshay Gowardhan, Philip Cameron-Smith, and Ronald L Baskett. Impact of meteorological inflow uncertainty on tracer transport and source estimation in urban atmospheres. *Atmospheric Environment*, 143:120–132, 2016.

[13] DC Estumano, FC Hamilton, MJ Colaço, AJK Leiroz, HRB Orlande, RN Carvalho, and GS Dulikravich. Bayesian estimate of mass fraction of burned fuel in internal combustion engines using pressure measurements. *Engineering Optimization IV*, pages 997–1003, 2015.

[14] Sebastian Mosbach, Andreas Braumann, Peter LW Man, Catharine A Kastner, George PE Brownbridge, and Markus Kraft. Iterative improvement of bayesian parameter estimates for an engine model by means of experimental design. *Combustion and Flame*, 159(3):1303–1313, 2012.

[15]Paul G Constantine, Qiqi Wang, Alireza Doostan, and Gianluca Iaccarino. A surrogate accelerated bayesian inverse analysis of the hyshot ii flight data. *American Institute of Aeronautics and Astronautics, Reston, VA*, 2011.

[16]Irene Vrbik, Rob Deardon, Zeny Feng, Abbie Gardner, John Braun, et al. Using individual-level models for infectious disease spread to model spatio-temporal combustion dynamics. *Bayesian Analysis*, 7(3):615–638, 2012.

[17]Richard P Dwight and Zhong-Hua Han. Efficient uncertainty quantification using gradient-enhanced kriging. *AIAA Paper*, 2276:2009, 2009.

[18]Serhat Yeşilyurt et al. Surrogates for numerical simulations; optimization of eddy-promoter heat exchangers. *Computer methods in applied mechanics and engineering*, 121(1):231–257, 1995.

[19]Ryan N King, Peter E Hamlington, Peter Graf, and Katherine Dykes. Adjoint optimization of wind farm layouts for systems engineering analysis. In *34th Wind Energy Symposium*, page 2199, 2016.

[20]Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate Bayesian Computation. *PLoS Comput Biol*, 9(1):e1002803, 2013.

[21]Paul Marjoram, John Molitor, Vincent Plagnol, and Simon Tavaré. Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328, 2003.

[22]Jean-Michel Marin, Pierre Pudlo, Christian P Robert, and Robin J Ryder. Approximate bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.

[23]Gregory B Rieker, Jay B Jeffries, and Ronald K Hanson. Calibration-free wavelength-modulation spectroscopy for measurements of gas temperature and concentration in harsh environments. *Applied optics*, 48(29):5546–5560, 2009.

[24]Gregory Brian Rieker. *Wavelength-modulation spectroscopy for measurements of gas temperature and concentration in harsh environments*. PhD thesis, Stanford University, 2009.

[25]Chris Greenshields. OpenFOAM | The OpenFOAM Foundation, March 2016.

[26]OpenFOAM/OpenFOAM-2.2.x.

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