ABSTRACT

WIGNALL, T.J. Development of a Data Fusion Methodology for Lineload Aerodynamic Databases for a Launch Vehicle during Liftoff and Transition. (Under the direction of Tarek Echekki and Jack Edwards.)

The need for databases for the distributed loading on launch vehicles during the early portion of flight necessitates the use of expensive computational flows in regimes where wake effects dominate. While also being expensive, this is a regime that computational tools tend to historically have problems simulating accurately. To help tackle this problem, a method of data fusion to combine computational results to wind tunnel derived force and moment data is developed. Using this method, significant reduction in computational costs and increases in confidence of the final product is possible and has been used to generate several databases for the Space Launch System (SLS) at NASA. While the full details of database generation are not part of this work, the crucial method at its core is developed here. Two SLS geometries are used throughout the work to demonstrate the techniques. These are two of the larger geometries and represent both planned crewed missions to the Moon as well as potential cargo missions to deep space.

The method uses principal component analysis (PCA) to generate a reduced ordered model (ROM) to help fill in the full parameter space. Other similar techniques are explored, but were not found to have a significant result on the predictions of the ROM. Because the full number of components are kept to generate the model, this lack of difference is expected. This method is then extended to ensure that predicted surfaces match trusted force and moment data derived from wind tunnel testing. This extension is done by setting up a constrained optimization problem in order to minimize the deviation from the surface resolved computational data while still integrating to the desired values. When generating the constrained optimization problem, a weighting factor to balance these competing needs is introduced. The work compares previously introduced weighting terms from similar work to the proposed terms and shows that the previously used terms do not have as desirable behavior in this flow regime.

This method is then expanded by developing a technique to incorporate uncertainty quantification into the developed data fusion methodology. This expansion takes a two pronged approach. One examines transferring the uncertainties in the force and moment database and characterizes how those adjustments change the predicted lineloads. The second looks at model form error and looks how rebuilding the model using slightly different data changes the predictions. These two terms are then combined in order to create an uncertainty model that takes both effects into account. The limitations of the proposed methods is then discussed as well as possible techniques to address these shortcomings.

© 2024 United States Government as represented by the Administrator of the National Aeronautics and Space Administration. No copyright is claimed in the United States under Title 17, U.S. Code. All Other Rights Reserved.

Development of a Data Fusion Methodology for Lineload Aerodynamic Databases for a Launch Vehicle during Liftoff and Transition

by
T.J. Wignall

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Aerospace Engineering

Raleigh, North Carolina

2024

APPROVED BY:

Tarek Echekki Co-chair of Advisory Committee	Jack Edwards Co-chair of Advisory Committee
Hong Luo	Ralph Smith

Jeremy Pinier

DEDICATION

To my parents.

BIOGRAPHY

After graduating high school, the author attended Old Dominion University where he received his Bachelors and Masters degrees. During his doctoral work, he accepted a job at NASA Langley where he continues to work. He currently resides in Richmond, Virginia.

ACKNOWLEDGEMENTS

I would like to begin by thanking Heather Houlden of TMC Technologies for her assistance on the uncertainty quantification chapter without which it would not have been possible. Between development of insightful graphics and guidance in methodologies it would not have been possible.

I am grateful to SLS Aerodynamics under Induced Environments team for providing the resources, data, and support necessary to accomplish this work. In particular the NASA Langley Research Center SLS CFD team. This work would not be possible without the help of both current and previous members. Without the contributions of Steve Krist, Nalin Ratnayake, Farhad Ghaffari, Karen Deere, Brent Pomeroy, and Michael Lee this work would be much more limited in scope. This team was essential in gathering the CFD data that was necessary to be able to perform this work as well as understanding and analyzing the results of the methodologies developed here. Special gratitude is extended to Steve Krist and Brent Pomeroy for their work in developing CFD solution visualizations many of which made it into this work. I would also like to highlight the contributions of Derek Dalle of the NASA Ames Research Center who provided significant inspiration, technical discussions, and several diagrams. The crucial wind tunnel data used in this work is thanks to the work of David Chan for both leading the test as well as the post-test analysis of the data. The data would also not be possible without the facility and its team led by Les Yeh at the time. Without the experimental data from the tunnel, understanding the complex aerodynamics of the SLS would not be possible. I would like to thank both Jeremy Pinier and Patrick Shea for helping me brainstorm, review technical content, and problem solve throughout this process.

On a personal level, I would like to first thank my long time partner Qui Nguyen for the support she has provided me. Without her making sure I took care of myself as well, the quality of this work would have greatly suffered. I would also like to thank all my other friends and family for their support as well.

Finally I would like to thank my advisor for his help and patience as I changed topics and took my time to complete this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Nomenclature

Acronyms

BSM	=	Booster Separation Motors
CFD	=	Computational Fluid Dynamics
CGT	=	Chimera Grid Tools
CRM	=	Common Research Model
DDES	=	Delayed Detached Eddy Simulation
DFROM	=	Data-fused Reduced Ordered Model
EDL	=	Entry Descent and Landing
EUS	=	Exploration Upper Stage
LES	=	Large Eddy Simulation
LOO	=	Leave One Out
LOT	=	Liftoff and Transition
LVSA	=	Launch Vehicle Stage Adaptor
PBROM	=	Physics Based Reduced Ordered Model
PCA	=	Principal Component Analysis
PCLAM	=	PCLAM Computes Lineloads and Moments
PC	=	Principal Component
POD	=	Proper Orthogonal Decomposition
RANS	=	Reynolds-Averaged Navier-Stokes
ROM	=	Reduced Ordered Model
RSS	=	Root Sum of Squares
SA	=	Spalart-Allmaras
SLS	=	Space Launch System
SRB	=	Solid Rocket Booster
SVD	=	Singular Value Decomposition
UQ	=	Uncertainty Quantification

Variables

a	=	ROM Coefficient
$\hat{a}$	=	DFROM Coefficient
$\overrightarrow{ll}_x$	=	Lineload in x direction
C_x	=	Coefficient of Force in x direction
C_p	=	Coefficient of Pressure
F_x	=	Force in x direction

l_{ref}	=	Reference length
S_{ref}	=	Reference Area
S	=	Solution on the surface of the vehicle
R	=	Covariance Matrix
Re	=	Reynolds Number based on centerbody diameter
V	=	Eigenvectors of the Covariance Matrix
w	=	weight
α_P	=	Total Angle of Attack
λ	=	Legrange Multiplier
ϕ_P	=	Roll Angle
σ	=	Stress Tensor (Chapter 2), Standard Deviation (Chapter 5)
Ψ	=	Principal Component

Superscripts, Subscripts, and Others

$\hat{\square}$	=	Normalized version (except for DFROM Coefficient, $\hat{a}$)
$\square_i$	=	Generic index
$\square_p$	=	Denotes Missile Axis Coordinate System
$\square^k$	=	Denotes Vehicle Component (Centerbody, LSRB, or RSRB)
$\square^l$	=	Denotes Variable (Pressure or Three surface friction directions)

CHAPTER

1

MOTIVATION AND OUTLINE

The driving motivation of this work is to improve the process of constructing lineload databases for the Space Launch System (SLS) during prelaunch, and the early stages of flight called liftoff and transition (LOT). As further explained in Section 2.1.1, lineloads are the distributed forces on a vehicle much like a lift distribution on a wing. While experimental methods such as pressure sensitive paint are constantly improving, the easiest and most accurate way to generate these lineloads is computational fluid dynamics (CFD). However, the fluid dynamics for this problem are dominated by massively separated wakes as the vehicle is exposed to ground winds, which are difficult to model accurately. These difficulties increase when many cases are needed to generate a database and accuracy is often sacrificed to reduce computational cost. Since it is impractical to run CFD at every desired database breakpoint, some form of surrogate modeling is necessary. To further decrease computational cost without sacrificing too much accuracy, a reduced order model (ROM) is used to fill in data between solutions. Other surrogate modeling techniques such as radial basis functions and linear interpolation have been used previously; however, the ROM that was formulated allows for feature extraction, which allows adjustment of these features to modify the integrated forces and moments.

While the distributed forces come from computational sources, the integrated loads are developed using extensive wind tunnel testing. This causes a slight mismatch between the predicted loading from the lineload database and the wind tunnel derived force and moment database. Because of the compromises in the fidelity to decrease the computational cost, the integrated loads from the wind tunnel test are better trusted. As a result, a methodology to adjust the distributed loads to match these more trusted data and remove the discrepancies between databases is desired. A data-fused ROM (DFROM) method is developed to bridge the gap between the two databases.

While nominal values are the primary values in an aerodynamic database, they do not allow for

follow-up analysis. By characterizing the uncertainty on the values in the databases, engineering decisions can be made in the event of unexpected circumstances. However, uncertainty quantification (UQ) is difficult especially on spatially distributed data. Further complicating matters, many solutions must be generated in order to properly characterize UQ for CFD. But, the data-fused method makes extensive use of the better characterized wind tunnel data, which allows for that uncertainty to be used in its place. The UQ methodology developed uses information from the uncertainty in the forces and moments as well as an estimate of the model form error. Finally, some of the outstanding limitations of the method are discussed and possible directions of future work.

1.1 Computational Cost and Reduced Order Models

One of the significant drivers for this research is the computational cost of individual simulations necessary to build a liftoff and transition database. The solver used, CREATE-AV Kestrel (see Sec 2.1.3.1), makes use of dual mesh approaches that allows one unstructured volume solver in the near body grid and a faster Cartesian solver in the far field, which also allows for feature based grid adaptation. Even with this efficient scheme, each case takes three to five days of wall time when using approximately 800 processors. The database these cases are used for is made up of about 450 breakpoints which would cost approximately 45 million processor-hours for a given configuration. The SLS program has six different planned configurations which have significant differences, meaning that separate databases are needed for each one. On occasion, a given configuration changes due to updated requirements or because more is known as the vehicle matures, which requires creating a new database. For example, the Block 1B Cargo has had different cargo fairing designs over the years to accommodate changes to planned missions necessitating the need for further analysis.

These huge computational costs have inspired research into methods to reduce the amount of data necessary to fully characterize the parameter space. ROMs have proven highly valuable in fluid dynamics to help reduce the cost of analyzing large parameter spaces. ROMs are often deployed for design optimization and uncertainty quantification, but here they are used primarily to evaluate the lineloads across the domain of interest. Chapter 3 gives the development and overview of the ROM used to fill the parameter space and reduce the computational costs of generating databases. Principal Component Analysis (PCA), also called Proper Orthogonal Decomposition (POD), methods are explored to provide predictions of the pressure and friction on the surface of the vehicle. These surface values are then integrated into lineloads, which are the quantities of interest. While the ROM provides predictions similar to the baseline CFD, there are still disagreements between the integrated loads provided by wind tunnel testing and the new ROM.

1.2 Differences Between Experimental and Computational Databases and the Implementation of Data Fusion

In a project, such as SLS, the differences between the distributed forces provided by this ROM and the integrated forces developed through wind tunnel testing must be addressed and accommodated given the importance of the two sources of data. When both are considered valid measures of forces for the vehicle, it is unclear which should be used if they give different results. Because of the difficulty of simulating a wake dominated flow and the compromises to computational fidelity in exchange for significant reduction of computational cost, the wind tunnel derived forces and moments are considered higher fidelity. As an added bonus, the wind tunnel derived forces and moments exist at all the required database breakpoints, which prevents the need of interpolation or other space filling methods. The rich spatial information provided by the CFD is informative and provides knowledge of local phenomenon around geometric features of interests, such as various supports and flanges that are necessary to accurately predict lineloads. The goal is to develop strategies to combine the best of both of these rich datasets.

Chapter 4 discusses in detail the data fusion method developed that allows for slight adjustments based on the underlying features that the ROM uses to predict new results. This method forces the predicted lineloads to integrate to the desired forces and moments while keeping as close to the original predictions of the ROM and the underlying CFD. Several variations of the methodology are developed and the one with the best characteristics is selected and used for further analysis. A simple constrained optimization problem is set up and used to balance the various requirements. After development of the DFROM, the performance is explored and a UQ model is developed.

1.3 Uncertainty Quantification on Distributed Loads

The last technical chapter of the paper explores the development of uncertainty quantification on the developed model. UQ is becoming increasingly necessary in aerodynamic modeling to allow for higher performing designs. By properly characterizing the uncertainty and the probability of each condition, tighter margins are possible allowing for cheaper fabrication of aircraft (and launch vehicles) through design or material changes. It also allows for proper risk assessment in the case of unforeseen circumstances. Before the first launch of the SLS for Artemis I, the vehicle was left on the launch pad during a hurricane. While winds stayed below designed maximum scenarios, by properly accounting for the possible error in database delivery, the necessary analysis could be preformed to decide if rolling the vehicle back to the vehicle assembly building was necessary, which would have caused significant delay to the flight schedule.

The UQ model developed in Chapter 5, takes advantage of the uncertainties created for the wind tunnel derived forces and moments and considers how variations in those affect the adjusted lineloads. This is followed up by estimating the model error. While this is normally done by using test points that are not used in model generation, the computational cost of each data point makes it difficult to create extra data. Model form error is also complicated by the data fusion method, which means that we do not

expect the DFROM to match new CFD data. To overcome this, the DFROM is generated and queried many times with a different point left out each time to get an estimate of the model error. These methods are further complicated by the spatially distributed nature of the output that requires a certain amount of cohesion in points that are adjacent to each other.

This problem of uncertainty on lineloads has been explored previously for SLS; however, the focus was on the ascent problem where compressibility effects dominate. Which means the focus is on capturing the location and magnitudes of shocks. Also the previously developed method does not take into consideration the nominal value, which means regardless of location in the parameter space, a one unit change will look the same. To help address this, the new method develops uncertainty as a function of ϕ_P. The leave-one-out technique is then expanded to a full bootstrap in order to get a confidence interval on the coefficients themselves allowing for a more integrated uncertainty model that is not increment based.

While the developed DFROM and UQ methods have shortcomings such as the linear assumptions in the presence of multimodal phenomenon, it has saved millions of processor-hours. Possible expansion of the developed method to address the shortcomings are proposed and discussed briefly in Chapter 6. Also highlighted are other possible aerodynamic designs with significant amounts of already developed data, both computationally and experimentally.

CHAPTER

2

INTRODUCTION

In aerodynamics, computational and experimental data sources are two of the most common types used in analysis, and they come with a range of benefits and challenges. Experimental methods, especially wind tunnel testing as used in this work, are able to provide accurate integrated forces and moments over a large and dense parameter space. Typically, parameter space is defined as a range of Mach numbers, angles of attack, and slide slip angle or roll angle. The dense parameter space is available since each condition with a set of parameters can be easily generated once the model is in the tunnel. Computational work, on the other hand, provides spatially rich data over a sparser parameter space because most of the computational costs are spent in actually running–as opposed to preparing– the CFD solvers. As a result, computational work is ideal for generating data necessary for structural analysis of vehicles at peak conditions. As both experimental methods and computational methods improve, there has been substantial work to merge these two major sources of data.

During the liftoff and transition (LOT) regime, which covers the vehicle sitting on the pad along with the first few seconds of flight, merging experimental and computational data is necessary to obtain the structural loads. In this regime, there are two databases with slightly different use cases: 1) the wind tunnel derived force and moment database, and 2) the lineload database. This dissertation is focused on improving this second database. Lineloads are the spatially distributed forces across the axial direction of the vehicle that are primarily used in structural analysis and have secondary uses for guidance, navigation, and control [Pan11]. The first and most commonly used tool for calculating lineloads is TRILOAD developed by Shishir A. Pandya and William M. Chan, which is distributed under the Chimera Grid Tools software package. Previously, during the NASA Constellation Program for the Ares I-X, the lineload database was provided through CFD data alone; however, the mismatch between CFD integrated forces and wind tunnel derived force and moment databases caused concern. As a result, researchers at NASA

Langley led by Jeremy Pinier and Amber Favaregh began exploring new methods to reduce or eliminate the mismatch between these two different databases. The significant computational cost of LOT CFD simulations encouraged the exploration of ROMs. The inadequacies of CFD capabilities and the resulting ROM, along with the previously mentioned mismatch between databases led to the exploration of data fusion techniques.

The dissertation here focuses on the new NASA Space Launch System (SLS) program and improving the LOT lineload database through the use of data fusion. The data fusion technique merges CFD data obtained through Kestrel [Mor09], a CFD code developed by the Department of Defense (DoD), and wind tunnel data collected at the NASA Langley Research Center 14- by 22-Foot Subsonic Tunnel. The research goal presented in this dissertation is to develop a robust model to combine parametrically sparse but spatially distributed CFD data with parametrically dense wind tunnel forces and moments to create a lineload database for the LOT domain that is both parametrically and spatially dense.

Combining similar sets of data was initially explored by Carlson et al. in Ref. [Car15]. Further work on adjusting lineloads to desired force and moment data has also been explored by Dalle et al. in Ref. [Dal18]; however, their adjustment method only considers the lineloads and does not look at the entire surface data. In addition, a paper discussing similar methods developed independently in Germany [Mif19], where data fusion is used in order to adjust CFD results to better match integrated forces and moments. The current work expands on the previous works by focusing on adjusting the predicted surface stresses in order to make variations in the lineloads. It also expands into new territory by focusing on quantifying uncertainty in the developed model.

A short overview of the SLS program and the data sources that will be used is given next. While the techniques developed are independent of how the data are gathered, the availability of data constrained some of the choices made in the research. This is followed by a quick overview of ROMs and their use in fluid dynamics. The final section of this chapter looks over data fusion and how other groups have explored combining multiple data sources in the field of aerodynamics.

2.1 SLS, Lineloads, and Source Data

The NASA Space Launch System (SLS) is a proposed series of launch vehicles being designed for deep space exploration. For an in-depth introduction and update on SLS analysis, see the work by Pinier, Herron, and Gomez [Pin19]; however, a brief overview is presented here to provide context for this work. The SLS has several different configurations, which allow for the flexibility to accomplish a variety of missions, yet all of them are comprised of a central core or centerbody and two solid rocket boosters (SRBs), port and starboard. The differences between the configurations are the choice in upper stages along with whether the vehicle is configured to carry the Orion spacecraft or cargo in a fairing. The variety in the SLS configurations can be seen in Fig. 2.1. The first SLS rocket launched for Artemis I is an example of the Block 1 Crew, which has the Launch Vehicle Stage Adapter (LVSA) that connects the core stage to the Orion space craft. There is also a designed Block 1 Cargo configuration, which has the LVSA connected to a cargo fairing allowing for transport of payload to deep space. A more

powerful version of the SLS that uses the Exploration Upper Stage (EUS) as the upper stage is known as the Block 1B. Like the Block 1, it is divided into crew and cargo configurations. The Block 1B Crew configuration is planned to be used in the Artemis IV and later flights to return astronauts to the lunar surface using the Orion spacecraft. The Block 1B Cargo configuration is again similar to the crew configuration except for the replacement of Orion with a cargo fairing. Eventually, the SRBs will be redesigned, which represents the significant update from the Block 1B series to the Block 2 series of configurations. Again they are split into crew and cargo configurations. These are some of the largest rockets ever designed with heights varying from 310 feet for the Block 1 Crew to a planned 365 feet for the Block 2 Crew. The SRBs are approximately 12 feet in diameter and the core stage is about 27.6 feet in diameter, which gives a maximum cross sectional width of about 55 feet.

For aerodynamic purposes, the SLS trajectories are broken into several sections, which can be seen in Fig. 2.2. The initial one combines prelaunch, liftoff, and transition, which characterizes aerodynamic forces on the vehicle while sitting on the launch pad and shortly after liftoff. This captures a wide range of wind angles and makes heavy use of incompressible flow assumptions for characterization. This regime is dominated by large separated flows in the wake of the vehicle in the presence of ground winds. The next section is the ascent phase, which covers the vehicle during most of its acceleration from Mach 0.3 to approximately Mach 5 when the SRBs are spent and booster separation occurs. This regime is heavily influenced by compressibility effects as shocks form around protuberances until the bow shock grows large enough to encase the entire vehicle. After the SRBs spend their fuel, they must be ejected

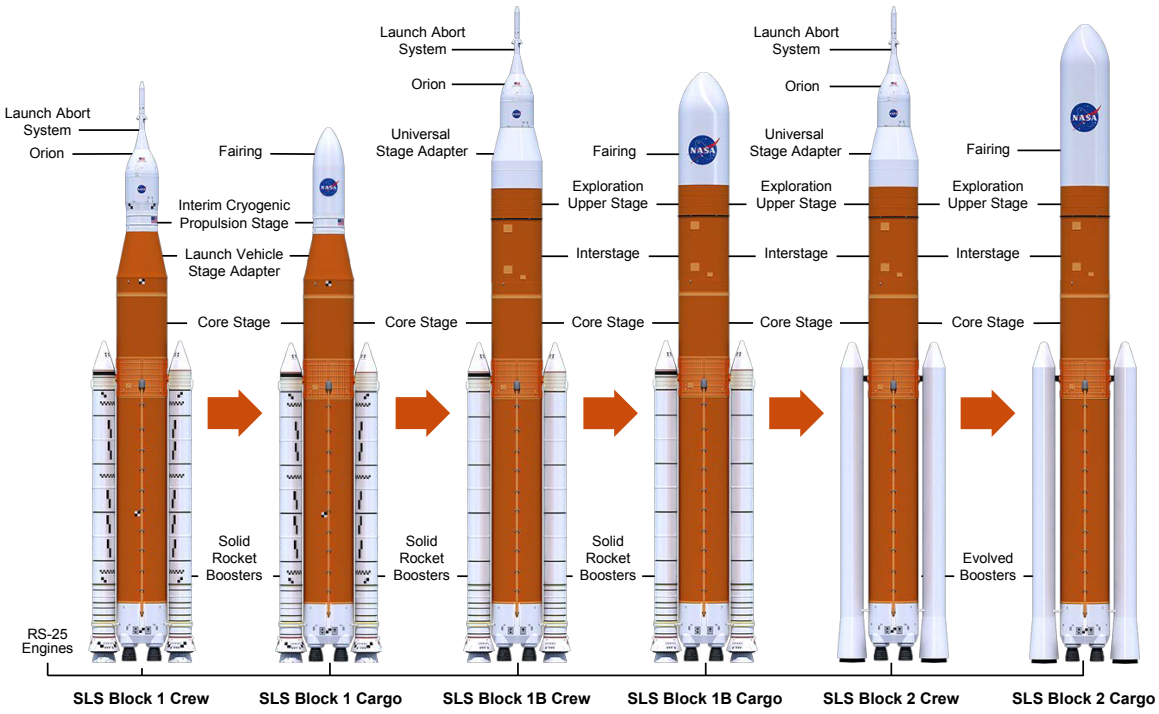


Figure 2.1 The various configurations of the Space Launch System [Nasb].

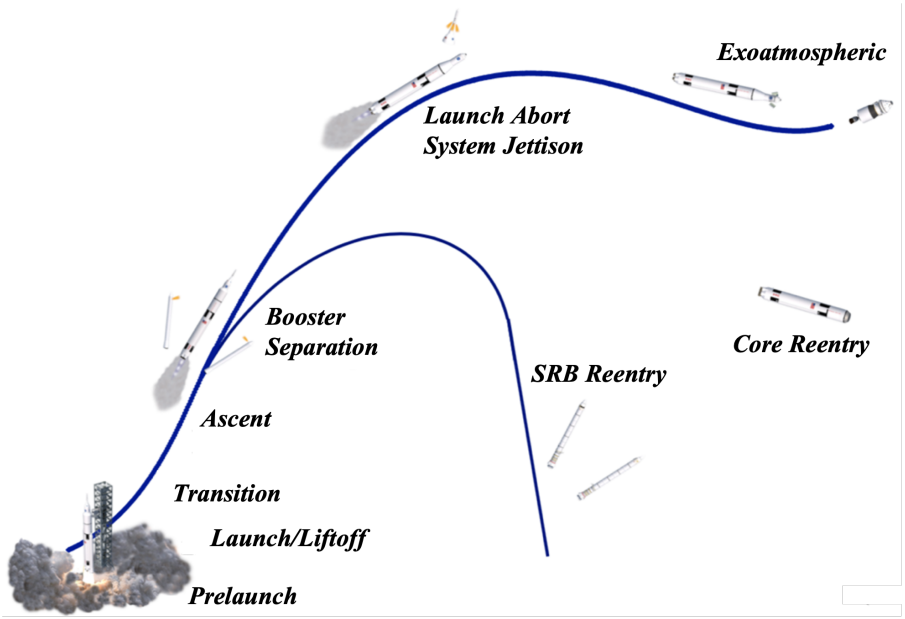


Figure 2.2 SLS mission profile with prelaunch, liftoff, ascent, SRB and Launch Abort System jettison, and reentry from [Ble14].

to help reduce vehicle weight. This is booster separation and is the most difficult portion of the flight to characterize. The initial separation event is characterized by the booster separation motors (BSMs) firing, which completely change the flow depending on how the BSM plumes impinge on the centerbody. As the boosters move away from the centerbody, the bow shocks from the three components start to interact, which can then be further complicated if either of the SRB enters the main engine plume. The next two regimes are the Launch abort system jettison and exoatmospheric flight. Both of these take place high in the atmosphere and so aerodynamic effects are minimal, which allow for simplified methods to be used. Since the SRBs and the core are not reused for SLS, the reentry for the jettisoned stages do not need to accurately predict heating and other thermal effects. Meaning these problems can be simplified significantly through inviscid assumptions.

The present work focuses on the development of the transition portion of the LOT lineload databases of the Block 1B configurations, both crew and cargo. The liftoff portion of the database covers the section of flight as the vehicle clears the tower, which is not covered in the dissertation. Tower clearance to Mach 0.3 is covered by the transition portion. The primary use for this database is for analyzing structural loads on the vehicle, and so accurate force integration is important. For reference, the transition domain covers all total angle of attack, α_P, and roll, ϕ_P, combinations in flow below Mach 0.3, a large domain that is mostly incompressible. These ranges of parameters represent the flight regime shortly after liftoff where crosswinds are a significant driver of fluid dynamics and thus aerodynamic forces and moments on the vehicle. The high angles of attack can be difficult to work with computationally because of the large wake flow that needs to be resolved. Because of the size of the launch vehicle, the size of the wake flow strains the accuracy and limits of available computational models, but they remain the only tool

currently available to get the spatially distributed data that are needed to predict lineloads. Because of the size of the vehicle and fidelity necessary, the computational costs are significant. One run takes on the order of 5 days using 1200 processors on the NASA Pleiades super computer, which is close to 150,000 CPU-hours.

The standard SLS axis system is presented in Fig. 2.3 for reference. Since the roll varies across all 360° and the total angle of attack goes from 0° to 90°, the aerodynamic analysis for the LOT domain is done in a variation of the missile axis to smooth out trends across the large changes in flow direction. Missile axis conventions also help the analysis by separating out the effects of asymmetric vortices from the normal force as roll angle changes. These asymmetries are often very localized to portions of the vehicle, which add significant difficulty to creating a model to predict lineloads.

2.1.1 Lineloads

Lineloads (also called sectional loads) databases are different from conventional force and moment aerodynamic databases because of their spatial distribution across the vehicle. This concept is not unique to this dissertation or SLS, but because of the importance a brief overview is provided. The use of computational tools to get these distributed forces dates back to the Ares program where Pandya et al. introduced TRILOAD as part of the Chimera Grid Tools [Pan11]. Lineloads are defined by integrating a stress tensor, $\vec{\vec{\sigma}}$, defined on the surface of interest in two dimensions (y and z usually) leaving a function in one

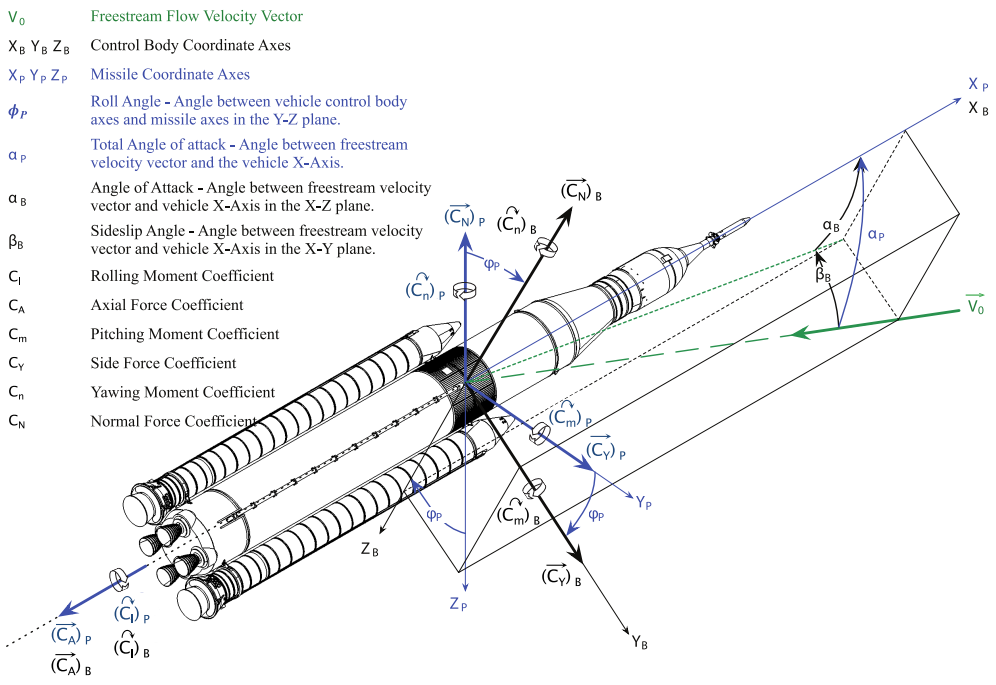


Figure 2.3 The standard reference frames used for the SLS program [Spe].

axis.

$$\overrightarrow{ll}_x \equiv \oint_S \vec{\vec{\sigma}}(x, y, z) \cdot \hat{n}\, dy\, dz\ , \tag{2.1}$$

where $\hat{n}$ is the unit vector normal to the surface. In practice, this continuous definition is discretized into a number of small sections sometimes called bins. An example of how the SLS is divided into these bins is shown in Fig. 2.4, which was graciously provided by Derek Dalle of NASA Ames Research Center. For this work, 1,000 slices are used on the centerbody and 500 are used on the SRBs. Within each bin, i, the forces distributed along the vehicle in the direction of interest, in this case the z direction, f_z, along the surface, s, are summed up between the two x boundaries, x_i and x_{i+1}, which gives the force in the z direction for the i-th slice, $\Delta F_{z,i}$, which in equation form is

$$\Delta F_{z,i} = \int_{x_i}^{x_{i+1}} \int_s f_z(x, s) ds dx. \tag{2.2}$$

This definition is however sensitive to the bin size, which makes comparing between configurations and vehicle components difficult. To address this deficiency, the resulting value is normalized by bin width when used.

$$\Delta \frac{F_{z,i}}{\Delta x} = \frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} \int_s f_z(x, s) ds dx \tag{2.3}$$

where Δx is the bin width. From here, a typical nondimensionalization of forces using the problem's reference characteristics is used[1],

$$\frac{\Delta C_{z,i}}{\Delta \hat{x}} = \frac{\Delta F_{z,i}}{\frac{1}{2}\rho_\infty V_\infty^2 S_{ref}} l_{ref} \tag{2.4}$$

where S_{ref} and l_{ref} are geometry specific reference area and reference length, respectively. Flow properties are captured by ρ_∞ and V_∞, which are the freestream density and velocity respectively. The addition of l_{ref} is done to normalize the Δx term. It is this coefficient form of the bin width normalized value that is plotted and analyzed throughout the work. For a given configuration and component, the bin width is constant, which means that the integrated load is simply

$$C_z = \sum_{i=1}^{n_{bins}} \Delta C_{z,i} \Delta \hat{x} \tag{2.5}$$

where $\Delta \hat{x}$ is the bin width normalized by l_{ref}. Subsets of this summation can also be used to get the loading across a subsection of the vehicle such as the Launch Vehicle Stage Adapter when needed by the database user. This need for spatial distribution means that CFD is the best tool available despite being such a difficult regime for CFD to model.

As part of the work, a new lineload solver was written in Python, which allowed for separate lineloads for the pressure and friction forces to be calculated. The framework and methodology of calculating lineloads from triangulated surface data are laid out in Ref. [Pan11], which lead to the creation of the lineload tool 'triload' in Chimera Grid Tools (CGT) [Cha18]. While conceptually straightforward, this

[1] Due to export control restriction on the data used in this work, a second normalization is done based on maximum observed values. See 2.1.4 for further details

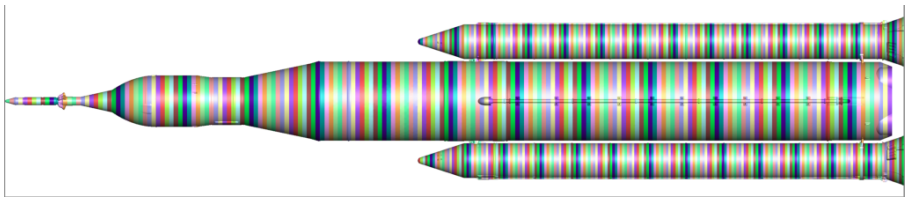


Figure 2.4 Sectional slices of the SLS Block 1B Crew vehicle used for calculating lineloads. Figure from [Dal18] provided by the author.

code requires handling a significant number of details that complicate the process. The most significant complication is cells that fall into multiple bins and must be divided into subelements and accounted for properly. As currently organized, triload requires the full CGT suite to be compiled and does not give separate outputs for pressure and friction components. This resulted in a complex workflow that required generating copies of the surface solution, which only had the pressure or friction components. Since the new solver is written in Python, it requires no installation and is stand alone along with the ability to output pressure and friction lineloads separately, which significantly streamlined the development and evaluation of the ROMs used in this work. Since initial development of this new lineload tool, it has been refined to become more robust, more computationally efficient and able to handle more edge cases [Lee22]. While line moments can also be calculated, they are not utilized in SLS databases.

The separation between pressure and frictional forces became necessary as the data fusion methodology developed in Chapter 4 adjusts these contributions separately. And while the math is all developed for the complete surface, the independence between spatially distributed information and parametrically distributed information allows the convenience to focus on the quantity of interest, which for this work is the lineloads.

A typical lineload generated from CFD data on the Block 1B Crew vehicle is presented in Fig. 2.5. This is from a solution at $\alpha_P = 70°$, $\phi_P = 0°$. The top section is the axial force lineload followed by the side force lineload below that. The third line is the normal force and at the very bottom is a simplified outline of the vehicle to help with orientation. The black lines are the lineloads for the centerbody, the red are the lineloads for the port SRB, and the blue lines are the lineloads for the starboard SRB. The side force shows a representative case where the asymmetric vortices (explained in the next section) causes a significant localized side force on the same order of magnitude as the the normal force. This flow phenomenon is common to slim bodies at midrange angles of attack and a significant source of difficulty in modeling this regime.

2.1.2 Asymmetric Vortices

The asymmetric vortices that appear at higher angles of attack are another reason the analysis stays in the missile axis system. These vortices are common to all slender bodies in midrange angle-of-attack flows and many resources have been devoted to characterizing and modeling them over the years [Fid75; Har90]. Because the domain of interest goes from $\alpha_P = 0$ to 90°, midrange here is considered from about $\alpha_P = 30$ to 70°. As mentioned briefly in the AIAA paper [Wigaa], the presence of asymmetric vortices

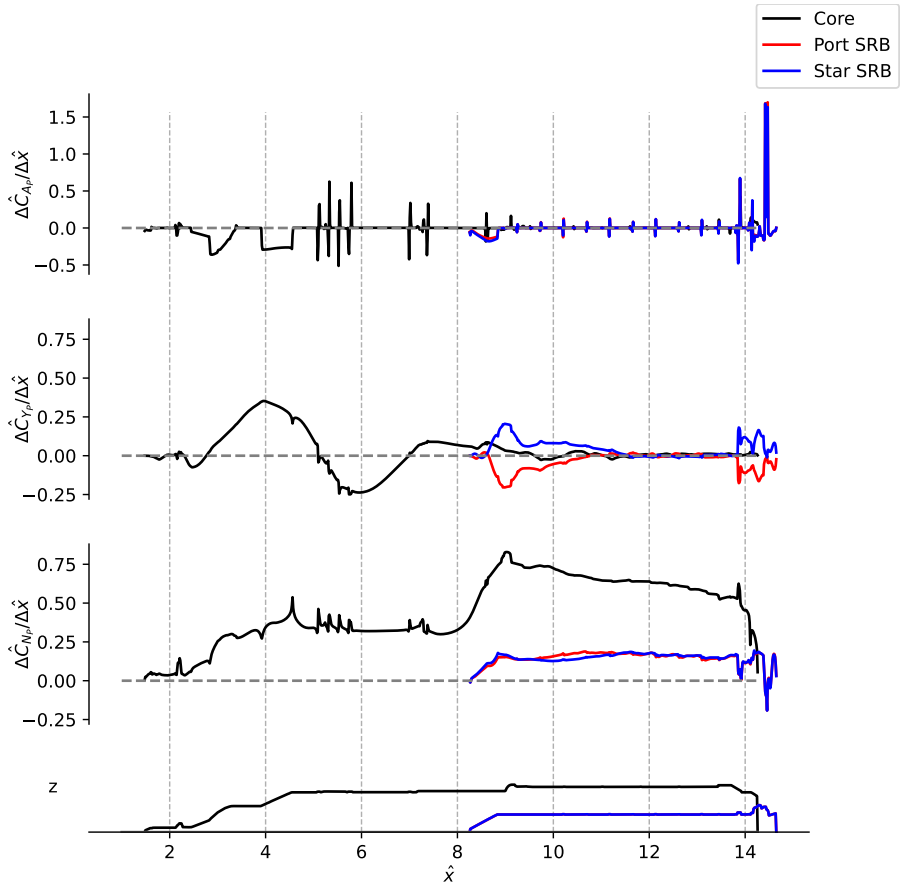


Figure 2.5 Representative set of lineloads for the Block 1B Crew vehicle at $\alpha_P = 70°$, $\phi_P = 0°$.

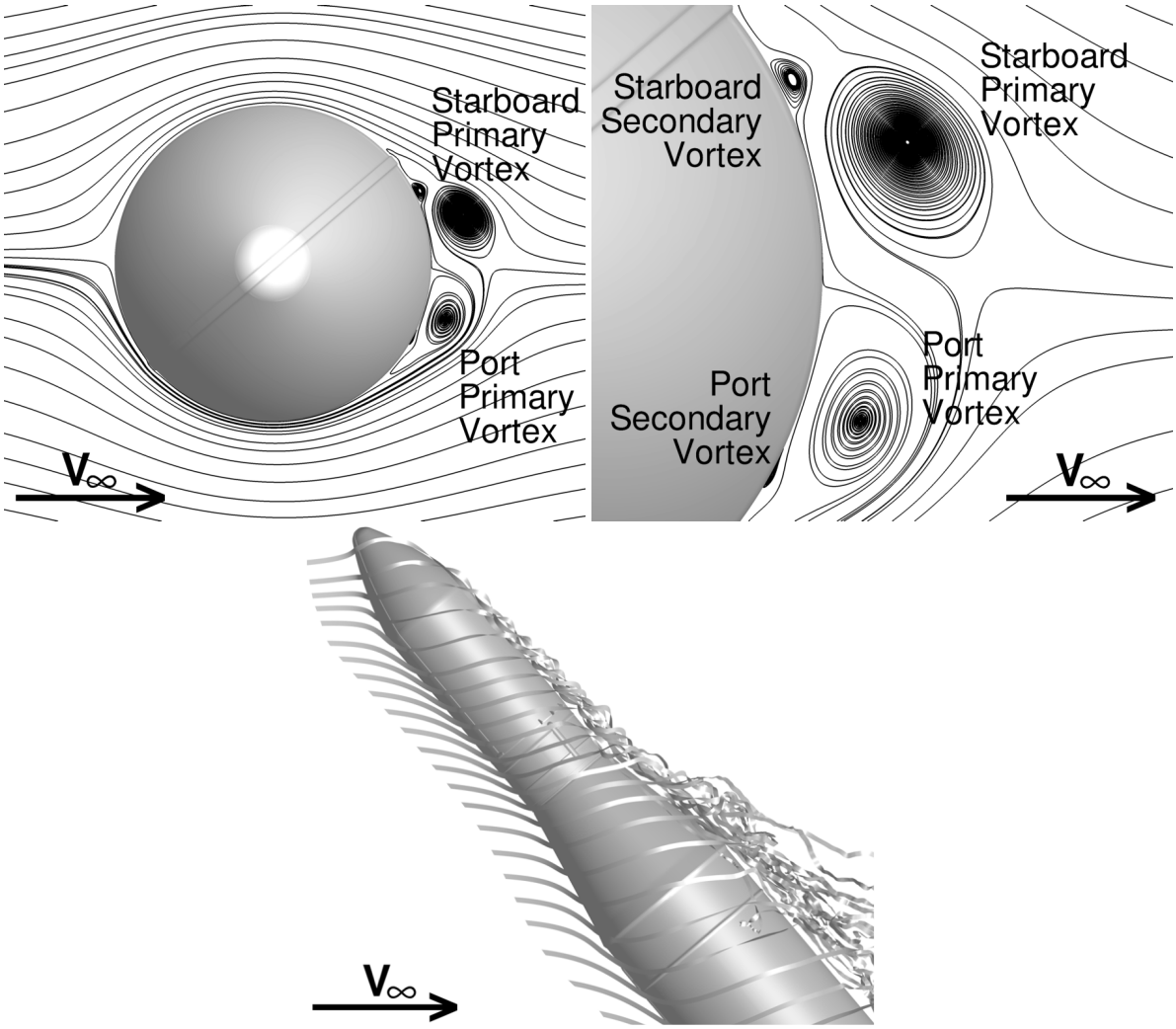


Figure 2.6 Separated flowfield showing primary and secondary asymmetric vortices in top left) wide, top right) zoomed, and bottom) profile views. Figures from [Pom] provided by the author.

caused significant complications due to the breakdown of linearity assumptions. Small perturbations can cause the vortices to set up on either side of the slender body, which makes prediction difficult.

Figure 2.6, which Brent Pomeroy developed, shows the asymmetric vortices forming on the Block 1 Cargo version of the SLS vehicle at $\alpha_P = 50°$, $\phi_P = 0°$. As can be clearly seen, there are two primary vortices with one on each side while one is significantly larger than the other. As can be seen in bottom subfigure, this causes the flow to stay close to the vehicle and move aft.

Under slightly different conditions both CFD modeling and wind tunnel testing, can see evidence of the asymmetric vortices forming on one side of the vehicle or the other. These vortices make drawing trends and predictions for side force and yawing moment difficult. Wind tunnel data often show a side force in the positive direction during one test while showing a negative side force the next test even when both model and tunnel used remains the same. While the vortex formation in the wind tunnel is dependent mostly on the actual model as built, computationally the vortex formation is sensitive to any

number of effects. In CFD, even something as simple as a slight variation in time step can result in a solution with the vortex forming on the opposite side. As a result, wind tunnel data are relatively smooth as a function of α_P while the corresponding CFD data can jump back and forth across 0 side force. A missile axis analysis helps constrain the nonlinearity and bimodality effects to the side force and yawing moment.

While computational work continues to be done to improve prediction and characterization of the asymmetric vortices, the focus of this dissertation is elsewhere. This brief overview is necessary because many of the difficulties encountered are related to this phenomenon.

2.1.3 Data Sources

2.1.3.1 CFD Data

As outlined in more depth in the work by Ratnayake, Krist, and Ghaffari [Rataa], the LOT domain is a difficult flow condition to model using common finite volume flow solvers. Recent work by Pomeroy, [Pom; Pom23], has gone into further characterization of the sensitivity of solutions to solver parameters and is recommended for an in-depth look of settings used in these simulations. Because of the size of the SLS vehicles, high angle of attack with large regions of separated flows, and each point being a major commitment of computational resources, significant effort was put into getting the most out of each CFD run. A brief overview of the flow solver and its capabilities is provided followed by an overview of how the computational data for this work was collected.

The CFD data were generated by other members of the NASA Langley SLS CFD team using the Kestrel flow solver. The NASA Langley CFD team has changed members over the course of the work, but I would like to highlight the work of Steve Krist for geometry preparation as well as grid generation. Other members Nalin Ratnayake, Karen Deere, and Farhad Ghaffari ran most of the simulations and helped analyze the results. While I helped with analyzing results and post processing, during the time of data generations I was developing the data fusion methods shown in this work.

The Kestrel flow solver is maintained under the Computational Research and Engineering Acquisition Tools and Environment, Air Vehicle (CREATE-AV) project which is part of the Department of Defense High Performance Computing Modernization Program (DoD HPCMP) [Mor09]. Kestrel uses a dual mesh approach that allows for an unstructured near-body grid and a Cartesian off-body grid. The flow solver within Kestrel that is employed on the near-body domain is KCFD, a cell-centered, finite volume method for solving the Euler and Reynolds-averaged Navier-Stokes (RANS) equations on mixed element unstructured meshes. Available turbulence models include a number of the Spalart-Allmaras (SA) variants and the Menter two-equation model with and without Shear Stress Transport. In addition to the traditional RANS capability, KCFD provides an option for delayed detached eddy simulation (DDES) for use in situations involving massively separated flow allowing for large eddy simulation (LES) turbulence modeling away from solid surfaces. The CFD code employed on the off-body domain is SAMAir, a finite difference method for solving the Euler and RANS equations on Cartesian meshes that is derived from an earlier code, ARC3DC, developed at the NASA Ames Research Center. SAMAir

is capable of implementing all the models available in KCFD, allowing for a uniform implementation of the governing equations in both solvers. Within the off-body grid, there is the option for adaptive mesh refinement.

The dual mesh approach provides a significant advantage in simulating flows that contain regions of massively separated flow. In these cases, the extensive wake on the leeward side of the body can be resolved in the off-body Cartesian mesh, for which operations are significantly cheaper to implement than operations on the near-body unstructured grid. Moreover, the Cartesian mesh is much more suitable for the use of higher-order spatial discretization and is amenable to the implementation of solution Adaptive Mesh Refinement. Both of these capabilities aid in obtaining the greatest possible accuracy while running on the smallest possible grid.

The typical form of the Navier-Stokes Equations are used as below.

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho u_i) = 0 \tag{2.6}$$

$$\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(\rho u_i u_j) = -\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} \tag{2.7}$$

$$\frac{\partial}{\partial t}\left[\rho\left(e + \frac{1}{2}u_i u_i\right)\right] + \frac{\partial}{\partial x_j}\left[\rho u_j\left(e + \frac{1}{2}u_i u_i\right)\right] = -\frac{\partial p u_j}{\partial x_j} + \frac{\partial}{\partial x_j}(u_i \tau_{ij}) - \frac{\partial q_j}{\partial x_j} \tag{2.8}$$

Typical notation where ρ, u_i, e, and p are density, velocity in the x_i direction, internal energy, and the pressure. The viscous stresses, τ_{ij} and the heat flux, q_j, also have standard forms. The equations are closed using the ideal gas law as the equation of state and as is typical in aerodynamics, gravity and other body forces are ignored. In practice, these are Favre averaged and closure is provided through an accommodating turbulence model. While Kestrel provides a variety of turbulence models, only one is used for this database.

For the results in this dissertation, the flow was assumed fully turbulent and used SA with DDES, a hybrid RANS-LES turbulence model, which allows for unsteady RANS modeling near the vehicle and LES results in the wake region. The Cartesian grid was adapted every 250 time steps based on the magnitude of vorticity within a cell. The threshold value remained constant throughout the work. Follow-up work showed little sensitivity to the vorticity value picked as explored in Ref. [Pom; Pom23]. The follow-up work also included exploration of other refinement variables such as Q criterion; however, since the desire was to capture the wake, as well as any vortices that formed, the magnitude of vorticity proved to be the most useful. Simulations were run time accurate for several thousand time steps before time averaging the results over two seconds of full scale. These time-averaged results are the CFD data used for this work. Freestream conditions were set to match the Reynolds number used in the wind tunnel and atmospheric conditions at sea level. Table 2.1, shows the typically used input parameters for the CFD simulations in this work. Credit must be given to Scott Morton and David McDaniels of CREATE-AV to providing the baseline setup for this difficult problem with further tweaking and debugging by Steve Krist.

The unstructured grids used were approximately 120 million cells with prisms in the boundary layer and tetrahedrons further out. With the typical grid refinement every 250 time steps, the Cartesian grid

Table 2.1 Table of values for a typical CFD run used in this work.

Parameter	Value
Kestrel Version	7.1.2
Reynolds Number (centerbody diameter)	600,000
Number of Time Steps	14,000
Time Step	0.0005 sec.
Nondimensional Time Step	0.00369
Time Averaging start	5 sec. (10,000 time steps)
Time Averaging Length	2 sec. (4,000 time steps)
Turbulence Model	SA DDES
Inviscid Fluxes Scheme	HLLE++
Spatial Accuracy	Second Order
Temporal Accuracy	Second Order Gauss-Seidel
Typical Unstructured Grid Size	120 Million Cells
Typical Cartesian Grid Size (After Refinement)	1 to 3 Billion Points
CPU-Hours	100,000

grew to a size between 1 to 3 billion cells depending on α_P. The NASA Advanced Supercomputing resources were used to complete the simulations [Nasa]. Typical solutions used on average 850 processors over approximately 5 days giving an approximate cost per run of 100,000 CPU-hours. While significant on its own, the necessity of generating a database over a large parameter space led to some compromises in simulation fidelity to reduce computational costs. The limitations are evident when $\alpha_P > 50°$, where the CFD results had difficulty matching the experimental data as will be shown in Fig. 3.9 and Fig. 3.10. The size and cost of these simulations prompted the initial investigation into ROMs to help manage the cost and data necessary to create the necessary databases for the SLS program.

As mentioned development of the methodology in this work was done in parallel to the data collection, so while I did not run any simulations I am grateful for my teammates Steve Krist, Nalin Ratnayake, Karen Deere, and Farhad Ghaffari for generating this CFD data. On newer SLS configurations, I have since run comparable simulations. Despite not generating the data, I was responsible for diving deeper into the results as they were generated. This included comparing new CFD data to previously collected data, especially the wind tunnel data (which is described in the next section) as well as other checks on the quality of solutions such as convergence criteria.

2.1.3.2 Wind Tunnel Data

To help guide the CFD, wind tunnel data from tests in the NASA Langley 14- by 22-Foot Subsonic Tunnel are examined and used [Cha19]. This tunnel is an atmospheric, closed circuit, low-speed wind tunnel that has an effective Mach range of 0 to 0.3. The goals of the test were focused on gathering integrated forces on the SLS vehicle during the liftoff and transition regime. Additional data were collected with pressure ports distributed on the surface of the vehicle; however, this was not a primary goal of the testing. While not used in this work, the wind tunnel testing also examined tower effects as seen in Fig. 2.7, which

shows the Block 1B Crew model next to the simplified mobile launcher that was tested. The data used in this dissertation were gathered using a different sting and attachmenet hardware than shown allowing for α_P range from 90° down to 0°.

The testing takes advantage of the common components of the SLS and allows simplified switching of the configuration by only having unique nose pieces. The Block 1 Crew, Block 1B Crew and Block 1B cargo configurations were all tested at a 1.75%-scale. Based on earlier testing, results above a dynamic pressure of 40 psf are nonlinear and so data were collected at 50 psf. This pressure corresponds to a Mach number of about 0.18 and a Reynolds number of 1.27 million per foot. The control system in the tunnel works by maintaining a constant dynamic pressure and so there were slight drifts in Mach and Reynolds number as atmospheric and tunnel conditions changed over the course of the test. The low speed allows for incompressible assumptions and so the drifting in Mach is not a concern. The Reynolds number variations are not a concern because the model is fully gritted forcing transition everywhere on the model at all model attitudes

The integrated forces and moment measurements from the balances were the data that gets processed into the force and moment database. Typical to aerodynamic testing, a single balance is used to get integrated forces and moments for the entire configuration. Unlike previous tests, this test included a set of load cells in the attach hardware of the SRBs. Due to the addition of these load cells, force and moment data are available for all three components of the vehicle and not just the total vehicle unlike previous testing. The additional force and moment information came at a cost of the removal of pressure port information on the SRBs. This change in available data is one of the contributions to the rework in methodology from previous SLS LOT lineload databases. The primary use of the collected forces and moment data in this work is the data fusion; however, they were also used in CFD point selection, as well as a quality check on the CFD and ROM outputs. The wind tunnel test also collected pressure port readings, which were used as a secondary check on the CFD and ROM data; however, because of the addition of the load cells, there are only pressure port data for the core on the configurations of interest.

The force and moment database generated from the wind tunnel testing is considered of higher fidelity than CFD because of the historical shortfalls in CFD when trying to capture flows with large wake regions. It also has the advantage of having well defined uncertainty quantification on these measured values, which are used in the uncertainty quantification of the resulting model.

These data were collected in 2017 by a team from NASA Langley headed by David Chan over the course of a few weeks. While Chan led the test effort as well as the post-test analysis, the test would not have been a success without the support of John Paulson and Patrick Shea for SLS and the entire team at the tunnel lead by Les Yeh at the time. While not part of the team responsible for the test, I did spend time in the test control room examining data as they came in as well as helping to take photographs of smoke flow in the tunnel. I was also a part of reviewing the data quality and the formation of the force and moment database that is used in this work as the source of high fidelity data in the data fusion step.

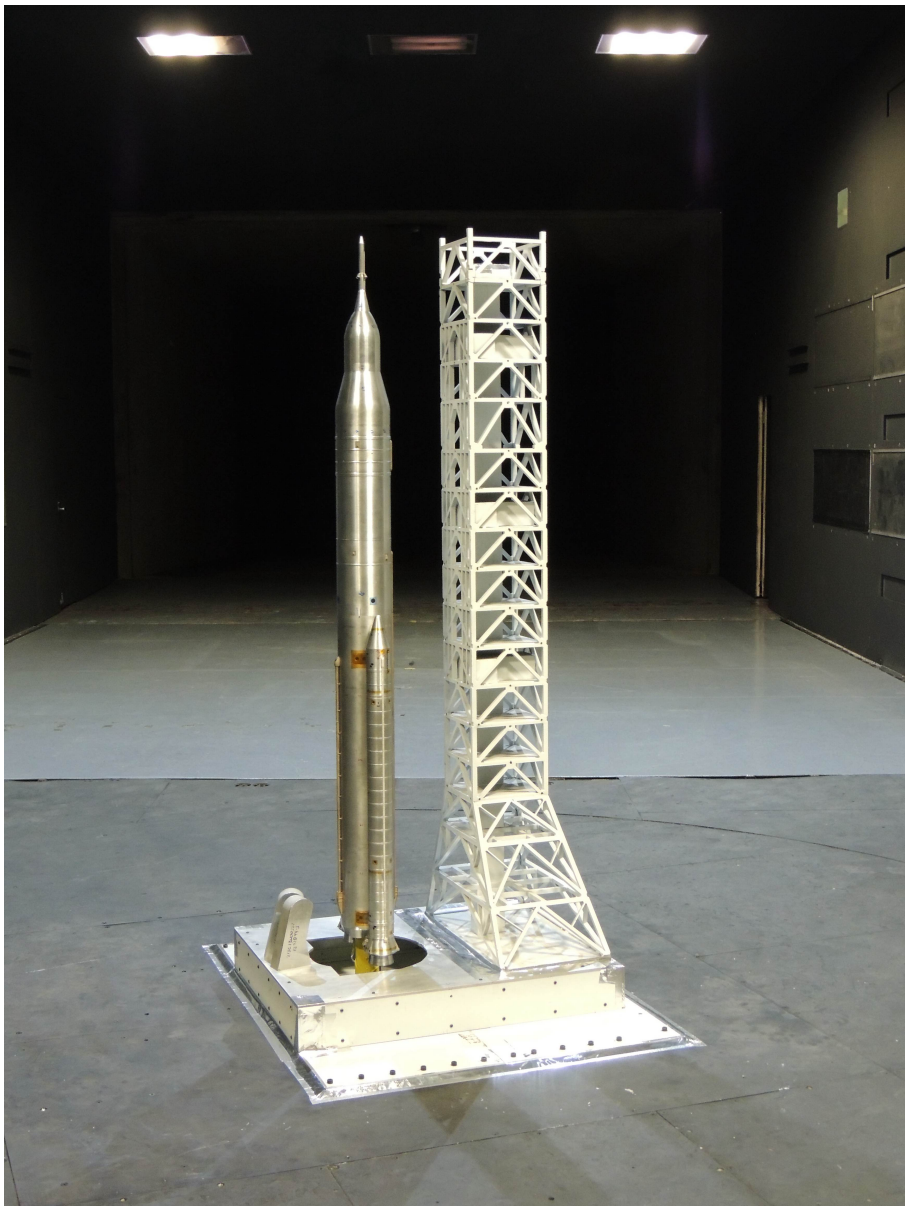


Figure 2.7 Model of the Block 1B Crew vehicle in the NASA Langley Research Center 14- by 22-Foot Subsonic Tunnel [Photo: NASA].

2.1.4 Data Normalization and Sanitization

Most of the data produced by the SLS program and all that are used here fall under International Traffic in Arms Regulations (ITAR), which are regulations that apply to defense and military related technologies. To allow publication and wide availability of this document, data normalization and sanitization is necessary. This section includes a brief overview of the normalization used throughout this document and an argument on how this sanitization of values does not diminish the impact or arguments within this work. It is also these regulations that prevent publishing the details of some of the parameters used to generate the source data. Normalization does not impact the outcome of the work and does not have a significant effect on the explanation of plots used. Sanitization here refers to the practice of removing the scales from some plots. For certain variables, such as coefficient of pressure, normalization is impractical, so when plotted, the scales need to be removed to conform to export rules. This methodology could be developed and explained with manufactured data; however, the choice of actual data provides a realistic implementation of the method as well as highlighting the limitations that future work will need to overcome.

2.1.4.1 Lineload Normalization

Since prediction of lineloads is the primary focus of this work, every effort is made to make comparisons and analysis of data as easy to understand as possible. Lineloads are already a nondimensional value; however, because of the aforementioned ITAR regulations, a further normalization is used to allow the usage of numerical scales making it easier to compare figures and data across the work. The normalization used is based on the peak lineload value at a single x station across all the SLS geometries analyzed including those not presented in this research. This normalization means that all lineloads presented have values that fall somewhere between -1 and 1, which allows for ease of comparison across the wide range of input parameters seen in the LOT regime. In essence, what this means is we slightly redefine Eq. 2.3 to become

$$\frac{\Delta\hat{C}_{N,i}}{\Delta\hat{x}} = \frac{\Delta C_{N,i}}{x_{i+1}-x_i}\frac{l_{ref}}{C_{N,ref}} = \frac{1}{x_{i+1}-x_i}\frac{l_{ref}}{C_{N,ref}}\int_{x_i}^{x_{i+1}}\int C_N(x,s)dsdx \tag{2.9}$$

2.1.4.2 Sanitized Scales

The pressure coefficient is the major sanitized data type. While there is no practical normalization to allow for using of numeric scales, for most of this work what is important are the trends and patterns in the pressure data which are still visible in the plots and in particular the contour plots that are used more for qualitative analysis as opposed to quantitative analysis. Also when quantitative analysis is important to compare two sets of pressure data, numerical differences are provided, which help highlight the necessary patterns found in the sanitized plots.

The other sanitized data are integrated forces and moments. These are primarily used to show the effectiveness of the data fusion methods, so relative change is the important information conveyed by these plots, which is still clearly seen. Since the data fusion method forces an exact match, these plots are mostly used to show that the new lineloads integrate to the desired forces and moments. Even with

these limitations on the presented results, the benefits and drawbacks of the various methods used in the work are clearly displayed.

This quick overview of data sources is presented for background information; however, the methods presented and developed in this work are independent of the data sources and can be easily applied to a wide variety of configurations, flow solvers, and experimental data. Part of the reason these techniques are source data agnostic is through the use of reduced ordered modeling, which can easily accommodate data from multiple sources as long as it has a similar format.

2.2 Previous Work

Many of the techniques and methods used in the SLS program have their start in the Ares program. While the space shuttle was developed before modern computational fluid dynamic tools, during the Ares program, the increase in computational tools and reduction in wind tunnel testing methods lead to work on combining these data for a launch vehicle. One such attempt at combining these datasets was the Ares lineload databases. These methods were fairly simple focusing on a two equation linear system to adjust a lineload. One equation was responsible for scaling to the right magnitude to match integrated force and the other tilting it to generate the correct moment. This method often predicted unbelievable results such as changing the sign of the lineload at the ends of the vehicle.

Uncertainty quantification in SLS aerodynamics can also trace their roots to the work done on the Ares program [Hou; Hem]. These approaches focused on a finding and describing the contributions from different effects, which are then added together to generate a total uncertainty. Some of these contributions were generated using CFD data while the nominal was rooted in experimental data.

Recent work, [Dal18; Car15], has focused on taking a step through a linear transformation to find fundamental modes to apply these adjustments. The work here combines this idea with other developments on PCA with interpolation to first build a model that expands the CFD data across the domain and then adjust to desired values. Finally as a further evolution, uncertainty quantification is performed on the resulting model, which is done in two parts. The first is the development of a simple increment based method followed up by a new method focused on the uncertainty in the transformation into PCs.

2.3 Principal Component Analysis

PCA is a linear transformation of the dataset that looks to transform the data into a new variable set with desirable characteristics [Jol02]. Generally, this transform is used to look for the least correlated variables as possible, which maximizes the variance within each transformed variable, called principal components (PCs). First introduced by Pearson in 1901, it was further developed by Hotelling, which developed the technique as it is known today [Hot33; Pea01]. In fluid dynamics, PCA was introduced by Lumley in 1967 under the name Proper Orthogonal Decomposition (POD) to help tackle stochastic turbulence [Lum67; Ber93]. [2] PCA in fluid dynamics tended to be first used in analyzing flow structures,

[2]The difference in name is seen across fields. PCA comes from the statistics domain while most of the work that uses the term POD is engineering-based. In climate related fields the equivalent Empirical Orthogonal Functions (EOF) is used.

but is now commonly used to predict data at new conditions. The PCs can either be trimmed down for dimensional reduction or be used as a form of basis extraction. When used as a means to extract a basis, the extracted features can then be recombined in different ways to predict new conditions.

PCA has applications in many subfields of fluid dynamics such as aeroelastic modeling [Sil18], aero-optics [Zil16] and cavity flows [Aru00; Car04]. A recent overview was performed by Taira et al. in Ref. [Tai17]. More relevant to the research being currently done is the use of PCA as feature extraction [Ril18; Lee10]. Riley et al. in Ref. [Ril18] used it to explore separation in a simulated scramjet while Lee in Ref. [Lee10] investigated it more broadly to compare to experimental flow visualization. PCA has also been explored in fluid dynamics as a method to speed up computation of turbulent combustion [Wigb; Mal22; Sut09]. One of the driving costs of turbulent combustion is the number of equations necessary to track individual species. By using PCA, an approximation of the species mixture can be achieved, which directly reduces the number of equations in the system.

Of interest is the use of PCA to recombine extracted features to generate new predictions. This falls under two broad categories. The simpler method, which is used in this work, uses the transformed variables and does interpolation in parameter space to make new predictions. This is often called POD with interpolation, which has found extensive use especially in design optimization due to its simplicity [BT; Gan13]. More advance methods also transform the governing equations allowing for temporal progression as well as incorporating other physical constraints [Kun02; Rey20; Lee]. This is very useful for unsteady flows where a relatively small number of time steps can be used to train a model to predict long term characteristics. For this work, I will be calling both a form of reduce order models (ROM); however, recently the field has been solidifying a distinction between the two, with a focus on calling methods that transform the governing equations as a ROM and in particular when coupled with PCA a projection-based ROM. The method used here with a focus on PCA with interpolation would fall better into the category of surrogate modeling.

ROMs come in a variety of formats. In general, they take a series of high fidelity observations and construct a lower fidelity model to fit the observed data. This new model is much cheaper to evaluate while maintaining most of the predictive capabilities of the original high fidelity model. Because of the ever increasing cost of computational tools, reduced order models are useful where the domain of interest is large. Examples of large domains are commonly found in preliminary designs and optimization as well as filling in aerodynamic databases as is the case here. While limited by the accuracy of the tools used to generate a ROM, they are useful for design optimization and exploring large parameter spaces. Once the initial investment is made in creating a ROM, they can be quickly and cheaply evaluated across the parameter space of interest. The full parameter space can be explored in minutes as opposed to days for each additional data point using high fidelity CFD data. The limited accuracy has sparked exploration of augmenting the ROM predictions with other data sources.

Throughout, PCA will be used since it appeared first in the literature.

2.4 Data Fusion

Data fusion and the related data assimilation is a set of methods that combine various datasets with the goal of improving overall information [Ras17; Cas13]. There are different subcategories of data fusion depending on input type, output type and how the data sets relate. Here the focus is on what Durrant-Whyte categorizes as complimentary data [DW88]. With complimentary data fusion, the focus is on taking two sets of data that attempt to measure the same quantity and combine them for a more accurate and reliable reading, most typically seen in a sparse high fidelity dataset combined with a denser lower fidelity set.

Most of the development of data fusion has focused on data that have quantified uncertainty through calibrated instruments. When the uncertainty in the measurements are known, the mathematics behind the fusion are fairly well defined but when the uncertainty is unknown it becomes less rigorous. This rigor breaks down further when two different methods are used entirely with different levels of fidelity. The end result is still a method that balances the strengths of the different datasets. When looking at fusing wind tunnel and CFD data, there are various methods that have been developed. Some examples are Ref. [Rei06] and Ref. [Zim14]. In Ref. [Rei06], Reisenthel et al. examine the use of response surface methodologies to combine experimental and computational data. This allows for a straight forward approach to take the trend seen in computational data and apply that trend to sparse experimental data. This is in a similar vein to multifidelity Gaussian process regressors, which use a correlation between different datasets to predict new conditions [Sco]. Unfortunately these methods focus on point predictions. As is often the case, integrated forces and moments are primary drivers of design and analysis; however, here the predictions of interest have a spatial component.

Work to combine experimental and computational data has been done to reconstruct flow fields around bodies of interest based on sparse experimental measurements, which naturally include a spatial component. Flow reconstruction has used a variety of method; however, more recently neural networks and other machine learning techniques have come to prominence [Leb; Buc; Xu20; Eri20]. Of more interest in the present work, in Ref. [Zim14], Zimmermann et al. use PCAs and constrain the ROMs to aerodynamic data through a least squares approach although the primary focus is predicting CFD solutions through residual minimization and not matching experimental data. Most of these methods are focused on providing more information about flow features in a qualitative sense. On the more quantitative side, Mifsud et al. in [Mif19] attempt to match pressure measurements in order to accurately predict integrated forces. The limitations of that method that this work address is further explained in 2.6.3.

The direct precursor to this work was done by Carlson et al, also for the SLS program in the same flow regime but an earlier configuration design. Like the method in Mifsud et al, Carlson adjusts surface information based on the pressure measurements, and like here the goal was to improve lineload predictions. This method successfully uses pressure measurements in combination with PCA to better match experimental forces and moments, while still keeping spatial information in order to calculate lineloads. Unfortunately the method increased disagreement in axial force and improvements were sought out. Further limitations that are addressed in this work are elaborated in 2.6.1, but of note is its difficulty of

dealing with the reduced number of pressure measurements in newer wind tunnel tests.

Further work is being done as part of SLS and NASA at large and while most of the techniques are not used in this dissertation they are relevant to know the context within. To begin with the work of Derek Dalle and the other members of SLS CFD team at NASA Ames have been heavily influential [Dal18]. Similar to the work here, Dalle adjusts lineloads to match target forces and moments in the transonic and supersonic regime; however, this is done on the lineload level as opposed to the entire surface as done here. Other limitations that I set out to address are expanded on in 2.6.2. More recently, work at NASA has focused on multifidelity type data fusion. Because of the nature of iterative design in aerodynamic vehicles there is a significant amount of data generated at each stage of the project that is often "thrown out" as the design matures and new data is gathered. To address this, I was involved in a project to help combine not just the latest wind tunnel and computational data but data collected over the course of a vehicle's design history. This project known as Aerofusion, focused on using machine learning methods such as neural networks to do this in a data driven method [Sny23].

While these multifidelty methods prove useful in reducing the amount of new data necessary for each design cycle, they are insufficient when trying to combine wind tunnel derived integrated loads and computationally derived lineloads, which is the goal of this work. Unlike other work in the field, instead of using a collection of spatially distributed measurements to predict a surface, point measurements represented by the forces and moments will be used. Current database generation methods create a disconnect between force and moment databases and the forces that the lineloads represent. By fusing the lineload data with the force and moment database directly, it is possible to remove this disconnect. This also helps to compensate for the reduced accuracy from using ROMs instead of using CFD at every point of interest as well as inaccuracies within the CFD itself.

2.5 Uncertainty Quantification

Uncertainty quantification (UQ) is best described as the field interested in identifying, quantifying, and hopefully reducing uncertainties in predictions. While directly related to statistics by measuring and quantifying errors and noise, propagating these errors and unknowns through complex systems can be considered its own field. Broadly, uncertainty can fall into two categories, aleatoric uncertainty and epistemic uncertainty. Aleatoric uncertainty being a fundamentally random error that can be characterized but not fully eliminated. Epistemic uncertainty on the other hand represents errors in the systematic representation of the problem. These can often be reduced and sometimes eliminated through sufficient resources such as performing direct numerical simulations instead of RANS simulations, which would eliminate errors associated with turbulence modeling. However the cost of elimination will often outweigh the benefits especially when said uncertainties are possible to characterize.

UQ is important because it allows for probabilistic analysis as well as being able to focus efforts on where the most analysis is critical. When only a nominal value is provided, any subsequent analysis is pass/fail; however, with UQ the chance of failure can be analyzed and the weighting of risks can be analyzed. When the UQ is not properly characterized, worst case can be failure, however more likely

is increased costs because of an increase to safety margins. For example during the Ares I-X launch, the uncertainty was over prescribed which led to the use of fly away maneuver on launch to ensure the vehicle did not impact the tower [Dav]. While it was a successful launch, this maneuver did significant damage to the launch complex.

For SLS, the focus tends to be on trying to fully quantifying the prediction interval of the quantities of interest. Usually a normal or uniform distribution is used with the total variance being the sum of discrete contributions from various sources. This is common to many NASA projects such as the Orion project, which Bibb et al. does a thorough job of outlining the various contributions and how they are applied [Bib11a; Bib11b]. These sources can be well defined such as balance calibration errors, or more engineering judgment derived such as accounting for Reynolds number effects. It is not uncommon to provide a percentage increase to cover any remaining aleatoric uncertainty not fully characterized. During an analysis, a single value from this distribution is drawn and used throughout; however, occasionally these pure-bias type methods provide an unrealistic result. For example during the Ares program, this pure-bias method on the rolling moment stressed the control capabilities; however, a more accurate UQ model was able to show that the current designs were sufficient [Pin12].

Recent work related to SLS have been focused on improving UQ across all aerodynamic regimes and unlike previous programs, UQ has taken a leading role in database planning and development [Pin19]. Picking up from the Ares program, Favaregh et al. [Fav16] lays out the basic framework the SLS program uses for 6-DOF aerodynamic databases. The techniques make use of adding the variance of different contributions of uncertainty in this case balance, database modeling, and experimental uncertainty. In particular during data collection, sufficient repeats are gathered to be able to characterize the experimental error. Finally an additional uncertainty factor is multiplied to represent unknown unknowns. Recent work by Shea et al [She19], goes into detail about building the SLS ascent force and moment databases which covers the vehicle M = 0.5 to M = 4. Much of the work is dedicated in describing and quantifying the uncertainty quantification methodologies, which have been improved upon over the years. In particular much work has been done to improve the merging of knowledge gained in experimental and computational data. While there is overlap here with data fusion the focus is more on quantifying errors over improving nominal predictions. The work in this dissertation represents just one portion of the research being done on UQ for the SLS program. And while the techniques used in previous SLS databases are not directly relevant to the current work, the framework and methodologies, influenced the framing and development of the UQ work here.

In this work, the aleatoric uncertainty has been incorporated into the force and moment database, which also takes into account the epistemic uncertainty associated with the limitations of the wind tunnel testing and is provided as a normal distribution with a mean of zero and a prescribed standard deviation. The UQ development in this work will focus on characterizing the uncertainty in the proposed ROM as well as how the force and moment uncertainty interacts through the data fusion process. Due to cost of fully characterizing uncertainties in the CFD source data, it is unfortunately neglected for this work. Methodologies derived from jackknife and bootstrapping will be heavily used to build a confidence interval on the ROM coefficients that are then propagated [Efr82]. While there have been attempts at

characterizing the error of these types of ROMs, extension to the resulting spatial distribution have not been characterized [Bra11].

2.6 Details and Limitations of Other Methods

Since there is significant work being done with similar ideas and goals, an in depth comparison between proposed work and other research is essential for discussion. The first subsection focuses on the work by Carlson et al. [Car15], from which the present work is directly built upon and inspired by. The second subsection looks at work being done by other members on the SLS team with similar goals of lineload adjustment to reduce the differences between various databases but for a different part of flight. The third subsection is dedicated to work being done in Germany, and is also closest to what is being proposed here. A summary of the key differences can be found in Table 2.2.

2.6.1 Comparison to Similar Work

Research by Carlson in Ref. [Car15] has been foundational to the research in this dissertation, however, the major difference is the process ascribed to data fusion. This dissertation also expands on the previous work by including an approach to incorporate experimental uncertainties.

For the work by Carlson et al., the first step is to globally adjust the CFD generated surface data by experimental pressure port data. These adjusted CFD data are then used to generate a ROM. This ROM is then fused with experimental force and moment data to create what they call a Physics-Based ROM (PBROM)[3]. Their adjustment creates a balance between the pressure port data and the integrated forces and moments. The method developed in this work matches exactly the integrated forces and moments while neglecting the pressure data. The reasons for this decision are elaborated on in Section 4.1.1, but are primarily driven by the introduction of load cells in the wind tunnel experiments that give direct measurements of the SRB forces and moments. Also, the collection of this pressure data was a secondary goal of wind tunnel testing and so there are conditions where data are missing and there was no effort to characterize uncertainty in the pressure data.

Carlson uses radial basis functions to guide the PBROM toward integrated forces and moments as opposed to the current work, which uses Lagrange multipliers to create a constrained optimization problem. Also, the inclusion of the pressure data to adjust the surface data before the ROM creation ended up creating difficulty during the adjustment to integrated forces and moments in particular the axial force. This work however inspired other members of the SLS team to explore similar techniques.

2.6.2 Other Method Developed in the SLS Program

The method developed by Dalle et al. in Ref. [Dal18] is closest from the data fusion point of view. This methodology focused on a similar problem for the SLS vehicle during the ascent phase of flight. This phase of flight is characterized by flow angles less than 10° and speeds from Mach 0.5 to 3+.

[3]The idea that a model built from other models is or is not physics based generates significant discussion. My opinion is that it incorporates the trends of the data but it is not actually physics based.

Compared to the flow regime in the present work, compressibility effects are a bigger concern. Also, instead of large wake dominated flow features, shocks tend to be the dominant feature, which creates significantly different loadings across the surface of the vehicle. This results in different priorities in what features should be maintained when adjusting lineloads. The technique I develop in Chapter 4 takes inspiration from Dalle by solving a constrained optimization problem by the method of Legrange multipliers; however, the constraints are significantly different.

The most significant difference is that Dalle's method works only on the lineloads. This loses out on the ability to look at how various forces and moments are correlated. For example, his method couples only normal force and pitching moment and then separately side force and yawing moment. The vehicle can be assumed to be symmetrical for most analysis, but in theory a mismatch in normal force on say the SRB noses can produce a net yawing moment, which is not caused by the side force.

Since Dalle's method does not work on surfaces, it lacks the ability to compare adjusted pressure data to wind tunnel data. While this pressure comparison is not critical for lineloads, it does allow for an extra check on the final result. Working with just the lineloads also loses the ability to adjust friction and pressure forces individually. This is another bit of information that is lost, which limits analysis on the final results. Dalle's method also has the ability to incorporate uncertainty values from the force and moment data. The implementation is limited though in that the adjustment is the same across the parameter space where we would expect the regions of higher uncertainty across a vehicle to vary at different flow angles. Overall this method is more robust than the proposed but sacrifices analytical capability to achieve it.

2.6.3 DLR Research

The paper by Mifsud et al. [Mif19] is closest to work presented here by directly predicting a full surface and using a similar constrained optimization problem to ensure matching target forces and moments. The work looks at an industrial wing-body aircraft at cruising conditions, which is a flow regime that most CFD solvers excel at. The technique presented by Mifsud makes adjustments to the surface solution by a least squares problem trying to optimize pressure measurement matching with constraints for integrated forces and moments. Across the parameter space, the different modes will dominate due to changes in flow features; however, Mifsud et al. rely entirely on this least squares fitting, which could in theory make it significantly different than the original CFD solution.

While full surfaces are being computed, the quantities of interest are integrated forces and moments, which are significantly different from lineload, which do not have a simple experimental analogue. This makes the quality check on the data fusion method much more difficult and requires more reasoning about what choices make sense. Also, uncertainty quantification is not addressed in Mifsud's paper.

While all three of these highlighted methods focus on fusing data from both computational and wind tunnel sources, the present work is unique in several important respects. For one, the presented data fusion method focuses on minimizing the changes from known high fidelity CFD surfaces explicitly. Across the parameter space, different PCs have different dominance especially in the liftoff and transition regime and so this constraint is crucial in keeping answers that are physical especially given the relative

Table 2.2 Summary of comparison between current work and previously done work.

Work	Flow Regime	Vehicle	Data Adjusted	Data Sources	Uncertainty Quantification
Carlson et al.	Subsonic Wake Dominated	Rocket	Surface Data	CFD, Pressure Ports Forces and Moments	None
Dalle et al.	Transonic, Supersonic, Shock Dominated	Rocket	Lineloads	CFD Forces and Moments	Yes
Mifsud et al.	Transonic	Airplane	Surface Data	CFD, Pressure Ports, Forces and Moments	None
Wignall	**Subsonic** **Wake Dominated**	**Rocket**	**Surface Data**	**CFD** **Forces and Moments**	**Yes**

sparseness of input data. The focus of this work on the liftoff and transition regime significantly increases the difficulty of the problem as well as constrains the possible solutions that can be used. This regime has significantly different flow features at low, midrange, and high angles of attack, which requires any methodology applied to it to be robust enough to handle this variety. Also, due to the bimodal stability in the asymmetric vortices, the linearity assumptions built into PCA are stretched to their limits. The final innovation of this work compared to similar research in the field is the robust uncertainty model. This model is a function of the independent variables in the parameter space and is sensitive to both the uncertainty in the force and moment data as well as the model form error. Hopefully, expansions to this work will include accounting for uncertainties in the CFD data as well.

CHAPTER

3

REDUCED ORDER MODEL

3.1 Reduced Order Model

As outlined in the introduction, the need for aerodynamic databases requires data all throughout the parameter space. While simulations can be run at each of these conditions of interest, the time and computational costs required for all the necessary points are impractical. To help overcome these limitations, reduced order models (ROMs) were explored and developed to help fill in the parameter space using a handful of points.

A ROM is a mathematical model derived from a sequence of high fidelity observations or simulations that has two coinciding benefits: the ROM preserves the essential physics and predictive capability of the original dataset and also can be interrogated at a much lower computational cost. In this work, time-averaged CFD results generated in Kestral [Mor09] (with details explained in Ref. [Rataa]) provide the high fidelity observations that build the ROM. As a result, the ROM then can approximate the surface grid solution at any angle of attack and roll combination. This method is solver independent and in theory could be implemented using solutions from multiple solvers. While not common to use multiple CFD software to generate aerodynamic databases, it sometimes happens as different physics become dominant. For example, in entry descent and landing, it is common to use a code designed for reacting flows at high altitudes and high Mach numbers while using a more typical Navier-Stokes equation solver at lower speeds. The major requirement is that all the solutions are on a common grid, which can be accomplished through interpolation. Once all the data are gathered and interpolated onto a common surface grid, model generation can begin.

This chapter splits the model generation into two parts, basis extraction and recombination. Principal component analysis (PCA) (aka proper orthogonal decomposition) [Jol02] is used as a means to extract

the basis of the desired solution space. The extracted basis, the principal components (PCs), are then linearly recombined to provide an approximation of the CFD solution at new conditions. While different variations on the PCA framework are explored, only one is selected for further development in the next chapter. The results for the ROM are then compared to wind tunnel data and CFD check points to prove its predictive capability. This comparison will show how this relatively cheap computational model is able to predict integrated forces and moments with similar accuracy to full fidelity CFD simulations.

3.1.1 Model Development

Since the force and moment database is split by component (core and the two SRBs), the ROM is split into the three vehicle components. These components do not interact with each other once separated, so the first step is taking the CFD results and breaking them into these components. For reference, a typical pressure distribution from CFD is presented in Fig. 3.1. This template, which is used frequently in this work was developed by Steve Krist for the SLS program. It has the upper surface of the vehicle in the top half and the lower surface of the vehicle in the bottom half as well as a front view insert. While at this roll angle the upper surface is the leeward side that is not always the case. The template goes back to earlier vehicle configurations; however, with the Block 1B configurations the front view was included to be able to better visualize the effects of the asymmetric vorticies. As can be seen on the upper surface, there are large regions of separation, as previously mentioned, that make this problem difficult to solve numerically.

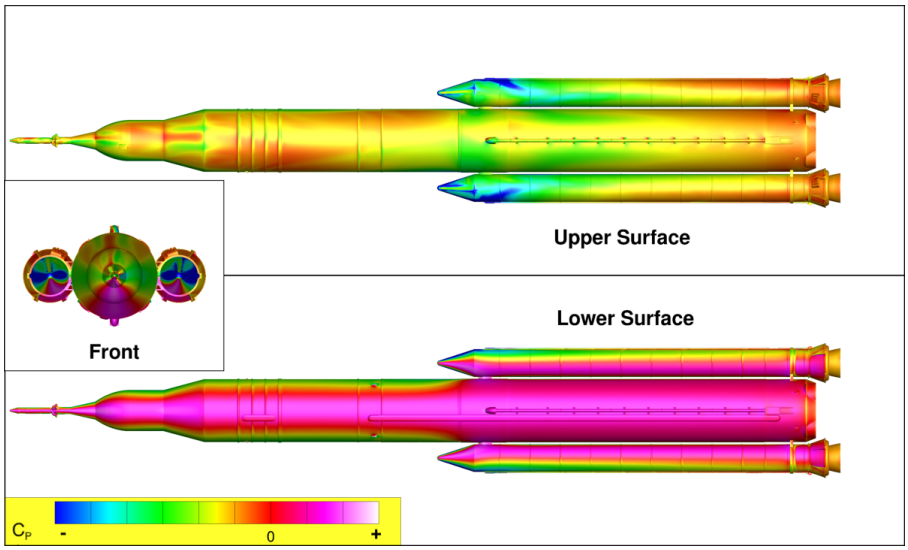


Figure 3.1 Time-averaged surface C_P distribution of Block 1B crew vehicle at α_P = 50°, ϕ_P = 0°. Template courtesy of Steve Krist.

The sample solution shown (as well as others) is then split into the three vehicle components: core, port SRB and starboard SRB, and then integrated to get the sectional loads as seen in Fig. 3.2. The top

graph contains the sectional forces acting in the axial direction, the middle graph contains the sectional forces acting in the side direction, the bottom graph contains sectional forces in the normal direction, and below the graphs is the outline of the launch vehicle for reference. The thin grey dashed line represents zero to help give reference. The x-axis in the figure coincides with the x-axis of the SLS system. For comparison purposes, the normal force and the side force lineloads have the same scale. Normal force and side force having the same scale will be consistent for all plots unless specifically noted otherwise. Due to the asymmetric vortices on a given section, the side force can be greater than the normal force and having a consistent scale helps highlight this phenomenon. This figure was based again on a template by Steve Krist when working on the Block 1 Crew vehicle. I however made changes to it, particularly the addition of the dashed guide lines to note the location of zero forces, which is particularly helpful for analyzing the side force, and also lining up with the outline of the vehicle. As mentioned in the introduction, the aerodynamic analysis for the LOT domain is in the missile axis system, which becomes more important at other roll angles.

Figures 3.1 and 3.2 represent the desired output of the ROM. The surface allows for further analysis of the flow physics in the LOT domain, which would not be possible due to resource constraints if only CFD was available, and the lineloads constitute the data to be delivered in the databases.

Once the CFD data have been gathered and split up, PCA is used for feature extraction. There are many techniques available that can be used to extract features, but PCA was selected due to its simplicity and use in previous work. The friction and pressure are kept separate during the PCA portion of the ROM creation to allow for the ability to analyze how the ROM predicts viscous contributions separate from the pressure components. Separation also allows for the ability to adjust them independently as seen later. For a given component, such as the core, the CFD solutions were further broken into 4 snapshot matrices, S, (one for each variable) with rows of cells and columns of solutions similar to the example presented in Fig. 3.3.

The resulting matrices have on the order of a hundred thousand rows with tens of columns. These matrices are then used to find the covariance matrix of the solutions, which is computed as follows:

$$R_{i,j}^{l,k} = cov(S_i^{l,k}, S_j^{l,k}) \tag{3.1}$$

where S_i and S_j represent the surfaces from distinct CFD solutions (columns of the S matricies), k represents the component under consideration (core, port SRB, or starboard SRB), and l represents the variable under consideration (C_P, C_{F_x}, C_{F_y}, or C_{F_z}). The resulting covariance matrices, $R^{l,k}$, are square matrices that are sized based on the number of input CFD solutions. That is, if 21 CFD cases are used, $R^{l,k}$ would be a 21×21 matrix. For reference, 21 CFD cases were used in the development of the Block 1B crew vehicle and 60 for the Block 1B cargo configuration. The difference in the number of points used was primarily due to the available computational resources, scheduling constraints, and increased complexity of flow around the Block 1B cargo vehicle. The database delivered for the Block 1B crew vehicle was developed from data generated for one quadrant ($0 \leq \phi_P \leq 90$) and reflected to the other quadrants ($90 < \phi_P < 360$) while the database for the cargo vehicle relied on data from all 4 quadrants.

Once the 12 covariance matrices (3 components with 4 variables each) are found, the next step is

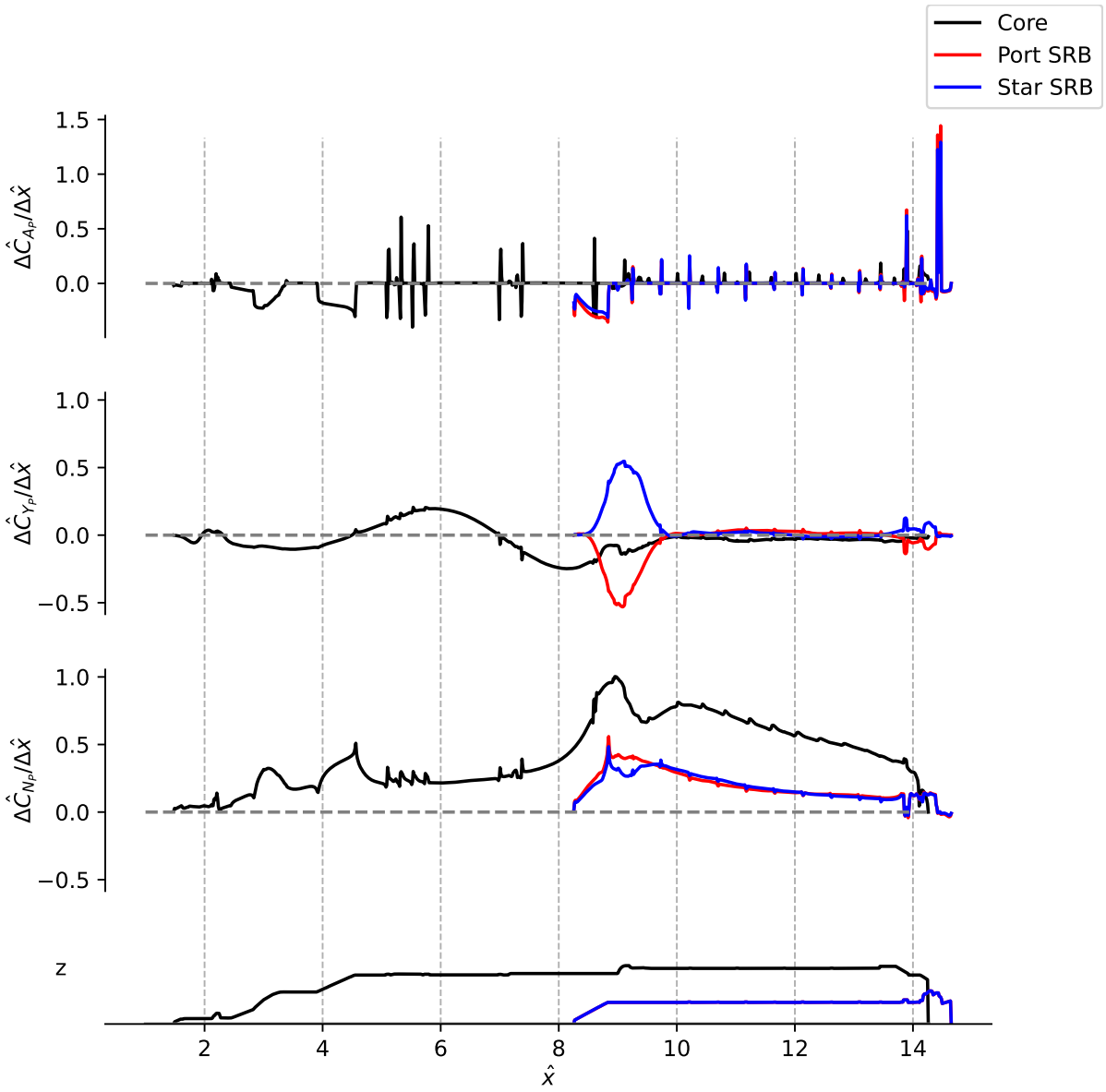


Figure 3.2 CFD Lineloads for Block 1B crew vehicle at $\alpha_P = 50°$, $\phi_P = 0°$.

to calculate their eigenvalues and eigenvectors. The eigenvectors are needed to transform the matrix of CFD solutions into PC space while the eigenvalues are used later in the data fusion portion. The surface distribution Ψ of a given PC is defined by:

$$\Psi_j^{l,k} = \sum_{i=1}^{n} V_{i,j}^{l,k} S_i^{l,k} \tag{3.2}$$

where the sum is taken over the input CFD solutions and $V^{l,k}$ are the eigenvectors of the covariance matrix $R^{l,k}$. There is one set of PCs for each variable and component.

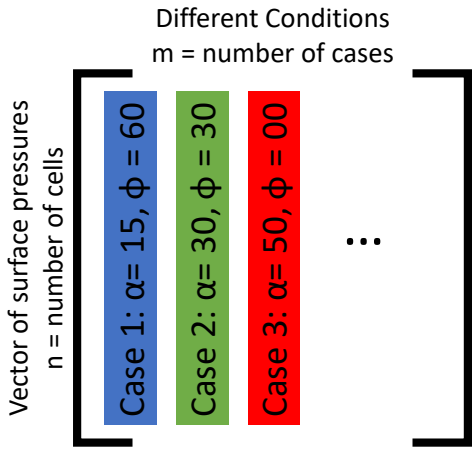


Figure 3.3 Setup of snapshot matrix used in PCA.

One of the advantages of PCA, and the primary reason it was selected, is how the transformed PCs are ordered. The nature of PCA orders the new PCs by the amount of variance they contribute to the system. This relative importance is reflected by the eigenvalues, which are plotted in Fig. 3.4. The eigenvalues are normalized by the sum of the eigenvalues within variables for the Block 1B crew vehicle and then plotted against the PC number. The values on the x-axis correspond to the numbering of the PCs. The trends shown in the figure help demonstrate how the first few PCs dominate the construction of the ROM surfaces. The first three PCs related to pressure contain almost 90% of the total pressure in the system. Since the normalizations are within variables, the difference between the eigenvalues of the different variables is lost. The smallest pressure eigenvalue is an order of magnitude greater than the largest eigenvalue that corresponds to friction. Other transformation options are discussed in brief in section 3.1.1.1; however, they all have a similar technique.

After converting to PC space, PC surfaces can be plotted as presented in Fig. 3.5, which shows the first PC related to pressure on the core of the Block 1B crew configuration. Since PC space is outside of the LOT domain, there is no corresponding angle of attack or roll angle. This surface shows large features, and could easily be mistaken for the pressure distribution from a CFD solution. As the PC number goes up, the feature sizes decrease and the overall surface becomes less like a CFD surface solution due to the lack of coherent structures.

Integration of PC surfaces into lineloads gives an idea of the shapes that individual PCs contribute to the lineloads. An example of these lineloads can be seen in Fig. 3.6. Figure 3.6 shows various PC lineloads corresponding to the integrated surface pressures for the Block 1B crew configuration including the PC shown in Fig. 3.5. PC lineloads are not in the missile axis system since they are outside of the LOT domain and as such the lineloads do not show axial, side and normal force but X, Y, and Z directional forces. In the SLS system, these are identical to the body axis forces; however, since the analysis and other results shown are in the missile axis, the distinction is made here to stress the difference. While integration into sectional forces hides some of the lack of coherence of the underlying surfaces, it is still clear that the higher numbered PCs tend to contribute much less to the output than the lower numbered

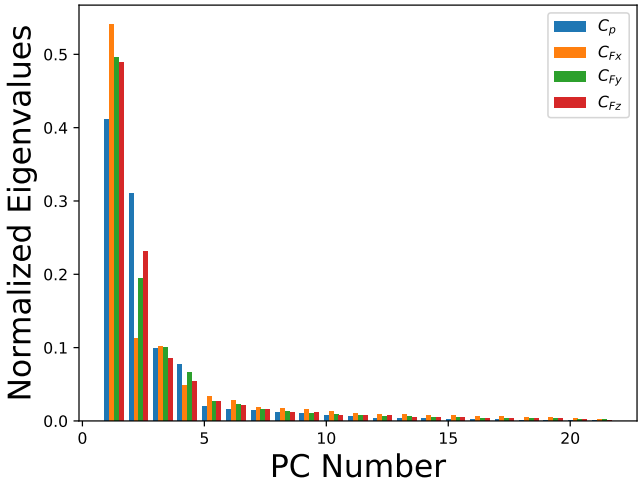


Figure 3.4 Normalized eigenvalues of the four variables for the core of the Block 1B crew vehicle.

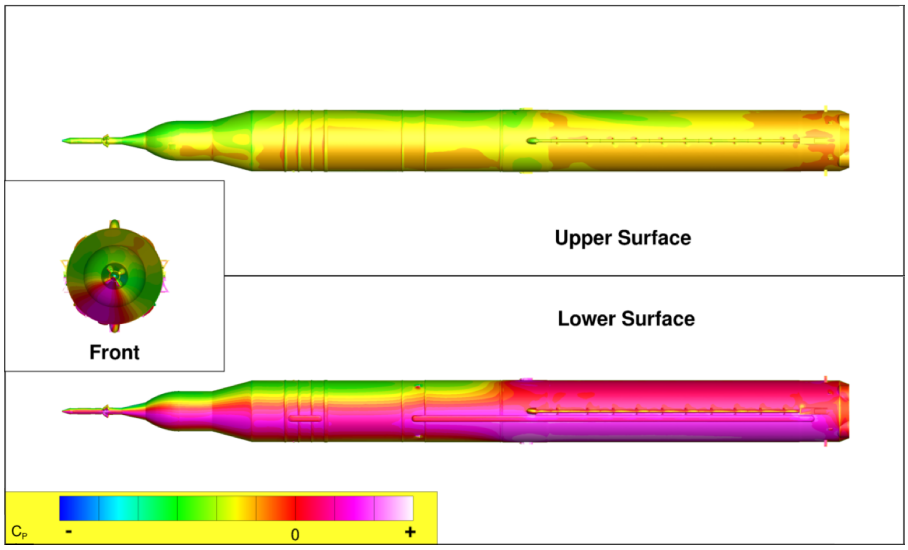


Figure 3.5 Surface distribution of the first Principal Component related to pressure on the core of the Block 1B crew configuration. Based on plotting template by Steve Krist.

ones. This further reinforces the results of the eigenvalue analysis above. Because the data are not mean centered beforehand, the first mode corresponds closely with the overall mean. Since all modes are retained, this does not significantly influence the ROM predictions.

By taking the inverse of the definition of the PCs in Eq. 3.2, a linear combination of the PCs can be used to recover the original CFD solution surfaces,

$$S_j^{l,k} = \sum_{i=1}^{n} \hat{V}_{i,j}^{l,k} \Psi_i^{l,k} \tag{3.3}$$

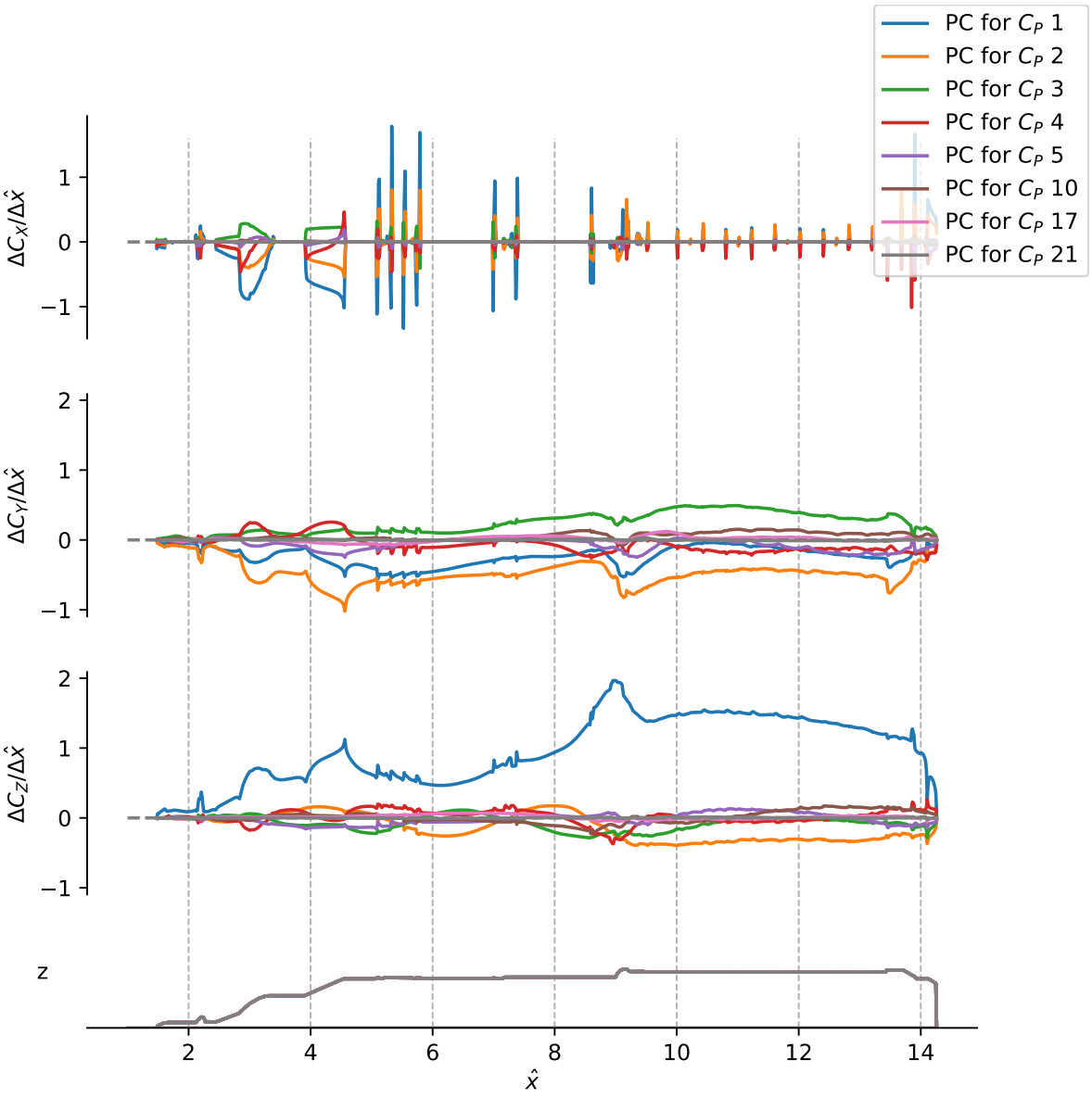


Figure 3.6 PC distributions related to pressure integrated into lineloads on the core of the Block 1B crew configuration.

where $\hat{V}$ is the inverse of the matrix of eigenvectors. The next step is to modify this equation to allow for calculating solution surfaces at conditions outside the CFD run matrix. This is done through replacing the inverse of the eigenvector with a coefficient, a, which is a function of α_P and ϕ_P.

$$S^{l,k}(\alpha_P, \phi_P) = \sum_{i=1}^{n} a_i^{l,k}(\alpha_P, \phi_P) \cdot \Psi_i^{l,k} \tag{3.4}$$

The remaining piece of the puzzle is to determine the ROM coefficients a, which apply proper weighting to each PC in the sum in order to arrive at the surface for each condition. The most straight forward

way to get the necessary values for a, is to take the inverted values from the matrix of eigenvectors and assign them to the corresponding conditions of the CFD solution. From those handful of scattered points, the rest of the domain is calculated through interpolation. An example distribution of the ROM coefficient can be seen in Fig. 3.7, which shows the first ROM coefficient associated with pressure on the core of the Block 1B crew vehicle as a function of angle of attack and roll. The circles show where there were originating CFD data in the domain. As can be seen, the coefficient is generally smooth with a peak in one corner (top left). It shows that a significant amount of the variance in the system is generated from the CFD solution at $\alpha_P = 70°$, $\phi_P = 0°$. As the PC number increases, the coefficients tend to get more peaks and valleys and as a result are less smooth.

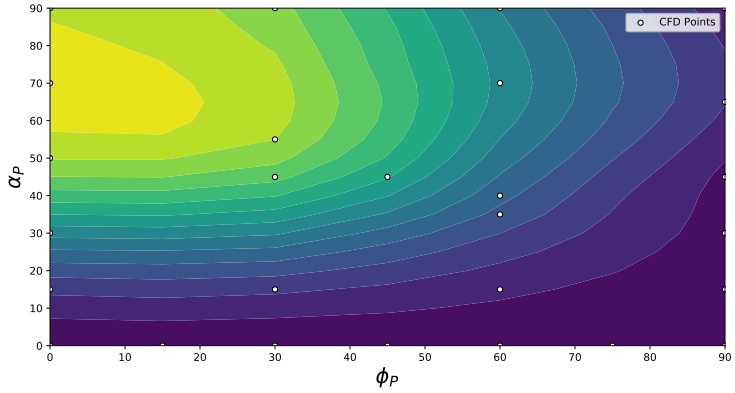


Figure 3.7 The values for the first ROM coefficient for pressure across α_P and ϕ_P for the core of the Block 1B crew vehicle. The dots represent locations of CFD points.

At this point, the basic ROM is complete. The model can take in desired angle of attack and roll conditions and return an approximation of a CFD surface as desired. A benefit to this form of ROM is that since the ROM coefficients are a function of condition only (and not any spatial variable), they can be removed from spatial integration. While the entire surface can be used to check for quality, only the lineloads are needed for the final database. With the ability to integrate the PCs spatially, this allows for smaller datasets to be saved while keeping the ability to fully create the database.

3.1.1.1 Other Transformation Basis

Along with using the covariance matrix, the correlation matrix can be used to transform the data, which would give highly correlated PCs instead [Jol02]. This difference in formulation has almost no effect on the ROM result unless PCs are dropped. However, because the eigenvalue distribution differs between these two decompositions, the data fusion approach, which relies on a weighting term derived from the egienvalues as described in the next chapter, will have a noticeable sensitivity to the modal decomposition method. PCA can also be performed by using the singular value decomposition (SVD) as well.

This will give a similar transformation to the covariance with different magnitudes of the eigenvalues (or singular values in this case). And again, the ROM portion will not have significantly different results from the covariance method presented here unless PCs are dropped, which is unnecessary due to the manageable number of snapshots.

Usually in PCA, mean centering is applied to help prevent the first PC recreating the mean of the dataset. This was also explored and had little effect on the PCs and the ROM generated. A likely reason is because most of the data are centered around 0 already. Because we examine conditions from all roll angles, a location on the vehicle with high values at one roll angle will have corresponding low values at another.

A possible modification to the current work could combine the friction components into one basis set, which would result in two sets of transformations instead of the four presented. This was not explored but would be in line with other work exploring similar methods where the three velocity directions are combined. This modification would also help ensure one could add other physical constraints to the system if necessary.

A linear method was selected due to its ease of implementation and understanding its inner workings. This relative simplicity becomes more useful when moving onto the data fusion method making it easier to trace the relationship between the various elements. The data fusion also helps address some of the limitations of the linear methodology as explored in the next chapter. Not discussed in this work, but the remaining limitations of the linear assumptions are addressed by ad hoc methods which were developed and implemented for the final lineload databases to cover the multimodalities, in particular the more extreme side force loading due to the asymmetric vortex system. Nonlinear transformations were not explored for this work but could provide possible solutions to the shortcomings of the presented method when dealing with nonlinearities in the solution space. Despite the significant nonlinearities, the ROM model as designed does a successful job of helping predict data at new points.

3.1.2 Model Results

Once the ROM coefficients and PC surfaces are calculated, it is possible to start outputting surfaces. The pressure surface in Fig. 3.8 is the output of the ROM for the Block 1B crew configuration at $\alpha_P = 60°$ and $\phi_P = 0°$, which is not a condition that was not simulated. This condition is relatively close to the condition shown in Fig. 3.1 and so similarities are expected. The most noticeable and most important is the increased area of separation as the angle of attack increases. This phenomenon is a significant contributor to the lineloads and so accurately predicting the extent of separation is necessary for accurate overall predictions. Additionally, on the upper surface, despite the core and SRBs being calculated separately, there are flow structures that emanate from the SRB noses and continue onto the core. The flow separation and the continuation of features from the SRBs to the core shows that the ROM maintains at least some of the physics of the original CFD. This level of analysis is possible since the ROM is built off and recreates entire surfaces and not just lineloads or forces and moments.

Since it is now possible to output surfaces at any condition desired, the ability to check each output quickly becomes overwhelming. As mentioned in Section 2.1.3.2, the availability of wind tunnel data

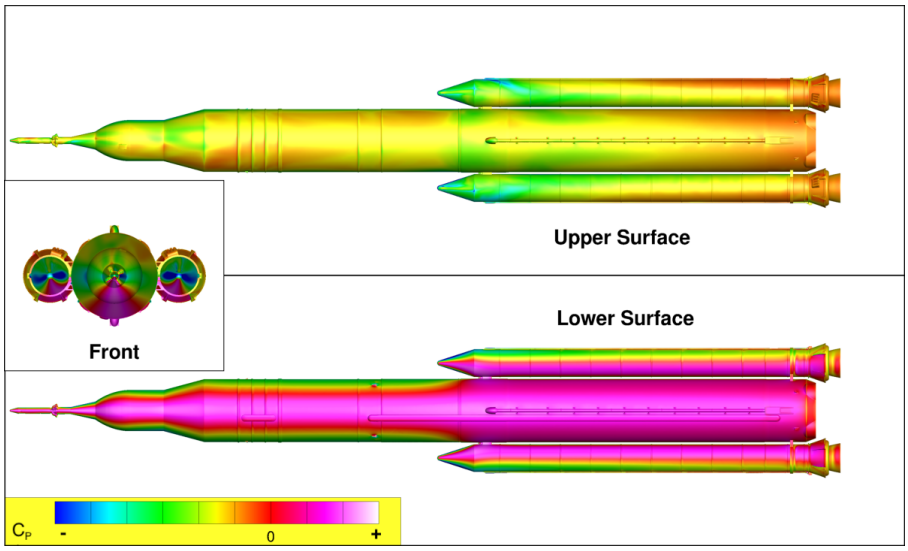


Figure 3.8 ROM output of surface pressures for the Block 1B crew at α_P = 60° and ϕ_P = 0°. Plotting template provided by Steve Krist.

allows for a check of quality through force and moment data. A typical force and moment comparison can be seen in Figs. 3.9 and 3.10. Force and moment data for the Block 1B crew configuration as a function of angle of attack at ϕ_P = 0° are seen in Fig. 3.9. Similar data at ϕ_P = 120° for the Block 1B cargo configuration are seen in Fig. 3.10. The black symbols represent data gathered from wind tunnel testing, the red circles are the data from CFD, while the line represents the output of the ROM. The top half of the figures are, from left to right, the axial, normal, and side force coefficients of the entire vehicle. The bottom half of the figures are, from left to right, the rolling, pitching, and yawing moment coefficients of the entire vehicle. The data shown are for the total vehicle, which means they compare the three component ROMs added together to data derived from the balance reading. As in previous plots, the normal force and side force coefficient plots have the same scale. The same holds true for the pitching and yawing moment coefficient plots. These figures show that the ROM exactly reconstructs the integrated force and moment coefficients generated from the CFD results, since the ROM coefficients are defined by the inverse of the eigenvector matrix. The figures also shows that the ROM is able to generate the same overall trends of the wind tunnel data. The locations that the ROM misses are also areas that the CFD misses, which are the mismatches that prompted the exploration of data fusion techniques. The biggest of these mismatches are the side force and yawing moment coefficients, which are an effect of the asymmetric vortices as displayed for the Block 1B crew forces and moments. The other important trend is that the CFD tends to underpredict peak normal and pitching moment coefficient values as displayed by the Block 1B cargo data. While rolling moment coefficients look to be poorly matched for the Block 1B crew vehicle, those values are very small and the level of disagreement is on the same order as the rolling moment coefficients seen in the Block 1B cargo vehicle.

As another check on the ROM's ability to recreate CFD level quality, leave one out (LOO) analysis was done. To clarify, the ROM was recreated while leaving one of the contributing CFD cases out,

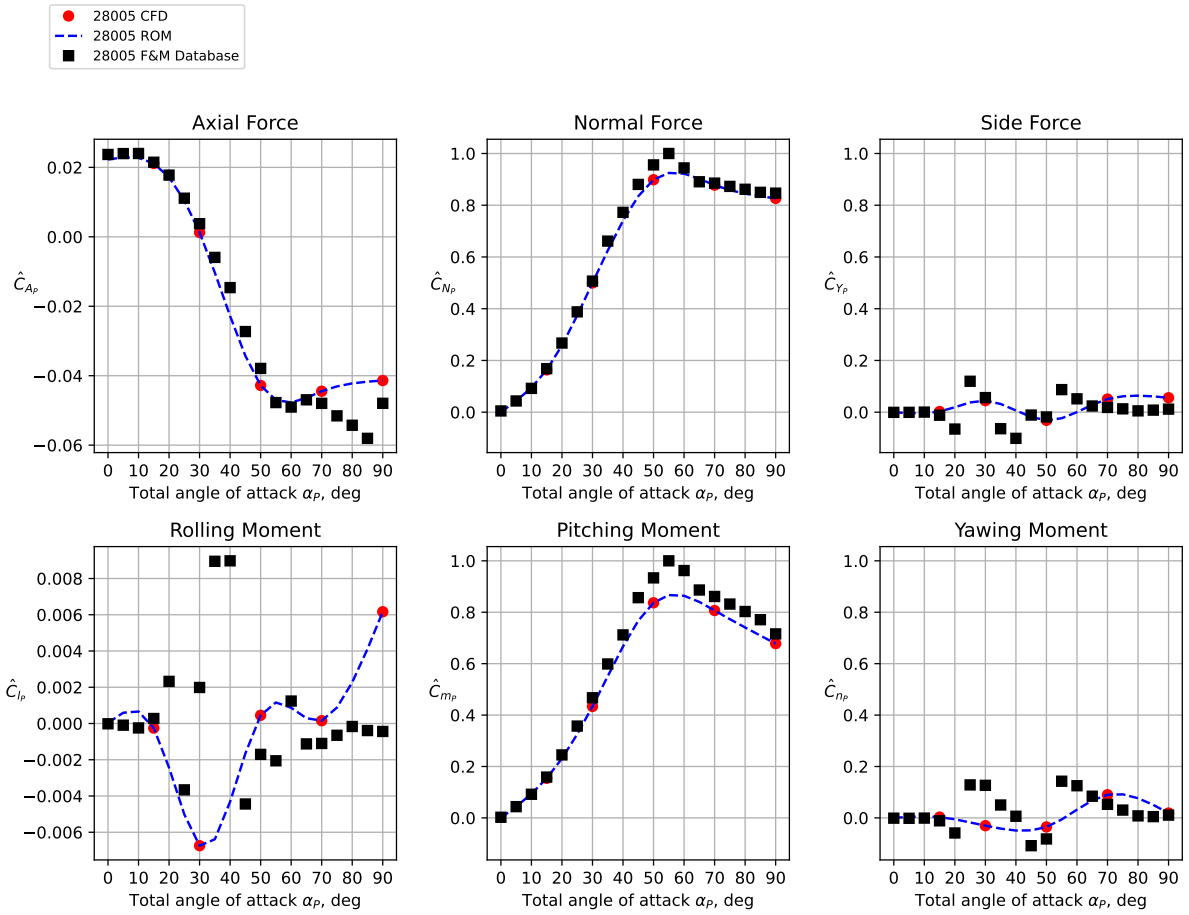


Figure 3.9 Comparison of wind tunnel force and moment coefficients versus angle of attack at $\phi_P = 0°$ for the Block 1B crew vehicle.

which could then be used as a check. The new dropped point ROM was compared to the ROM at the CFD condition that was dropped to examine the ROM sensitivity to missing information, as well as being able to compare it to CFD data. As a reminder, the ROM exactly recreates the CFD data when at a point with CFD. Two example cases are presented through lineload comparisons. The Block 1B crew vehicle shown in Fig. 3.11 shows a flip in the vortex prediction as seen in the side force sign change based on dropping the CFD point at $\alpha_P = 50°$, $\phi_P = 0°$. This sign flip does not move the dropped point ROM outside the realm of CFD-like data since CFD was able to recreate the vortex on either side by changing settings. The Dropped Point ROM prediction for the normal force accurately recreates the forebody portion of the full ROM almost exactly, with only slight differences near the aft end. The axial force lineload has relatively little change except a general decrease in axial force coefficient near the aft end, which mirrors the changes observed near the aft end in the other lineloads. Overall, the dropped point ROM for this case confirms the ROM's ability to approximate CFD solutions.

The Block 1B cargo vehicle result shown in Fig. 3.12 shows one of the worst cases that was encountered at $\alpha_P = 45°$, $\phi_P = 270°$. Here once again, there is a vortex prediction flip when dropping the CFD

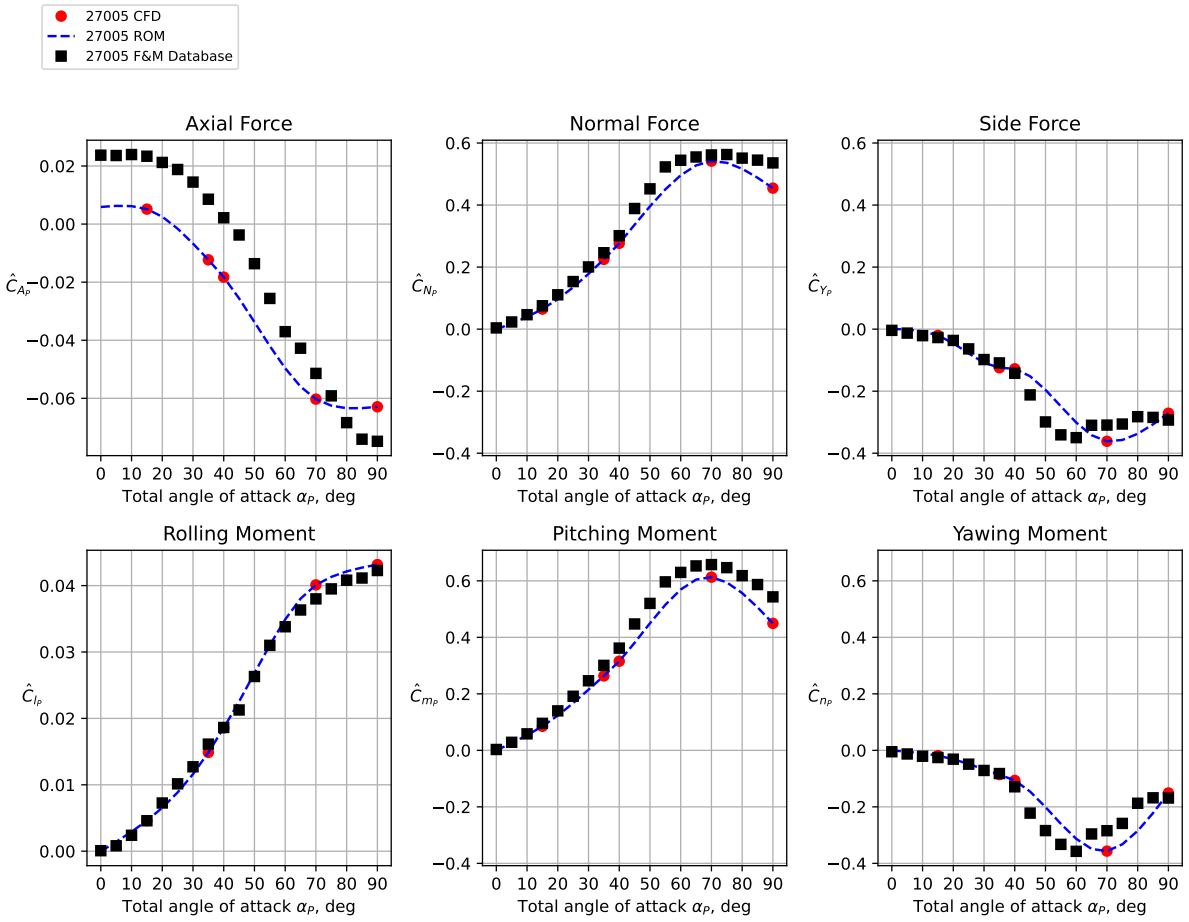


Figure 3.10 Comparison of wind tunnel force and moment coefficients versus angle of attack at $\phi_P = 120°$ for the Block 1B cargo vehicle.

point, however, the effects are much greater in this case. Since the whole surface is affected by the vortex flip, the changes in the normal force can also be attributed to this especially since most of the changes are in the middle section of the vehicle forward of the SRBs, which is the section of the vehicle that is affected most by the forebody vortices. Despite being a considerably different form, the integrated values for the ROM and the dropped point ROM are fairly similar. Thus, while the ROM may not always reproduce CFD data as accurately as desired, it can adequately give an approximation of what the CFD would produce.

A ROM was developed that is able to provide results at a variety of conditions that compare favorably to CFD results and database forces and moments. This was demonstrated through comparison to the wind tunnel force and moment database where it captures the trends seen while keeping some of the shortcomings in the base CFD. The robustness of the ROM is demonstrated through LOO analysis that shows that the ROM can adequately recreate CFD data.

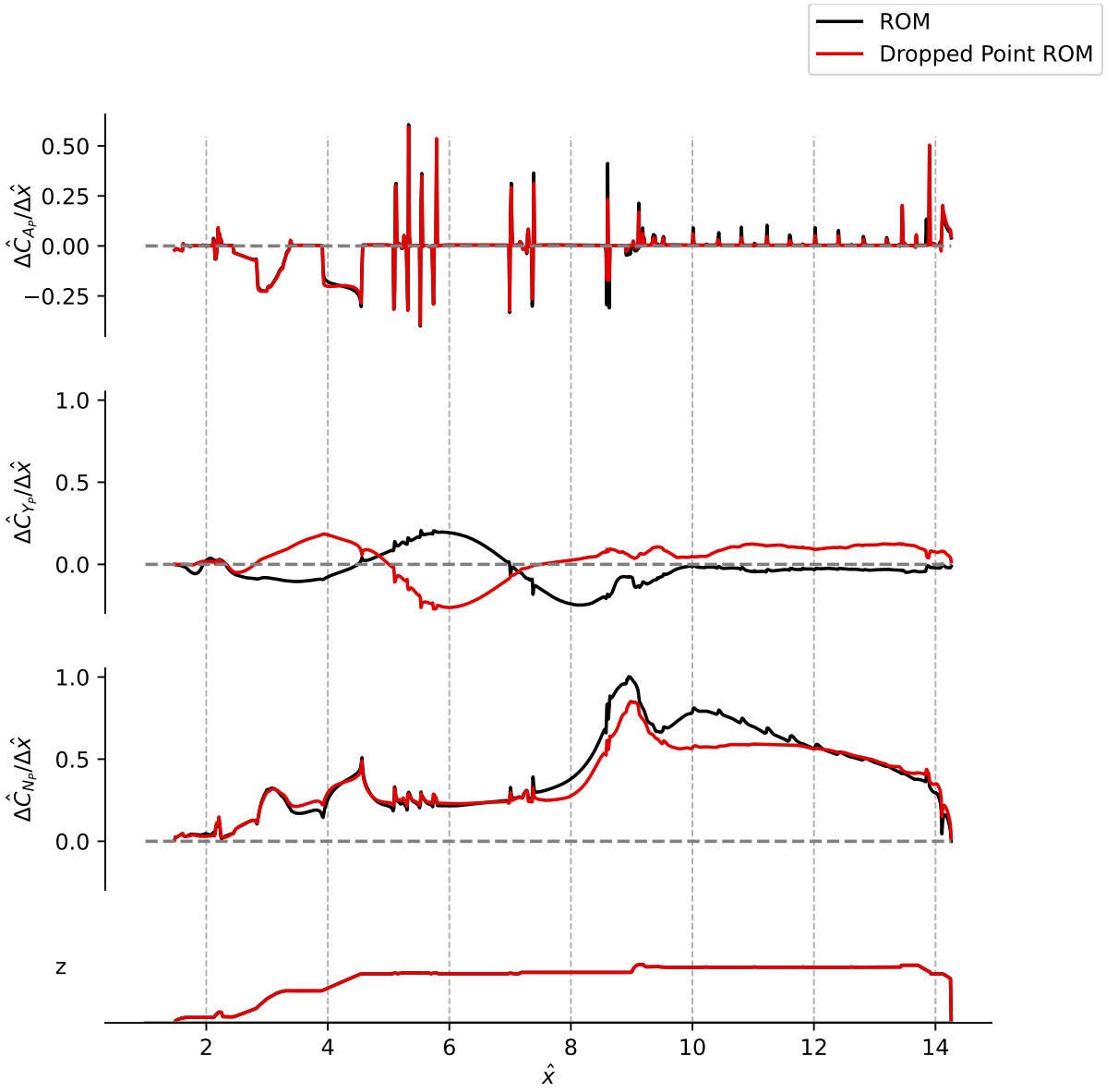


Figure 3.11 LOO ROM lineloads for the core of the Block 1B crew configuration at $\alpha_P = 50°$, $\phi_P = 0°$.

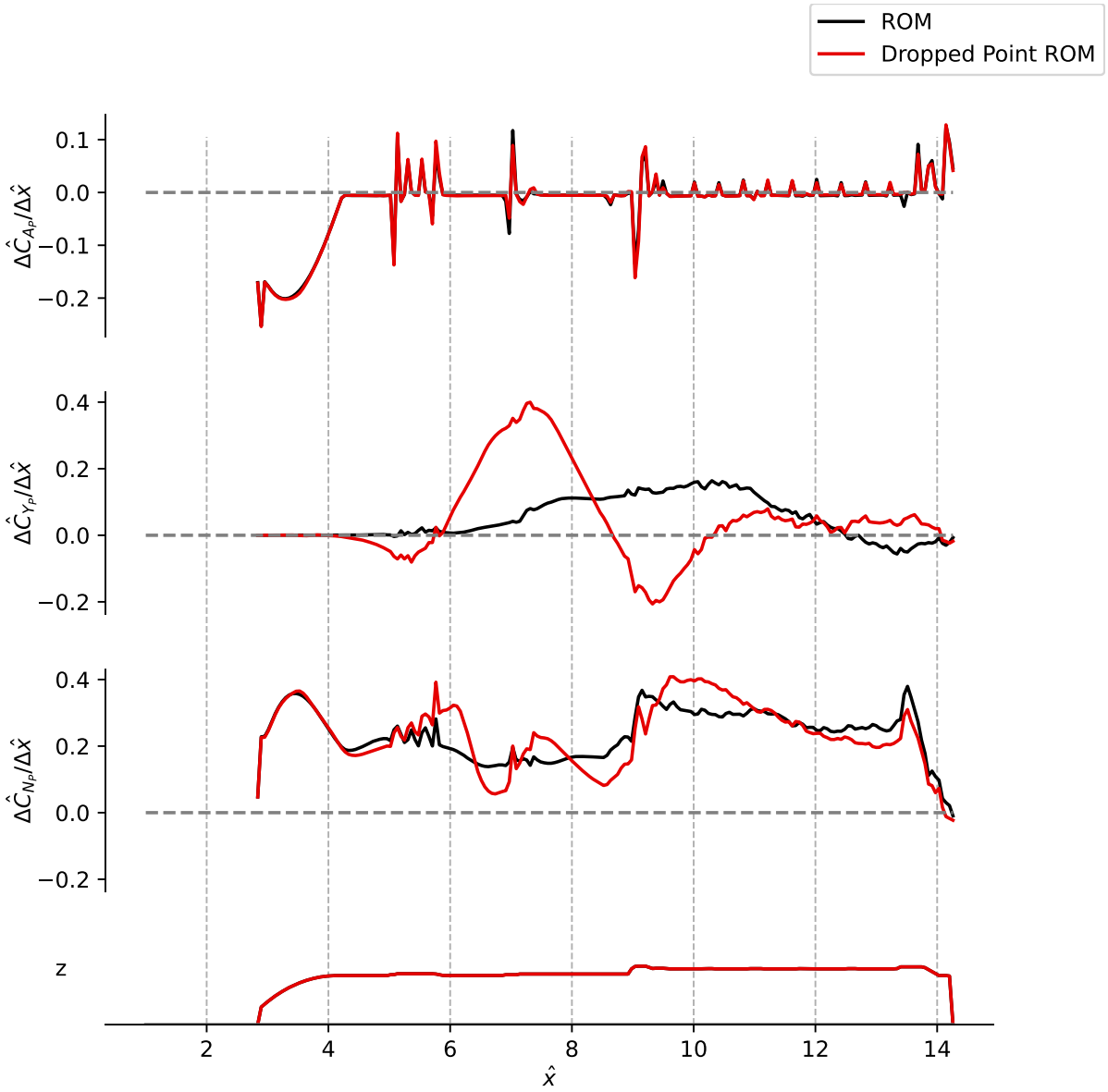


Figure 3.12 LOO ROM lineloads for the core of the Block 1B cargo configuration at $\alpha_P = 45°$, $\phi_P = 270°$.

CHAPTER

4

DATA FUSION

The data fusion method developed here takes the model as described in section 3.1.1 and finds a new set of data-fused ROM coefficients (called DFROM coefficients after adjustment) that forces the generated surfaces to integrate to desired forces and moments. As stated previously, while CFD is trusted, it is computationally impractical to fully resolve the wake in these flows with massive separation and produce enough data to generate the required lineload database. The limitations in the computational fidelity to decrease costs led to the consideration that the wind-tunnel-derived force and moment database is higher fidelity than the CFD data. The inadequacy of CFD is displayed in the previous section particularly Fig. 3.9 where the CFD, and as a result, the ROM had difficulties in matching the integrated forces and moments observed in the wind tunnel. While further computational resources could be spent to improve these simulations, with each case already taking on the order of 100,000 CPU-hours, cheaper methods of improving predictions was desired. The goal was to develop a method to modify the ROM such that the integrated forces and moments match the values from the wind-tunnel-derived force and moment database without significant changes to the ROM lineloads.

This chapter goes over the methodology of taking the ROM developed in the previous chapter and combining it with the wind tunnel data. It sets a cost function that balances several competing needs and then uses that cost function to set up a constrained optimization problem. The derived cost function has a weighting term that can be adjusted and several different choices are explored and one is highlighted for its simplicity and quality of its results. After selecting the weighting term and solving the resulting constrained optimization problem, the resulting DFROM predictions are analyzed.

4.1 Data Fusion Development

Since there are many ROM coefficients that can be optimized and only a limited number of constraints are available, methods for constrained optimization were explored with the method of Lagrange multipliers being selected due to its simplicity. Determining the constraints of setting up the constrained optimization problem was a straight forward process. Since each component is analyzed separately, the superscript denoting component, k, will be dropped in this section. For force (or moment) j, the constraint is defined as:

$$\sum_{i=1}^{npc} \sum_{l=1}^{nvar} \hat{a}_i^l \Delta c_{i,j}^l - C_{F_j} = 0 \tag{4.1}$$

where $\hat{a}_i^l$ are the new DFROM coefficients of the i^{th} PC related to variable l and $\Delta c_{i,j}^l$ are the integrated j^{th} force (or moment) values of the i^{th} PC related to variable l. The higher fidelity force or moment data are represented by C_{F_j}, which are the j^{th} force (or moment) values from the database. This constraint ensures that the new coefficients of the DFROM return surfaces and lineloads that integrate to force and moment database values. Because of the difference in asymmetric vortex behavior between the Block 1B cargo and crew configurations, the forces and moments used for constraints varied slightly. For the Block 1B crew configuration, all force and moment constraints were used. For the Block 1B cargo configuration, only the normal force and pitching moment were constrained for the core vehicle but all six were used for the SRBs. Remember missile axis is used in the analysis,so side force and yawing moment will be the primary victims of the asymmetry regardless of ϕ_P. Notably, the differentiation between Block 1B crew and cargo configurations does not substantially change the methodology but does show the flexibility and ease of adding or removing constraints.

After determining the constraints, a cost function was defined to complete the optimization. Three characteristics were selected as most relevant for a cost function. First, since the ROM produces physically plausible results, it was undesirable to stray far from the initial ROM coefficients. Large changes in the coefficients represent large changes to the output lineloads; which while not as accurate as desired, are the best spatially distributed data available. Second, a cost function that is easy to use with both the method of Lagrange multipliers and also to differentiate with respect to the DFROM coefficients is an important characteristic to retain. Finally, to account for differences in variance across PCs, a PC weighting term was introduced. Given the above characteristics, the cost function developed for this work was:

$$f_{cost} = \sum_{i=1}^{npc} \sum_{l=1}^{nvar} w_i^l (a_i^l - \hat{a}_i^l)^2 \tag{4.2}$$

where w_i^l is a weighting term that allows for different PCs to be adjusted based on their relevance. Several options were explored for the weighting term before settling on using the corresponding eigenvalues from the PCA. This weighting method favored small changes to the PCs that correspond to large amounts of variance in the space, while allowing large changes in the PCs that contribute small amounts of variance to the space. As a result, the weighting proved most reliable in maintaining the overall shape of the lineloads without additional complexities.

4.1.1 Pressure-based Adjustments

One avenue explored that was not discussed in Ref. [Wigaa] was trying to match the wind tunnel pressure data from the wind tunnel experiment. The pressure ports are arranged in rings in several stations down the vehicle with approximately 8 ports per ring depending on local geometry. The data fusion method was modified so that the pressure at the port locations would match the data from the wind tunnel. This constraint only affects the pressure PCs and ROM coefficients associated with them.

Testing looked at restricting the study to a constrained optimization problem and only the Block 1B crew configuration, which had less CFD data. This meant only using a handful of the ports. One and two rings of ports and all the ports in a line down the vehicle were tested. Exact matching of the pressure ports was achieved at a cost of accuracy at unconstrained pressure ports. For example, when the second ring of pressure ports was used as a constraint, the third ring had significantly higher differences between the DFROM and the wind tunnel data. This choice of pressure port constraints also created significant gradients in the surface pressure data. One of the conclusions made at the time was that the sparsity of pressure port data prevents meaningful trends from being captured. In reflection, it could be that there were not enough PCs associated with pressure to allow for fine tuning of surface level details. As a result, revisiting this with the Block 1B cargo configuration may be valuable due to the increased computational data available. Another possible avenue to explore is using all available pressure port measurements; however, this would change the nature of the system. The system would become over-constrained, which would require looking at different methodologies and finding a way to balance all the constraints.

Also of note, compared to the work in [Car15] there are less pressure ports available. As part of the updates for the newer wind tunnel tests, pressures ports in the SRBs were replaced with load cells. This allowed for additional force and moment information at the cost of surface pressure information. Even looking at the results in [Car15], adjusting to the pressure information could cause some locally large changes. These large changes are reflected in the decrease accuracy in the axial force predictions.

4.1.2 Solving the Constrained System

Once the constraints and cost function are in place, the constrained optimization problem can be solved using the method of Lagrange multipliers. To give a brief overview, the constraints are multiplied by new variables, λ_j, which are called the Lagrange multipliers, and then added to the cost function to create the Lagrange function.

$$L = \sum_{j=1}^{ncon} \lambda_j \left[\sum_{i=1}^{npc} \sum_{l=1}^{nvar} \hat{a}_i^l \Delta c_{i,j}^l - C_{F_j} \right] - f_{cost} \tag{4.3}$$

The solutions to the problem are then found to be at the stationary points of this new equation. The stationary points are defined as the points where all the partial derivatives of a given function are zero, so the next step is to calculate all the partial derivatives and set them to zero.

The partial derivative with respect to the Lagrange multipliers, which represents the constraints has

the form:

$$\frac{\partial L}{\partial \lambda_j} = \sum_{i=1}^{npc} \sum_{l=1}^{nvar} \hat{a}_i^l \Delta c_{i,j}^l - C_{F_j} \tag{4.4}$$

while the partial derivatives with respect to the DFROM coefficients have the form:

$$\frac{\partial L}{\partial \hat{a}_i^l} = 2w_i^l(a_i^l - \hat{a}_i^l) + \sum_{j=1}^{ncon} \lambda_j \Delta c_{i,j}^l. \tag{4.5}$$

Because of the choice in cost function, the DFROM coefficients can be solved for relatively easily. After setting the above equations to zero, a system of linear equations is formed with the unknowns being the Lagrange multipliers, λ, and the DFROM coefficients $\hat{a}$. In practice, the resulting linear system was solved through the creation and use of a Python script. Since the linear system has on the order of 200 equations, the efficiency of the SciPy linear algebra solver was sufficient to solve the system [Vir20]. There were no problems with the condition of the constrained system and so no steps were taken to adjust it to improve numerical performance. Each parameter point of interest is a separate optimization problem, which means this system is set up and solved approximately 400 times for a single database.

Once the new DFROM coefficients are determined, the model in equation 3.4 is used the same way as the original ROM but with the new DFROM coefficients. With this new adjustment, the surfaces generated by the new model will integrate to match the more trusted source of wind-tunnel data.

4.1.3 The Choice for Weighting

Before settling on using the eigenvalue from PCA several other weighting terms were explored. Along with the uniform weighting case, three different terms plus their inverses were explored. The three cases were using the eigenvalues directly, the ratio of peak value to the corresponding eigenvalue and finally the peak value alone. The eigenvalues are an obvious choice because they are directly related to the "mathematical importance" of each PC. The ratio of peak value to eigenvalue is derived from the work of Dalle et al., which uses that weighting term for adjustment of supersonic lineloads [Dal18]. The third option uses the peak value as the middle point between the other two. The inverse is also explored to confirm if it is better to have higher or lower variance PCs change more. To test the ideal weighting term strategy to use, the Block 1B Crew vehicle was used and the data-fusion process was done on the core of the vehicle and all six forces and moments were constrained.

4.1.3.1 Uniform Weights

The first implemented version was the trivial case where all weights were equal to one, which as a result, tends to favor changes to PCs that have larger integrated forces and moments. However, a small change to the corresponding ROM coefficient can have a larger effect. This is questionably desirable behavior because while integrated forces and moments tend to decrease as PC number increases, there is no guarantee of that assumption.

4.1.3.2 Eigenvalue Weights

The second method examined was using the eigenvalues corresponding to each PC. This weighting favors changes to the higher number PCs since they have smaller eigenvalues. The outcome of this technique is a monotonic weight function that is mathematically related to the variance of each PC.

Figure 4.1 shows the eigenvalue weights normalized by the sum of the weights within each variable. To clarify, the weights associated with the pressure PCs are all added together and used to normalize the eigenvalues associated with the pressure PCs. The sum used in normalization can be seen in the legend next to the corresponding variable. This normalization is done to improve comparisons between weighting strategies. Because they are all "competing" within a strategy it is more important to pay attention to the trends and relative values as opposed to absolute numbers. The PC number associated with each weight is found on the x-axis. Notice that the sum of the weights for the friction terms are significantly lower than the pressure one. Because of the lack of normalization between pressure and friction terms before the PCA is performed, the friction terms have smaller eigenvalues by several orders of magnitude or more.

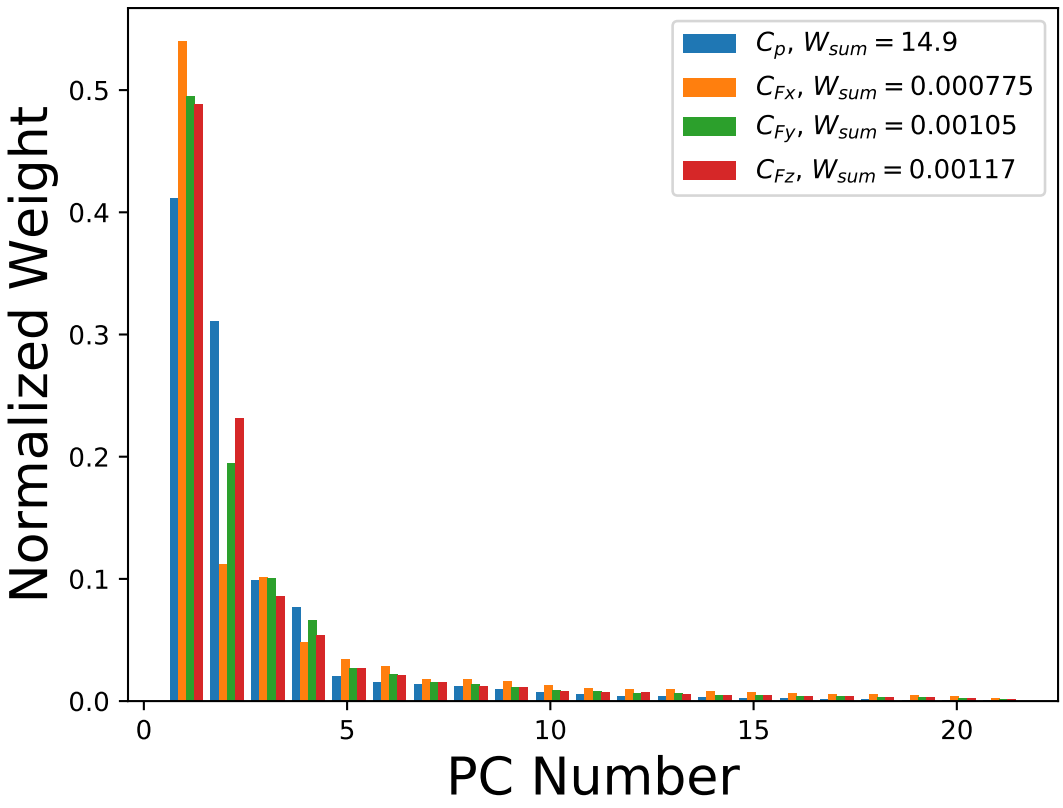


Figure 4.1 Eigenvalue-based weights normalized in variable by their sum (displayed in legend) for each PC.

4.1.3.3 Ratio of Peak Value to Eigenvalue Weights

The next method heavily considered was derived from the work of Dalle et al., which was developed for the adjustment of lineloads in the transonic and supersonic regime instead of the subsonic one. In this regime, PCs can have a number of local peaks due to shocks, which can be a problem when performing adjustments. This weighting was developed to help compensate by penalizing PCs that have one or two large "spikes" in the corresponding lineload. Peak value is defined here as the maximum of the absolute value across the corresponding PC. A bar chart of the normalized weights can be found in Fig. 4.2. Similar to the eigenvalue weight chart above, the normalization shown is the in-variable sum of weights. Unlike the previous eigenvalue weights, the weighting term tends to increase with PC number although not monotonically. Importantly, the friction terms now have two orders of magnitude greater weight sums. This results in a system that favors changing pressure terms over the friction terms. One thing to keep in consideration is that while the method developed by Dalle solves a similar constrained optimization system, the form of the cost function is different in that it simply has $\hat{a}^2$ and not $(a-\hat{a})^2$. The full results of this difference have not been explored.

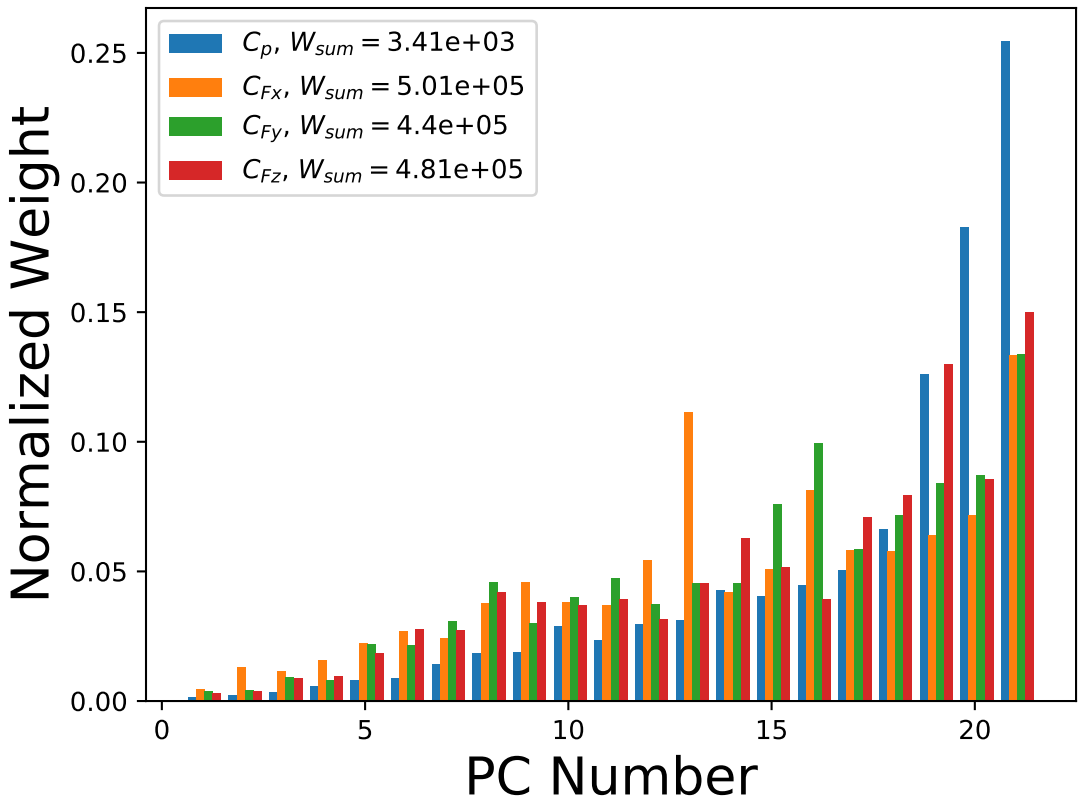


Figure 4.2 Ratio of peak value to eigenvalue based weights normalized in variable by their sum (displayed in legend) for each PC.

4.1.3.4 Peak Value Weights

Because of the introduction of the peak value term in the previous methodology, it was of interest to explore what the effects that using the peak value itself would cause. Figure 4.3 shows the values of the normalized weights of this method. The figure shows the weights being more evenly distributed than either of the previous two methods while maintaining similar trends to the eigenvalue weighting. Like the ratio weighting, the trends are monotonic with a few poke outs like PC 13 for the friction component in orange. Overall, it is expected that this weighting strategy would perform very similar to the eigenvalue weighting strategy.

4.1.3.5 Inverse Weights

To finish the examination of weights, the previous three weights were inverted. While the different weightings as presented test whether it is better to have high weights for the low numbered PCs or high numbered PCs, there is some confounding variables. This inversion is to help control for that and settle the issue.

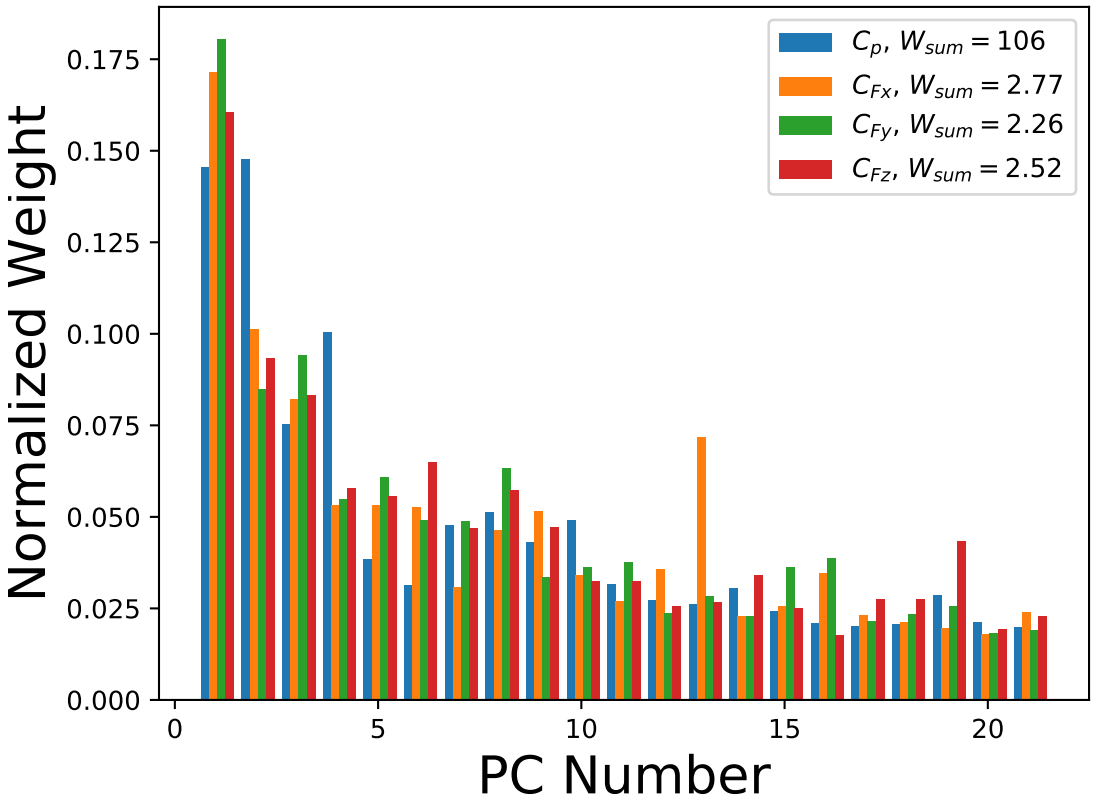


Figure 4.3 Peak value weights normalized within variable by their sum (displayed in legend) for each PC.

4.1.3.6 Results of Weight Variation

Table 4.1 summarizes the previous proposed weighting terms. It takes note of whether the weights are increasing or decreasing with PC number as well as the sum of the weights for each variable. The inverted versions are left off. For many cases, especially when the ROM predicts values very close to the desired forces and moments, the choice on weighting term has little effect. However, on other cases, the results can be significantly different pointing to the importance of carefully choosing the weighting term.

The lineloads shown in Fig. 4.4 are a good example of where weighting plays a large role. The figure shows data fusion results using all weighting strategies explored in lineload form. The top shows the axial followed by the side force and finishing with the normal force lineload on the bottom. The inverted weights are dashed to help with distinction. The region of most interest for this case is near the radius change at about x/L = 4.5 and is broken out in Fig. 4.5. Here we see a disqualifying result for many of the weighting strategies. The sectional axial force goes positive for many of the options, which is not a physically realistic outcome. If the desire of the data fusion process is to have higher fidelity answers, then it would be a poor idea to choose one of these options.

The options that can be excluded by examining this case are all systems that have higher weight values for the low variance PCs. This shows that it is best to minimize the change to the ROM coefficients related to the lower PC numbers. Remember one of the goals for the data fusion process is to keep the lineload as close to the original ROM as possible since it is the best source of spatially distributed data available. Figure 4.6 shows the remaining weighting options and the original ROM, in black, for the same case as above. While the lineload for the side force coefficient shows all remaining strategies presenting about equally valid results, the normal force coefficient lineload shows clearly that the eigenvalue weighting strategy remains closest to the original ROM lineload. With this case in mind, eigenvalue weighting is clearly the preferred weighting strategy for this data fusion process in this regime.

4.2 Results

Before going in depth on the results, a difference in how this methodology is applied to the two configurations needs to be made explicit. The centerbody for the Block 1B cargo is much more sensitive to asymmetric vortices due to its cleaner geometry. As a result, the side force and yawing moment are not used as constraints except at α_P = 90° where the asymmetric vortices are not present. When used as a constraint, the DFROM predicted a large lateral loading on the nose of the vehicle even when the ROM and the constraining forces and moments predicted the same vortex state. The SRBs still use all six force

Table 4.1 Summary of the weighting strategies used in the developed DFROM.

Weighting Strategy	Change with PC Number	C_pW_{sum}	$C_{F_x}W_{sum}$	$C_{F_y}W_{sum}$	$C_{F_z}W_{sum}$
Uniform	Constant	21	21	21	21
Eigenvalue	Decrease	14.9	.000775	0.00105	0.00117
Ratio of Peak Value to Eigenvalue	Increase	3.41e3	5.01e5	4.4e5	4.81e5
Peak Value	Decrease	106	2.77	2.26	2.52

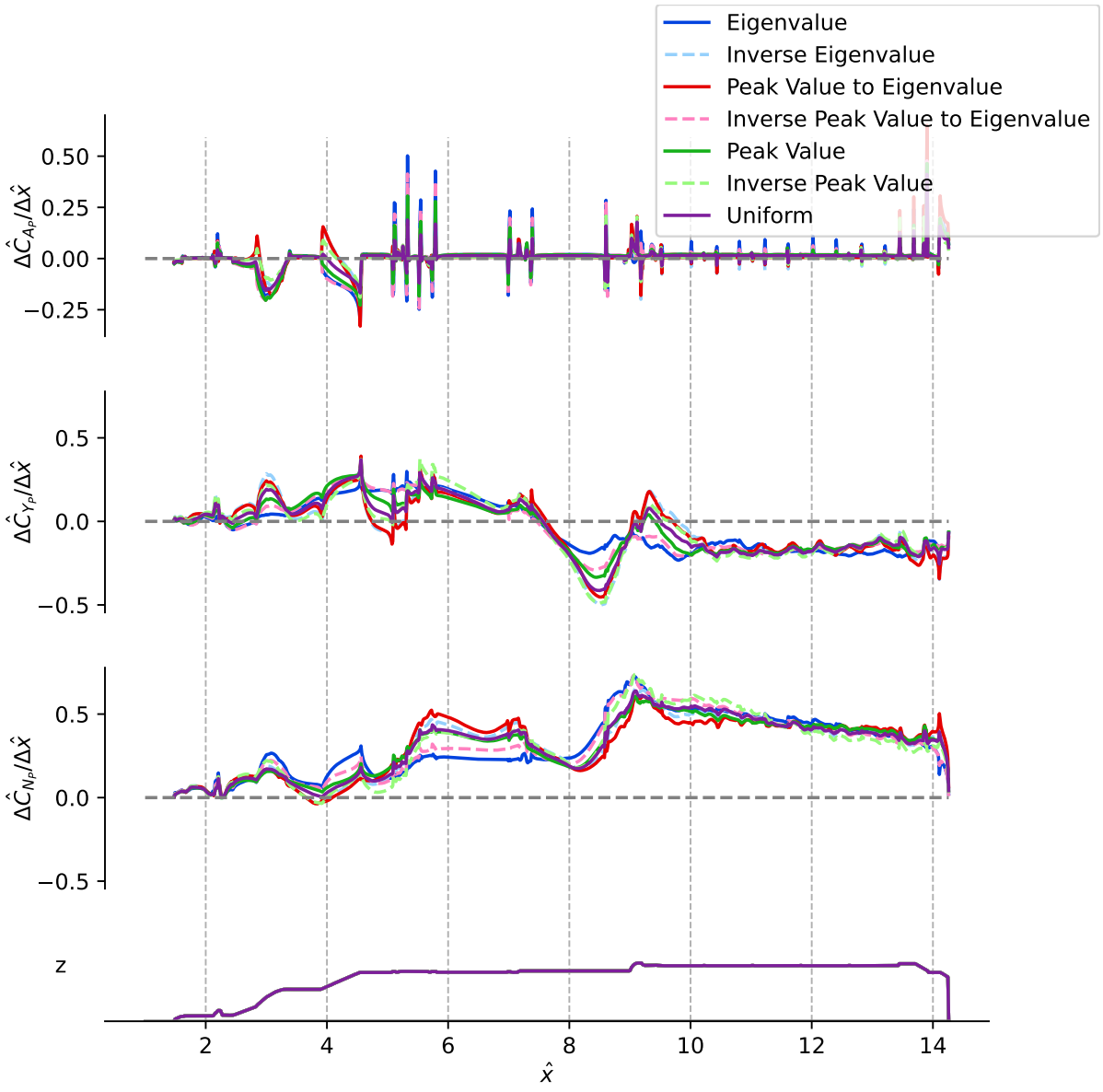


Figure 4.4 Normalized lineloads at α_P=35°, ϕ_P=0° showing results of data fusion using different weighting terms.

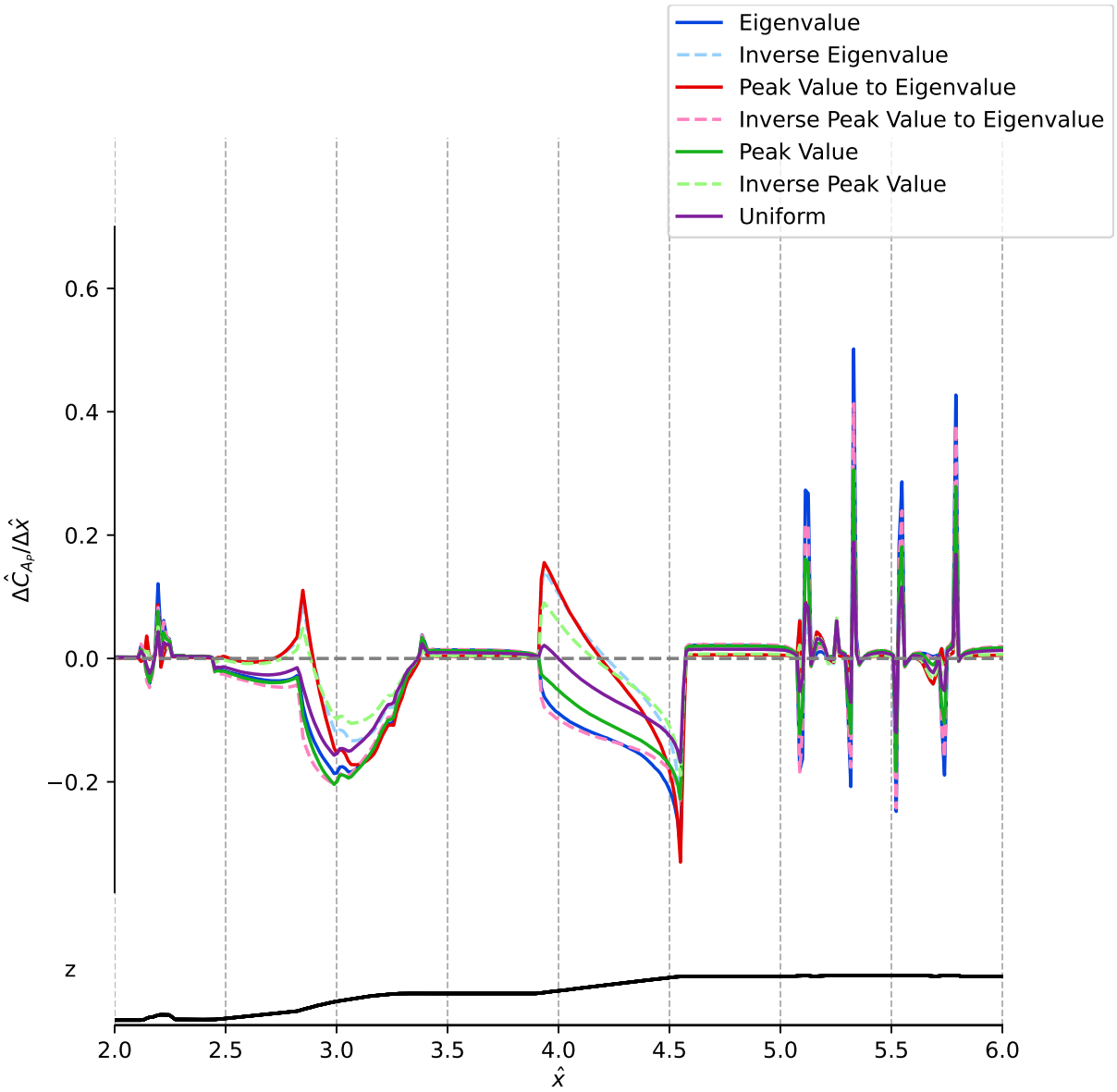


Figure 4.5 Normalized lineloads at α_P=35°, ϕ_P=0° showing results of data fusion using different weighting terms. Focusing on the axial lineload near the front of the vehicle.

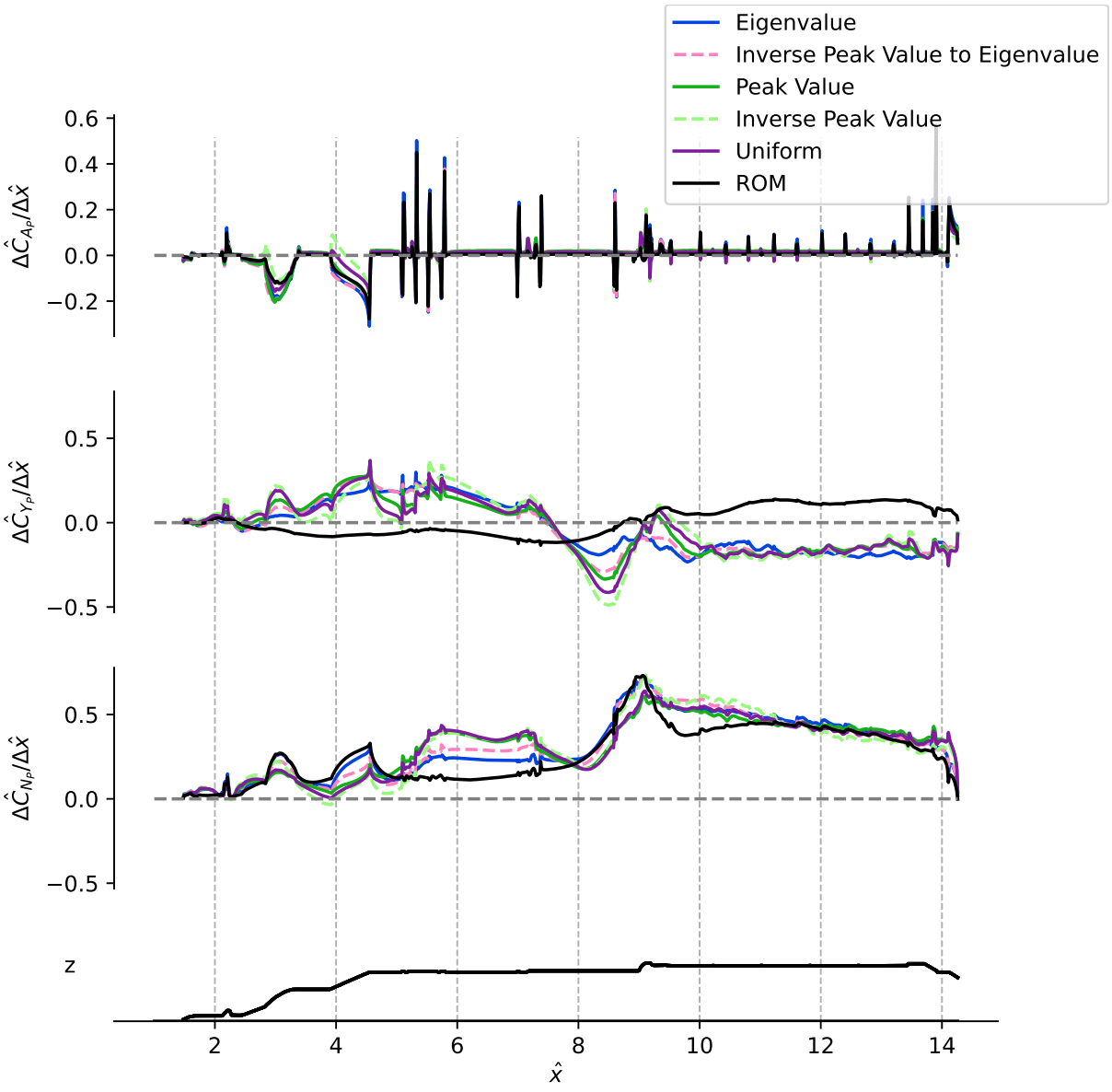


Figure 4.6 Normalized lineloads at α_P=35°, ϕ_P=0° showing results of data fusion using different weighting terms compared to original ROM.

and moment constraints since they are not as strongly affected by the phenomenon.

Since the DFROM works by adjusting the ROM coefficients, it is important to see the changes that they undergo. A side by side comparison of the ROM and DFROM coefficients for the first pressure coefficient of the Block 1B cargo vehicle can be seen in Fig. 4.7. The leftmost graph is the original ROM coefficient as seen before and the center is the new DFROM coefficient. These plots use the same contour scale. The third plot (on the right) is the difference between the two ROM coefficients. Because the weighting term used in the adjustment is equivalent to the eigenvalue associated with the first PC, it is expected that this variable changes the least, which is confirmed in these plots. The values of this change tend to increase further from the CFD points, which reflects the increased inaccuracies in predictions further from the source data.

In the more extreme case, the coefficients associated with the friction components change a significant amount due to their much lower eigenvalues. This is presented in Fig. 4.8, which plots the coefficient values for the first PC associated with the Z direction of friction across the domain for the Block 1B cargo vehicle. While some of the friction coefficient values have changed significantly more than initially expected, they prevent large changes in the pressure, which has a larger effect on final lineloads and one of the goals is to minimize the change in the overall shape of the lineload between the ROM and DFROM. Consequently, the increase in viscous forces has been deemed an acceptable side effect.

As a quick check to confirm that the DFROM coefficients are creating surfaces with the desired values, the forces and moments can be plotted against the wind tunnel values. The forces and moments can be seen in Fig. 4.9 and Fig. 4.10, which show the same conditions as were shown in the ROM section. Again, while the rolling moment coefficient plot in Fig. 4.9 looks extreme, these values are relatively close to zero. The normal and pitching moment for the DFROM now match with the wind-tunnel-derived data for both configurations. Since the other four coefficients are allowed to drift for the Block 1B cargo configuration, they have slight changes, but overall they keep their values close to the ROM prediction

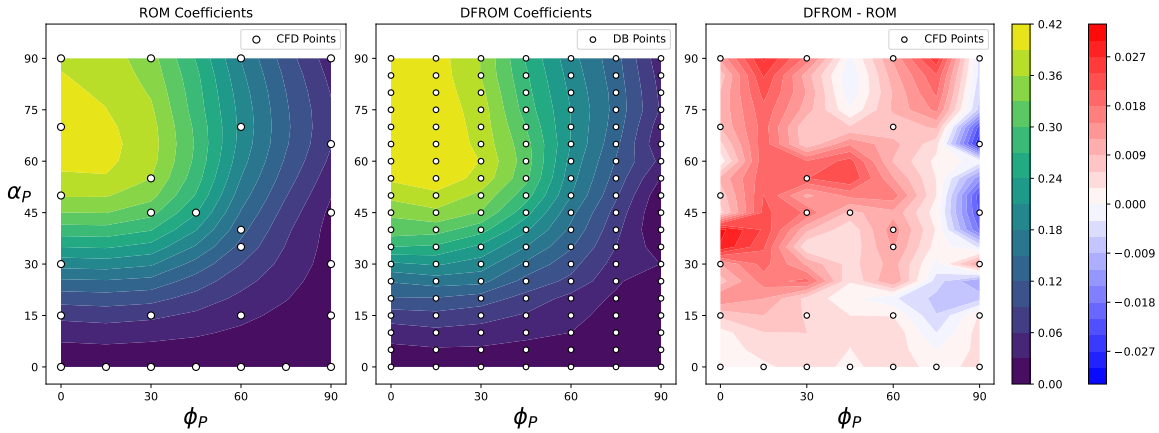


Figure 4.7 Comparison of ROM and DFROM coefficients for first PC related to pressure for the core of the Block 1B crew vehicle.

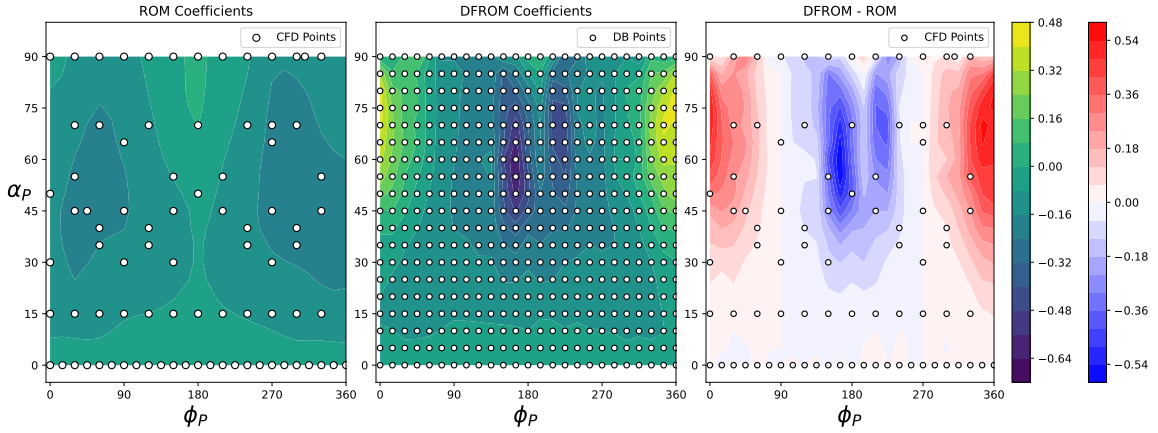


Figure 4.8 Comparison of ROM and DFROM coefficients for first PC related to friction in the Z direction for the core of the Block 1B cargo vehicle.

as seen in Fig. 4.10

To show how the adjustment changes the surface prediction, two sets of three figures are used to show the ROM, DFROM and the difference between the two for a single condition. For the first set, an example from the Block 1B cargo vehicle at $\alpha_P = 75°$ and $\phi_P = 270°$ are shown. The first figure, Fig. 4.11, depicts the pressure distribution output by the ROM at α_P of 75° and a ϕ_P of 210°. The second figure, Fig. 4.12 is the equivalent DFROM surface, with the pressure delta (DFROM minus ROM) in Fig. 4.13. There is a slight decrease in pressure on the upper surface and a slight increase on the lower surface of the core, which may not look significant, but can significantly change the lineloads. On the SRBs, a much larger change can be seen near each nose, which is not unexpected because this is a region where the CFD has trouble. The effect of the slight change in surface pressure along the core can be analyzed in lineload form presented in Fig. 4.14. Here the ROM lineloads are in black, while the DFROM lineloads are in blue. The major difference is in the forebody, which points to significant underprediction in pitching moment as well as normal force in the ROM that is now corrected. This holds true to the general trend of the ROM underpredicting these two coefficients. The adjustments made have a slight effect on the side force sectional loads, which is expected from the constraints implemented for this vehicle.

The second set shows a more extreme change, which can partially be contributed to the asymmetric vortices. The next three figures show the Block 1B crew configuration at $\alpha_P = 40°$ and $\phi_P = 60°$. As before, the first figure, Fig. 4.15, shows the output of the ROM. Nothing is anomalous here even when comparing to Fig. 4.16, which shows the DFROM results. However, the difference shown in Fig. 4.17 reveals significant areas of the pressure have been changed in order to meet the constraints. The good news is that when looking at the lineload level, Fig. 4.18, the adjustments are not as drastic. Most of the adjustments on the core reflect a change in the side force. Here the ROM calculated a vortex on one side while the database shows a side force in the other direction. As the ROM and DFROM do not have a way to compensate for nonlinear effects, it creates these types of corrections. This type of correction exemplifies the strength of the DFROM adjustment method selected by demonstrating how even sign

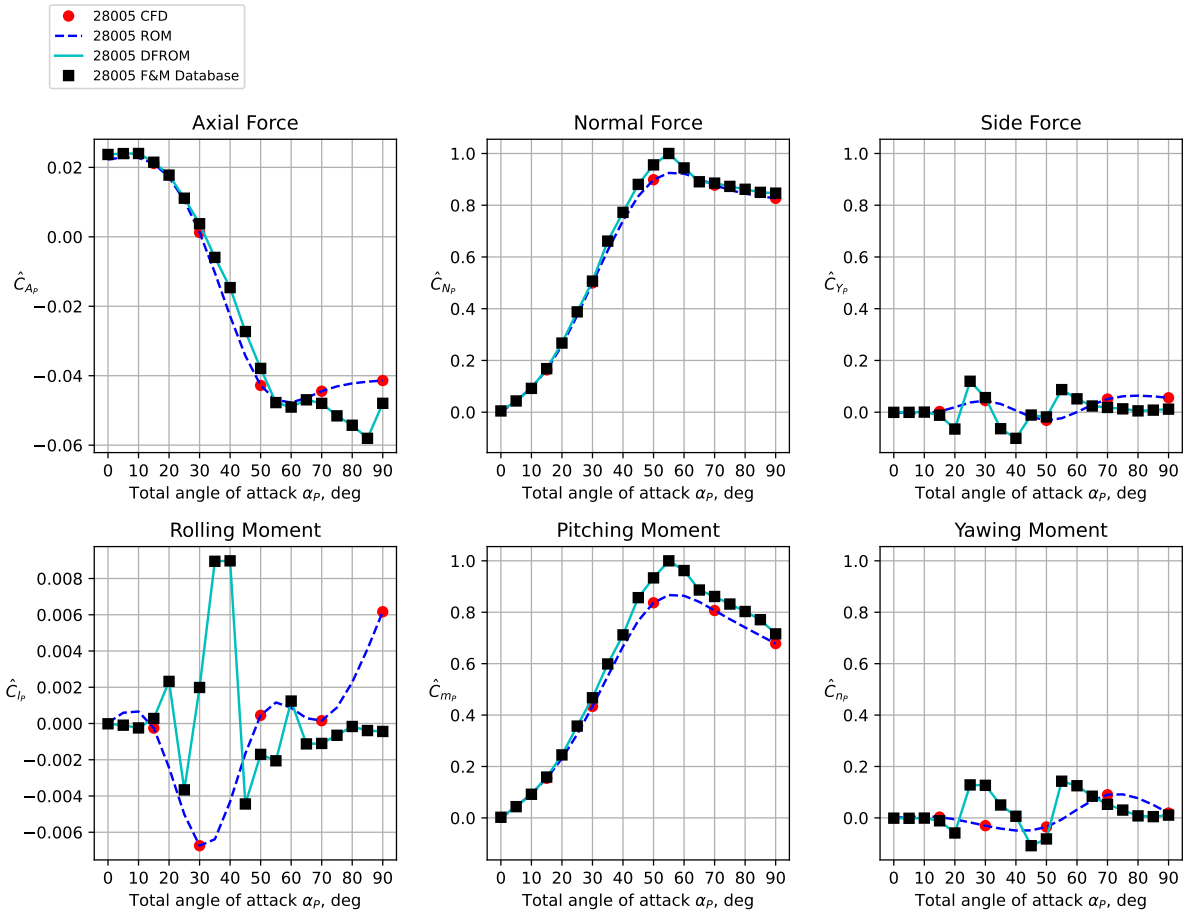


Figure 4.9 Comparison of ROM force and moments to DFROM versus angle of attack at $\phi_P = 0°$ for the Block 1B crew vehicle.

changes do not disrupt the basic structure of the lineload.

To show that the DFROM improves the predictive capability and robustness of the ROM, the LOO analysis was repeated. The same points as before were examined. The new dropped point DFROM was compared to the DFROM at the CFD condition that was dropped to examine its sensitivity to missing information. The first, Fig. 4.19, is the Block 1B crew vehicle at $\alpha_P = 50°$, $\phi_P = 0°$. The DFROM is the blue line and the dropped point DFROM is the red. The most significant change compared to the previous LOO analysis occurs at the aft end. Previously, there was a significant difference in the lineloads near the aft end particularly in sectional side force, which showed up as an offset. The sectional normal force is also more robust with a much smaller difference between the two DFROMs.

The second point, Fig. 4.20, is the Block 1B cargo vehicle at $\alpha_P = 45°$, $\phi_P = 270°$. This looks almost identical to the LOO analysis done on the ROM. The differences are subtle but there is some movement in the lineloads near the aft end. This case points out one of the weaknesses in the adjustment method and difficulty in adjusting lineloads in general, which is that there are an infinite number of lines that integrate to any desired value. As before, these two sets of lineloads integrate to similar forces, which

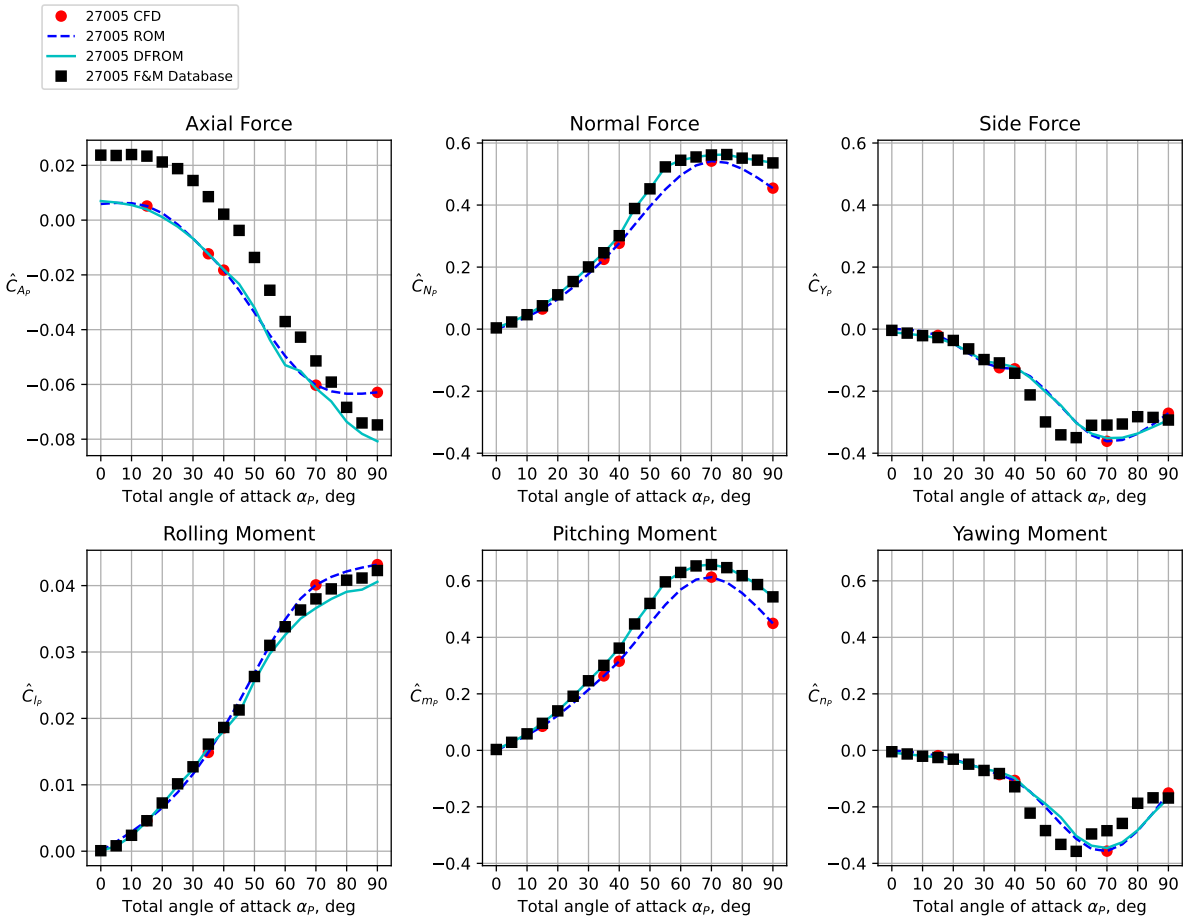


Figure 4.10 Comparison of ROM force and moments to DFROM versus angle of attack at $\phi_P = 120°$ for the Block 1B cargo vehicle.

further reinforces this point. While the DFROM has the ability to help fill in some of the holes in the initial dataset, it lacks the ability to fill in "potholes".

The DFROM provides lineload results that are similar in shape and characteristic to the initial ROM output; however, they now integrate to the desired values. The DFROM output surfaces have similar characteristics to the ROM surfaces, which means they have similarities to the original CFD data. These DFROM lineloads tend to be a bit more robust to missing data but still not able to function without those essential data.

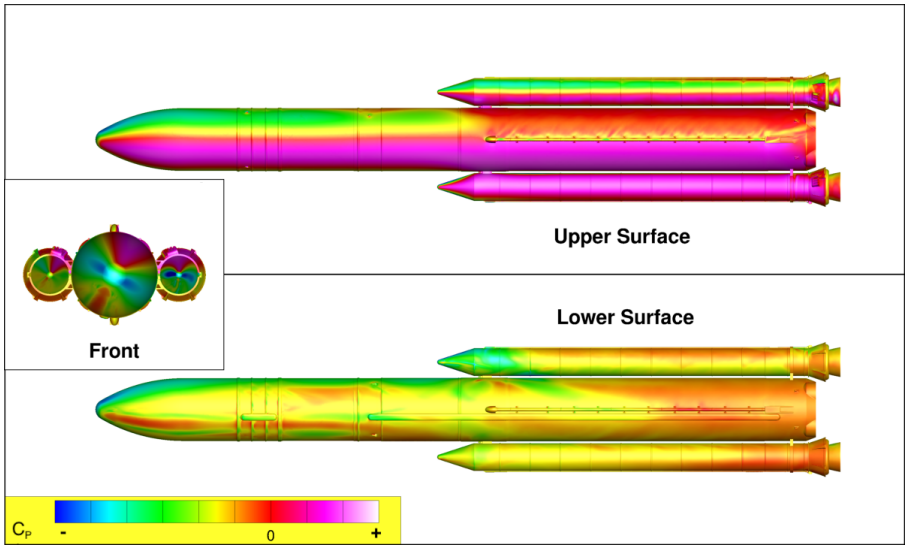


Figure 4.11 ROM generated surface pressure distribution before adjustments for Block 1B cargo configuration at $\alpha_P = 75°$, $\phi_P = 210°$. Plotting template provided by Steve Krist.

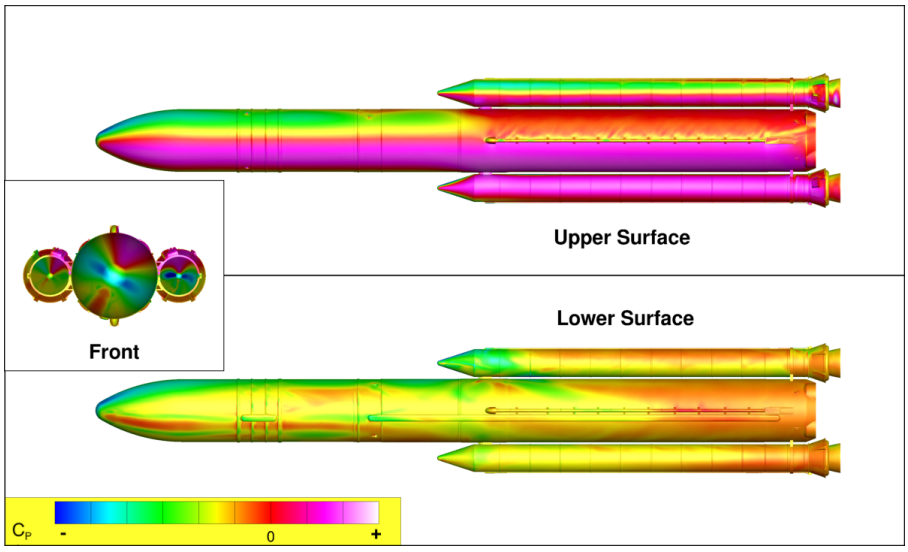


Figure 4.12 DFROM generated surface pressure distribution after adjustments for Block 1B cargo configuration at $\alpha_P = 75°$, $\phi_P = 210°$. Plotting template provided by Steve Krist.

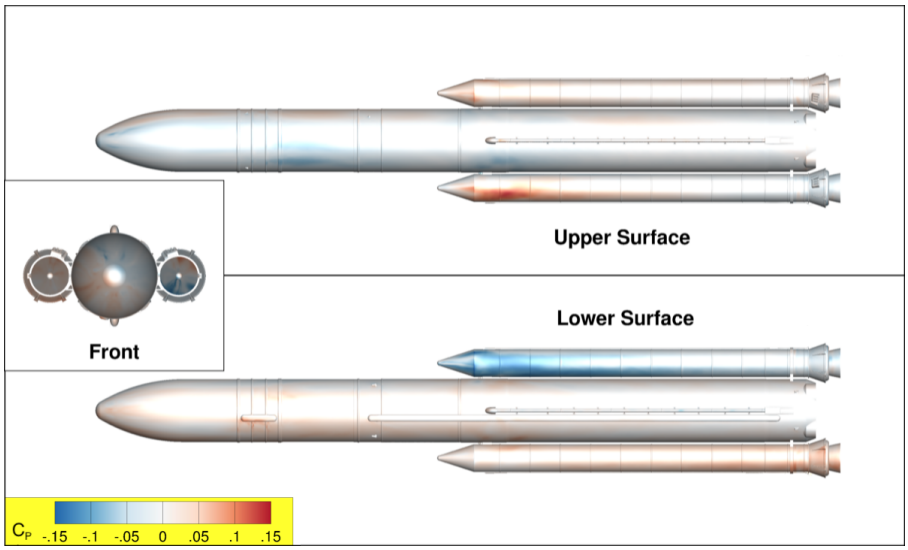


Figure 4.13 Change in generated surface pressure distribution after adjustments for Block 1B cargo configuration at α_P = 75°, ϕ_P = 210°. Plotting template provided by Steve Krist.

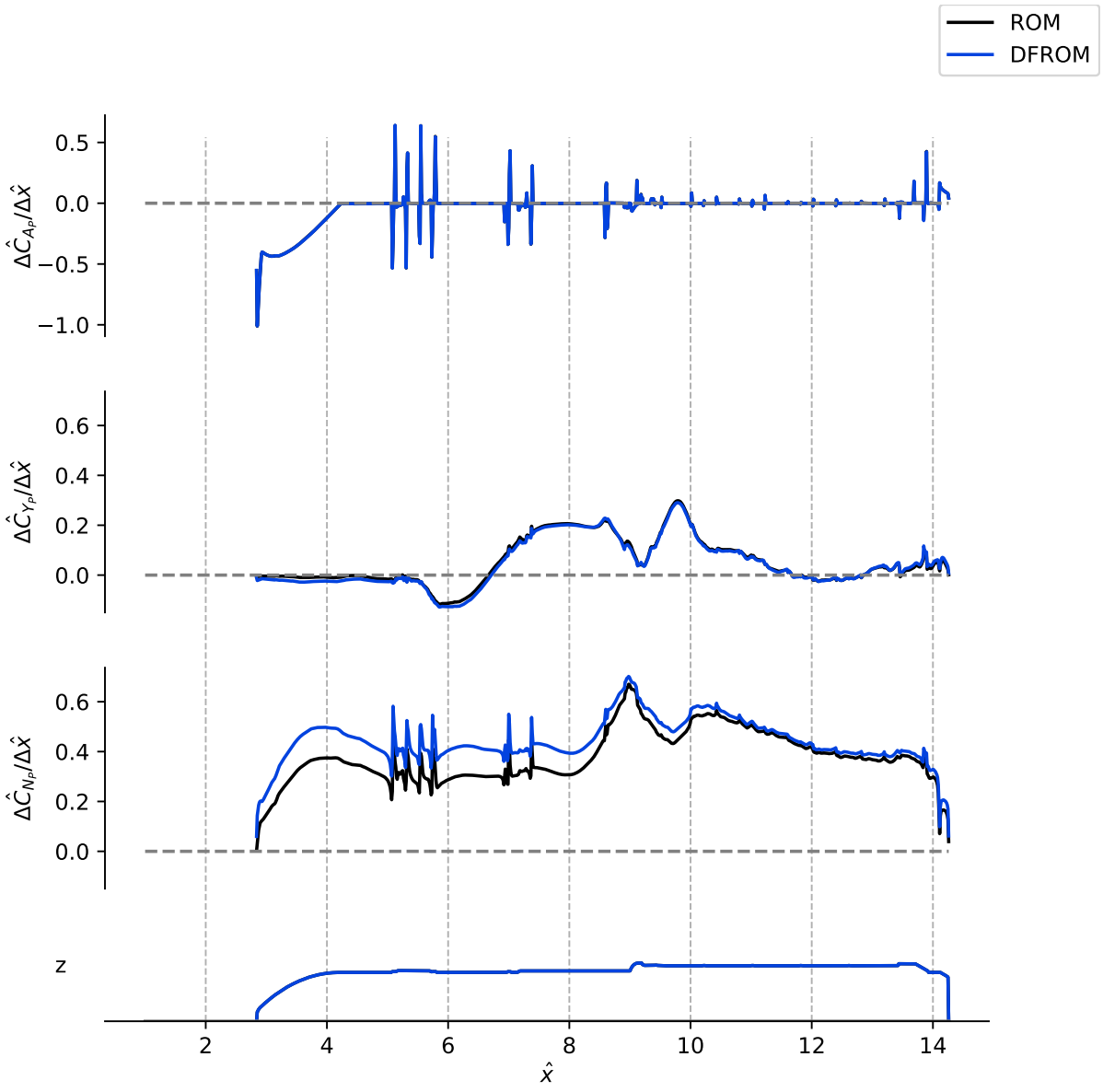


Figure 4.14 Comparison between ROM and DFROM lineloads for the core of the Block 1B cargo configuration at $\alpha_P = 75°$, $\phi_P = 210°$.

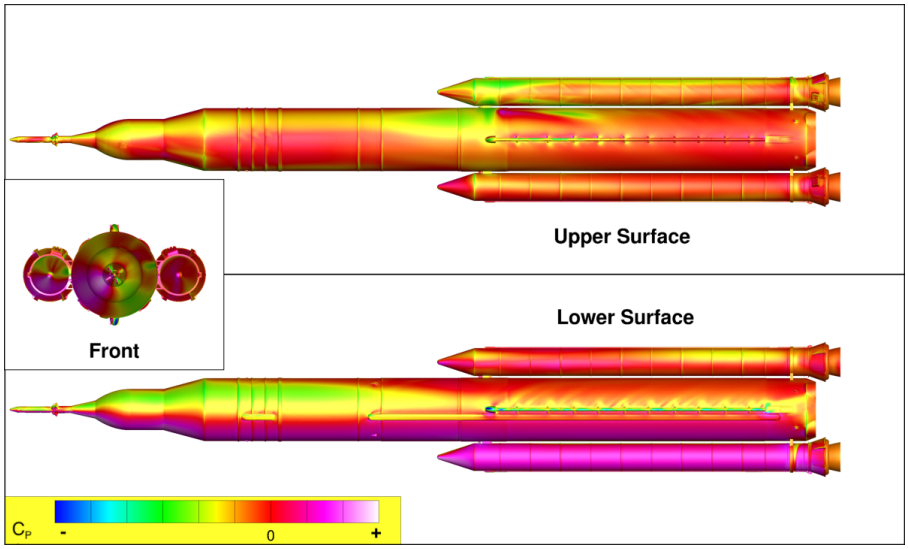


Figure 4.15 ROM generated surface pressure distribution before adjustments for Block 1B crew configuration at $\alpha_P = 40°$, $\phi_P = 60°$. Plotting template provided by Steve Krist.

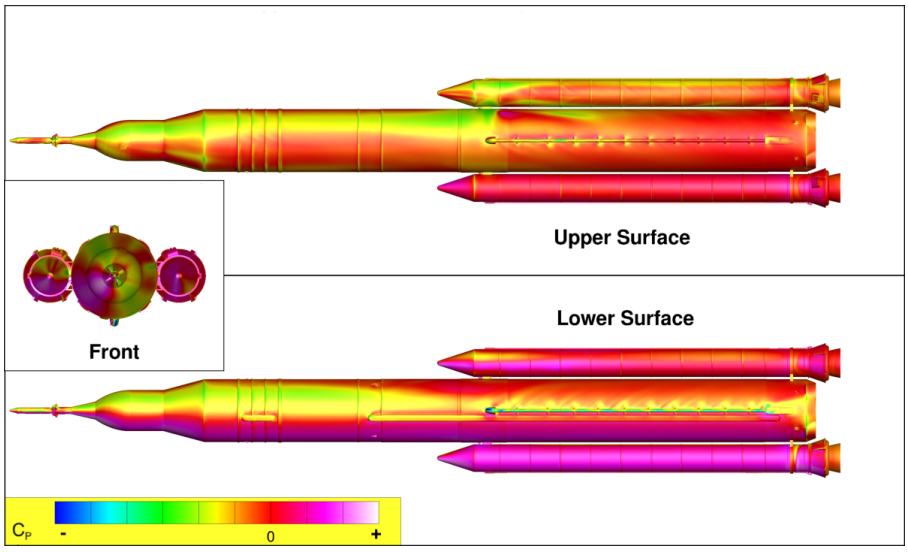


Figure 4.16 DFROM generated surface pressure distribution after adjustments for Block 1B crew configuration at $\alpha_P = 40°$, $\phi_P = 60°$. Plotting template provided by Steve Krist.

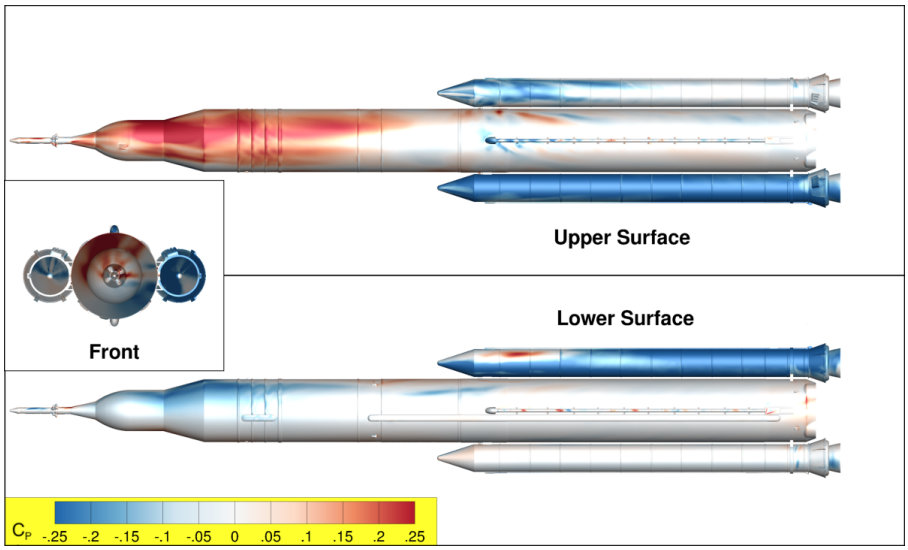


Figure 4.17 Change in generated surface pressure distribution after adjustments for Block 1B crew configuration at $\alpha_P = 40°$, $\phi_P = 60°$. Plotting template provided by Steve Krist.

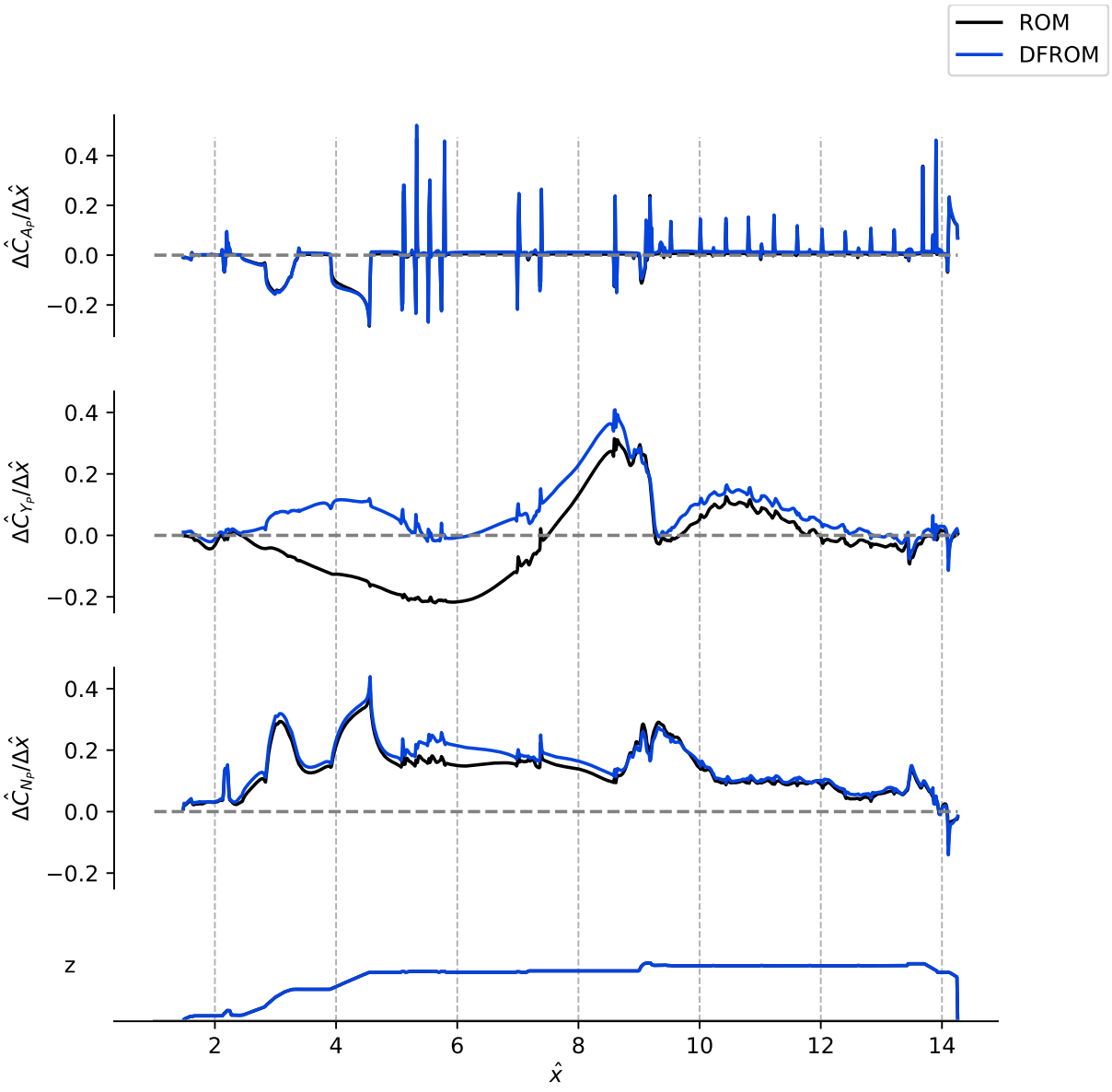


Figure 4.18 Comparison between ROM and DFROM lineloads for the core of the Block 1B crew configuration at $\alpha_P = 40°$, $\phi_P = 60°$.

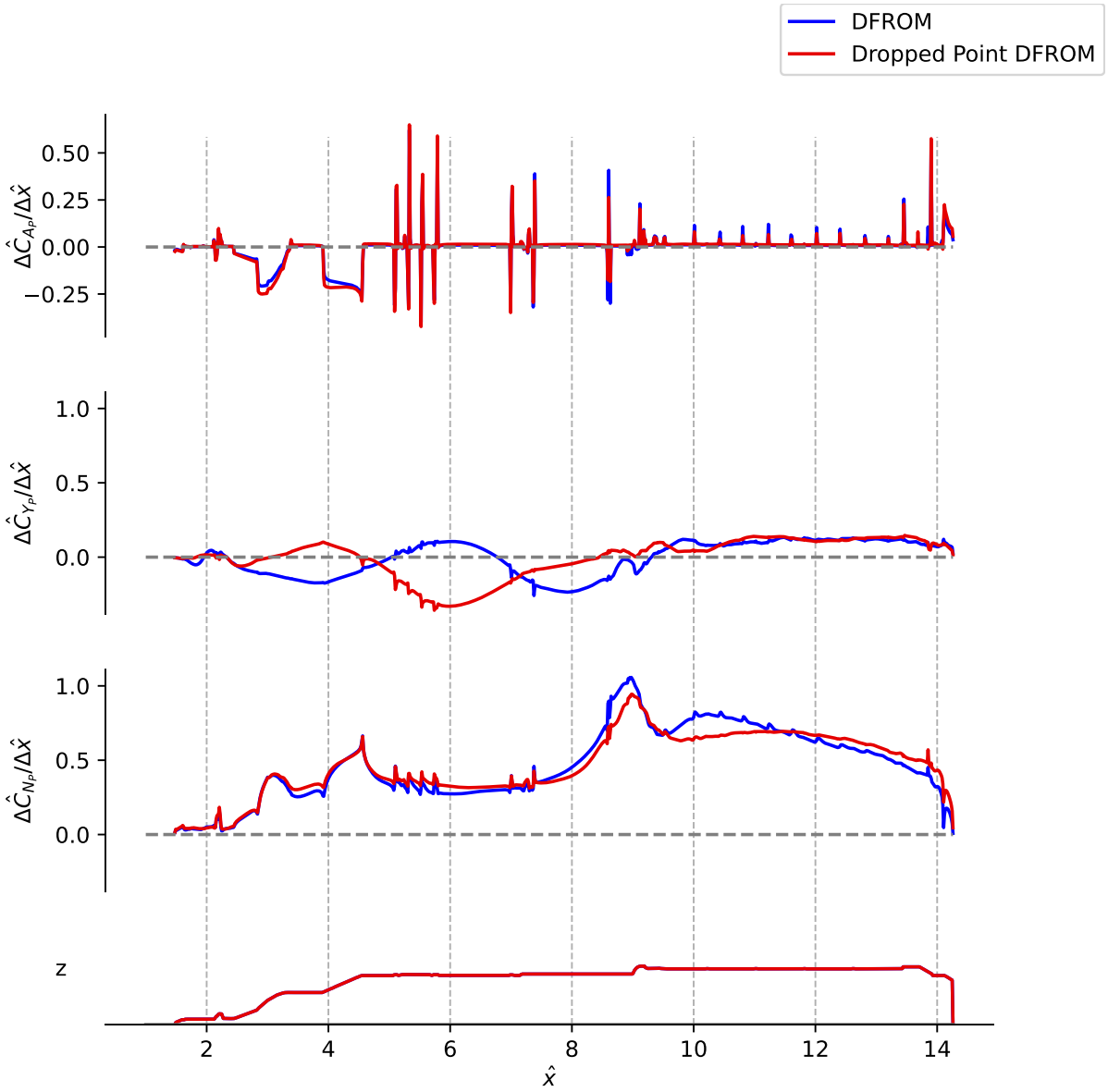


Figure 4.19 LOO DFROM lineloads for the core of the Block 1B crew configuration at $\alpha_P = 50°$, $\phi_P = 0°$.

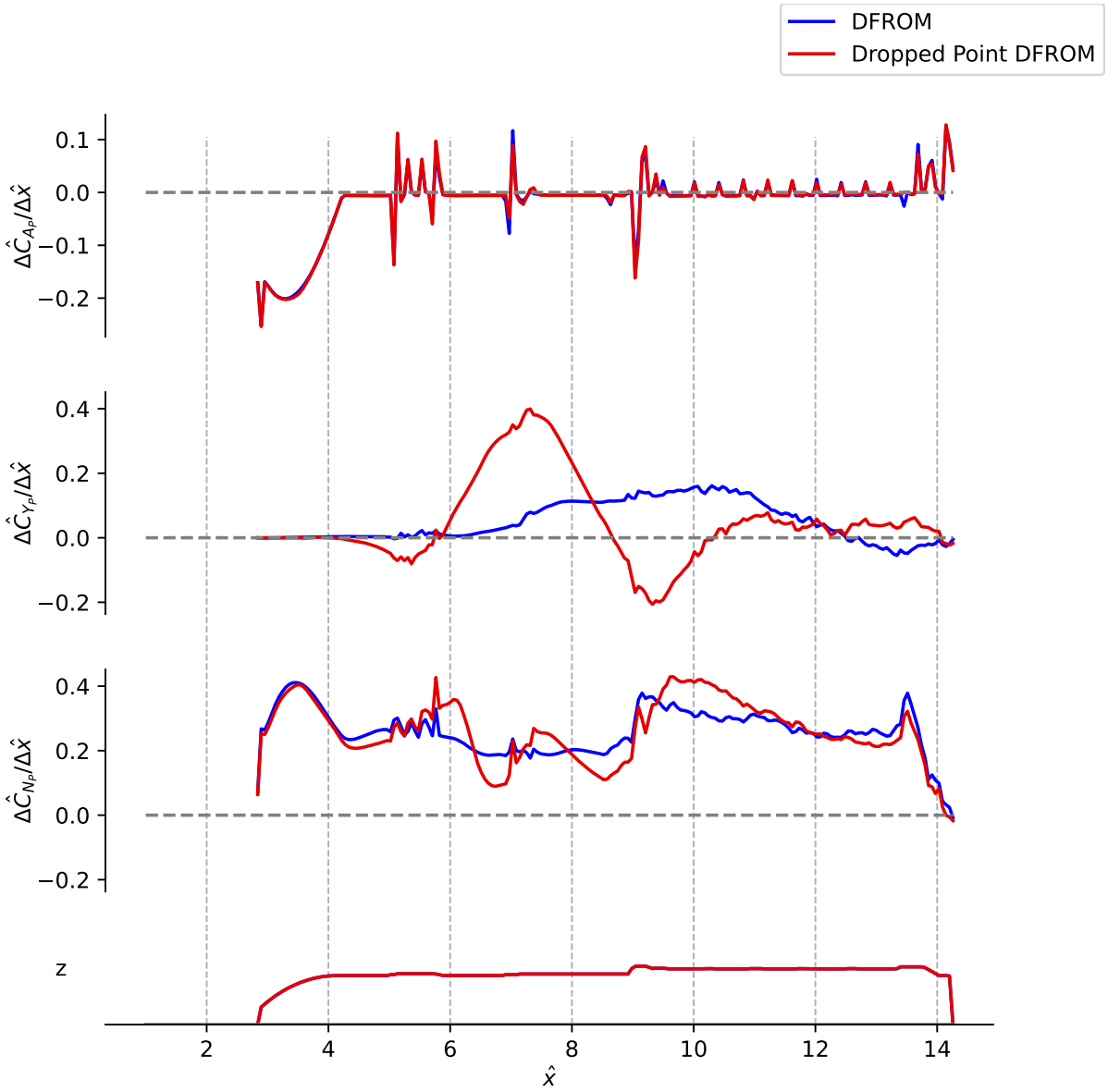


Figure 4.20 LOO DFROM lineloads for the core of the Block 1B cargo configuration at $\alpha_P = 45°$, $\phi_P = 270°$.

CHAPTER

5

UNCERTAINTY QUANTIFICATION

As model complexity increases, researchers have sought to understand the limitation of such models and explain the uncertainty associated with them. In particular, looking at distributed loading as seen in lineload databases is especially difficult. This problem has only recently been addressed by Dalle et al. in Ref. [Dal18]. One of the difficulties in modeling uncertainty in aerodynamic lineloads is the lack of work that focuses on spatially distributed outputs. While in many cases, the traditional methods and techniques can be used, there is always increased difficulty as the number of dimensions increase. The unique characteristic of spatially distributed outputs is the need to incorporate realistic variations in spatial derivatives in the final output. While an incremental increase can test some design constraints, it is still important to capture the effects of different "shapes". For example, testing for shear limits is affected more by the difference in forces between two sections, whereas net forces are more relevant to bending. Capturing such a high level of detail in the uncertainty is difficult and requires a complex uncertainty model. Because of the difficulty of the problem at hand some simplifications to the problem space have been made.

The work described here focuses on the normal force lineloads of the Block 1B Cargo configuration specifically the centerbody. The normal force is the focus because it is less influenced by the effects of the asymmetric vortices. There are two different uncertainty models generated. The first is a more traditional "increment-based" uncertainty quantification (UQ) model and there is no attempt to capture the spatial variations. The increment based model breaks up the uncertainty into two portions. The first form of uncertainty examines the model form error, that is error that is associated with the limitations with the model under consideration, which would not have been possible without the assistance of Heather Houlden. The second form is related to the uncertainty in source data used in the data fusion process,

which is brought in by exploring how the uncertainty in the original force and moment databases are transferred to the lineload database. These two sources of uncertainty are then combined through a simple root sum square addition.

The second UQ model focus on using bootstrapping in order to develop a distribution on the values of the ROM coefficients. After fitting a Laplace distribution, randomly drawn ROM coefficient values are used to generate forces and lineloads. After that the variations are than propagated through the data fusion method.

5.1 Incorporating Model Form Error

Model form error is a function of both the choice of model and how well the model predicts new data. One measure of the model accuracy can be captured by looking at model predictions at a set of validation and test points. The other part of model form error can be captured by comparing different types of reasonably effective models. Where these models disagree, a higher level of uncertainty is needed. To begin with a "leave-one-out" analysis similar to cross-validation and bootstrapping was performed on the DFROM [Koh95].

A study was conducted to assess the sensitivity of the DFROM to the input, or training, data. A "leave-one-out" approach was used to compute an alternate lineload database with one of the CFD input points omitted. Subsequently, 46 sets of alternate lineloads were computed, with a different CFD solution omitted for each DFROM. The differences between the nominal and each alternate "leave-one-out" lineload were computed. These residuals were computed at every total angle of attack and roll angle in the database for the axial, normal, and side force sectional loads. This resulted in a total of 10,488 alternate lineloads for each force coefficient, generating an enormous amount of data. A three-dimensional surface plot of the normal force leave-one-out residuals is presented in Figure 5.1, which was developed thanks to Heather Houlden, to illustrate the size and density of these data. In this figure, the axis labeled "Lineload Index" is numbered from 1 to 10,488, corresponding to the number of residual lineloads. A detailed analysis of these residuals was conducted in an attempt to quantify the model form error, to identify regions of the database space where the lineloads were most sensitive to input data, and to assess the sufficiency of the quantity and distribution of the CFD solutions selected as training points for the model. This leave-one-out sensitivity analysis enables a characterization of possible lineload values from which a prediction interval can be derived [Smi13].

The leave-one-out residuals were a function of X/L_{Ref}, total angle of attack, roll angle, and the omitted CFD solution. Initial exploration of this substantial collection of data consisted of examining the maximum variation, due to omitting an input point to the data fusion model, in the load distributions at each breakpoint. An example of this is shown in Figure 5.2 for a total angle of attack of $50°$ and roll angle of $120°$.

With 10,488 lineloads computed for each force coefficient and each load distribution consisting of 200 X/L_{Ref} coordinates, it quickly became necessary to condense and/or pool the data in order to make it manageable. Thus, it was decided to begin with visually inspecting the residuals to identify trends or

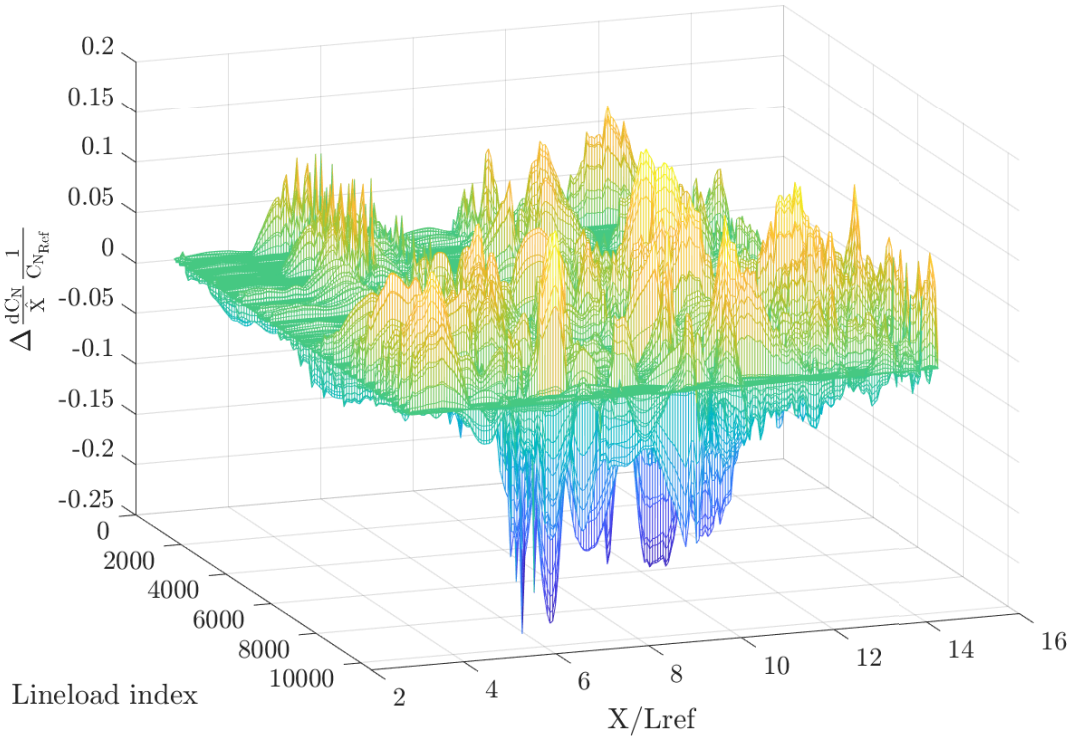


Figure 5.1 Carpet plot of $dC_N/\hat{X}$ leave-one-out residuals.

patterns. These data were visualized by generating a three-dimensional contour plot of the residuals for each leave-one-out case (each CFD solution omitted from the DFROM). Figure 5.3, again developed by Heather Houlden, presents several of these contour plots. The x-axis represents the body coordinate of the sectional load (X/L_{Ref}), the y-axis represents α_P, the z-axis represents ϕ_P, and the actual residuals are represented with colored contour lines. The filled black circle in each plot indicates the pitch and roll angle of the CFD training point that was omitted from that particular solution. Note that there are few, if any, contour lines visible above or below the roll angles indicated by the black dots. This is because residual values close to zero are represented by the color white. This indicates that the effects of omitting an input point are fairly localized in the data-fused model. This suggests that a sufficient number of CFD points were selected to run for this database, and they were fairly well placed within the data space. Additionally, Figs. 5.3b and 5.3c show that omitting training data at roll angles of 150° and 60° appears to have less impact on the DFROM. There also does not appear to be a readily identifiable pattern or trend in the signs of the residuals along the length of the vehicle. In other words, the contours in Fig. 5.3 appear to randomly vary between red and blue when viewed along the x-axis for any leave-one-out case. Also of note is that most of these residuals show limited influence across roll angles, which points to a possible division of the uncertainty.

For reference, the leave-one-out residuals are described symbolically in Eq 5.1, where *i* represents the

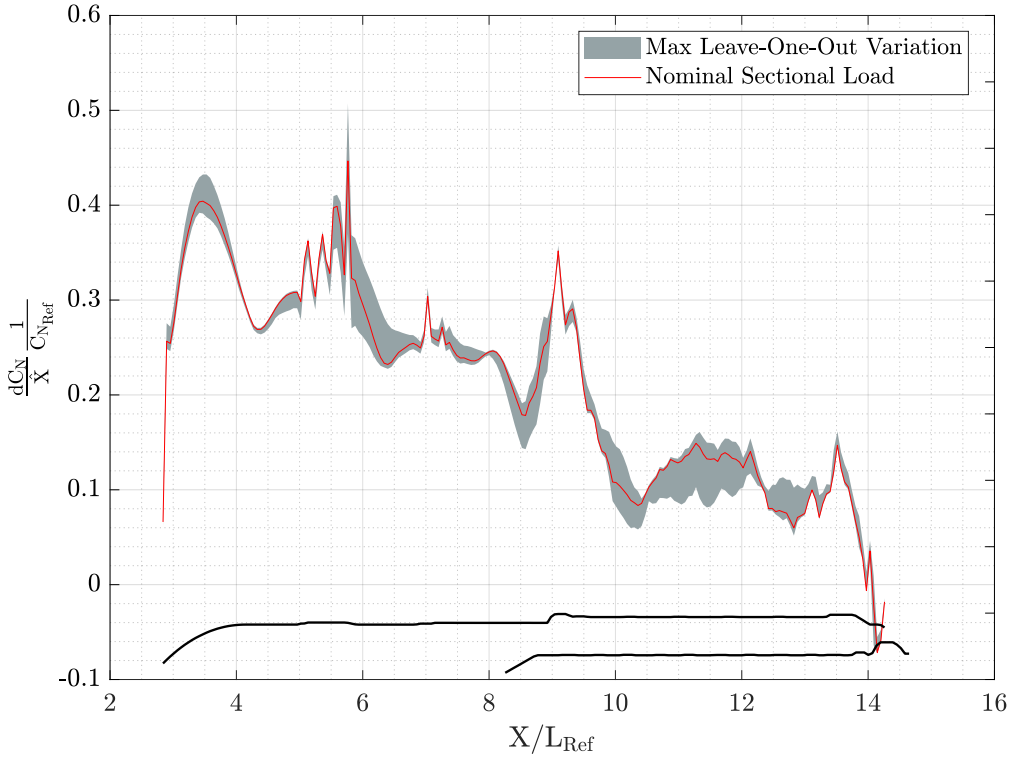


Figure 5.2 Nominal lineload for $\alpha_p = 50°, \phi_p = 120°$, enveloped by maximum variations from leave-one-out analysis.

index of the local load along the body, j represents the number of α_{P} database breakpoints, k represents the number of ϕ_{P} database breakpoints, and l represents the number of leave-one-out cases run (l =1 to 46).

$$\Delta y_{i,j,k,l} = y(X/L_{Ref}, \alpha_P, \phi_P, LOO_{\alpha,\phi}) - y_{DB}(X/L_{Ref}, \alpha_P, \phi_P) \tag{5.1}$$

The decision was made to model the uncertainty as a function of X/L_{ref} and ϕ_{P}, by pooling the residual data across α_{P}. The total model form error was then estimated by combining two levels of variation. One level is the variation in model predictions at individual points in the database due solely to omitting training data from the DFROM. This is similar to a within-group error, as the only source of variation accounted for so far is the 46 leave-one-out differences at each point in the database. The other level accounts for the fact that the overall variation can change across the physical space of the database (i.e., different roll angles). A discussion of combining multiple levels of variation in experimental aerodynamic data is presented in Ref. [Hemaa].

The standard deviation of the variation in the DFROM results due to omitting training data was computed at every point in the database (see Eq. 5.2).

$$\sigma_{\mathrm{LOO}}(\mathrm{X/L_{Ref}}, \alpha_{\mathrm{P}}, \phi_{\mathrm{P}}) = \sqrt{\frac{1}{\mathrm{N}-1}\sum_{\mathrm{l}=1}^{\mathrm{N}}(\Delta \mathrm{y_{i,j,k,l}} - \Delta\bar{\mathrm{y}}_{\mathrm{i,j,k}})^2}, \tag{5.2}$$

where N is the number of alternate models (46 in this case). The standard deviations at a constant roll angle were pooled together and averaged to develop an estimate of the leave-one-out standard deviation (Eq. (5.3)). This term is referred to as $\bar{\sigma}_{\mathrm{local}}$ because it is meant to account for only the parametrically local variation in the lineloads due to changes in input points to the DFROM.

$$\bar{\sigma}_{\mathrm{local}}(\mathrm{X/L_{Ref}}, \phi_{\mathrm{P}}) = \mathrm{mean}(\sigma_{\mathrm{LOO}}) \tag{5.3}$$

To account for the effect of vehicle orientation (i.e., roll angle) on the leave-one-out variation analysis, all of the residuals at a constant roll angle were pooled together (46 leave-one-out cases and 19 pitch angles for each roll angle) and the standard deviation was computed. This resulted in standard deviations computed for each roll angle and vehicle body coordinate: $\sigma_{\mathrm{DB}} = \mathrm{f}((\mathrm{X/L_{Ref}}, \phi_{\mathrm{P}})$. Figure 5.4 presents color contour plots of these two standard deviations. The body station is represented on the horizontal axis, and the roll angle is represented on the vertical axis. The color contours represent the standard deviation values, and a planform outline of the vehicle is shown along the bottom of each plot for reference. As one would expect, the average local standard deviations depicted in Fig. 5.4a are slightly smaller in magnitude than the σ_{DB} values shown in Fig. 5.4b.

The total model form error is computed using Eq. 5.4, where σ_{global} is the standard deviation for the increment in variation due to vehicle orientation (across ϕ_{P} errors), and is referenced in Eq. 5.5.

$$\sigma_{\mathrm{ModelForm}} = \sqrt{\sigma^2_{\mathrm{global}} + \bar{\sigma}^2_{\mathrm{local}}} \tag{5.4}$$

$$\sigma_{\mathrm{DB}} = \sqrt{\sigma^2_{\mathrm{global}} + \frac{1}{\mathrm{N}}\bar{\sigma}^2_{\mathrm{local}}} \tag{5.5}$$

Equation 5.5 is rearranged to give Eq. 5.6. Substituting Eq. 5.6 into Eq. 5.4 gives the equation for the final model form error shown in Eq 5.7, where N=46.

$$\sigma_{\mathrm{global}} = \sqrt{\sigma^2_{\mathrm{DB}} - \frac{1}{\mathrm{N}}\bar{\sigma}^2_{\mathrm{local}}} \tag{5.6}$$

$$\sigma_{\mathrm{ModelForm}} = \sqrt{\sigma^2_{\mathrm{DB}} + \left(1 - \frac{1}{\mathrm{N}}\right)\bar{\sigma}^2_{\mathrm{local}}} \tag{5.7}$$

A contour plot of the final model form standard deviations, computed as a function of $\mathrm{X/L_{Ref}}$ and ϕ_{P}, is presented in Fig. 5.5. This approach to modeling the lineload model form error accounts for different levels of variation that occur across the physical locations along the length of the vehicle, and it accounts for variation in the errors when the vehicle changes roll orientations.

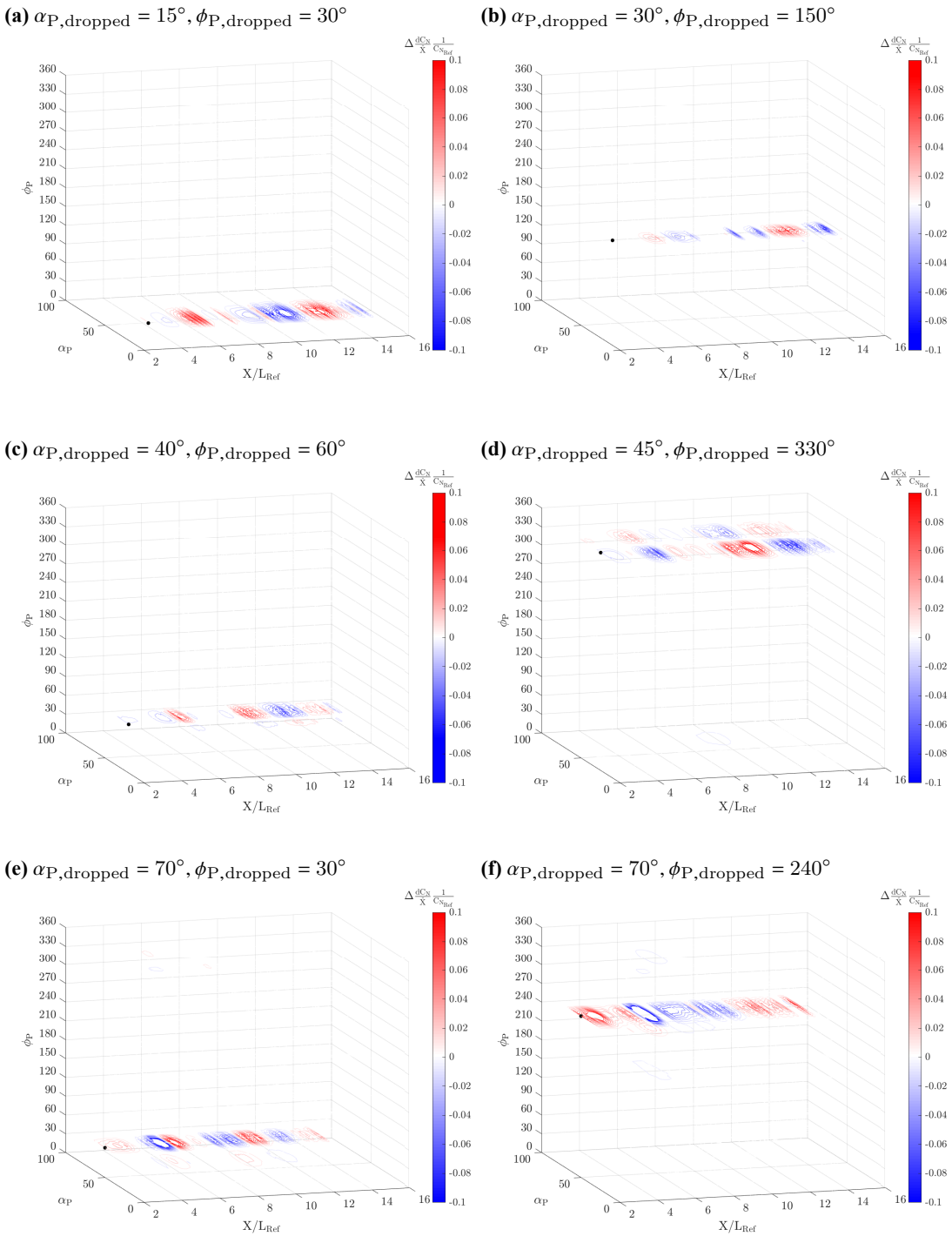
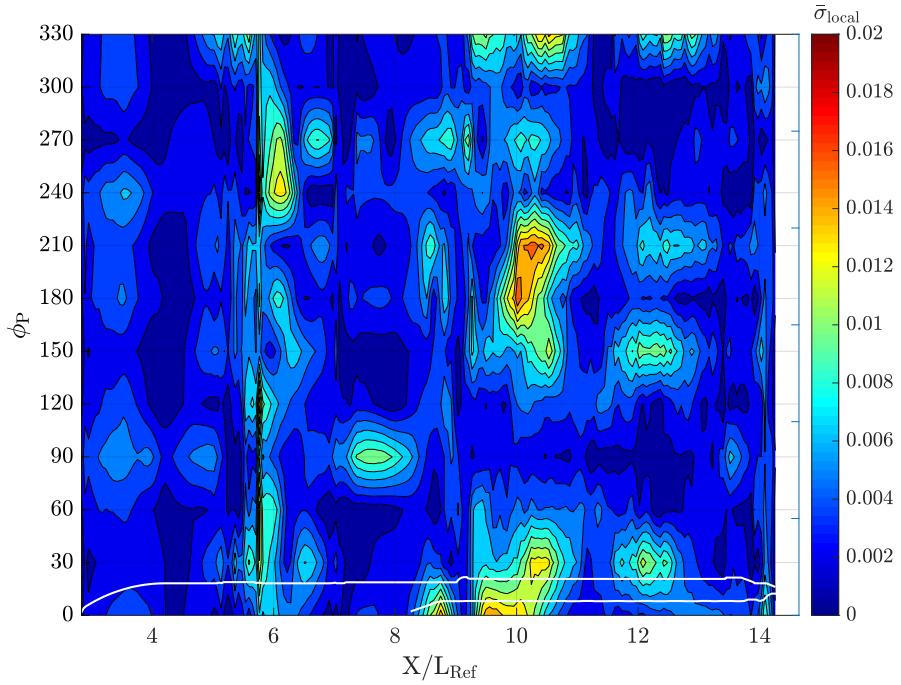
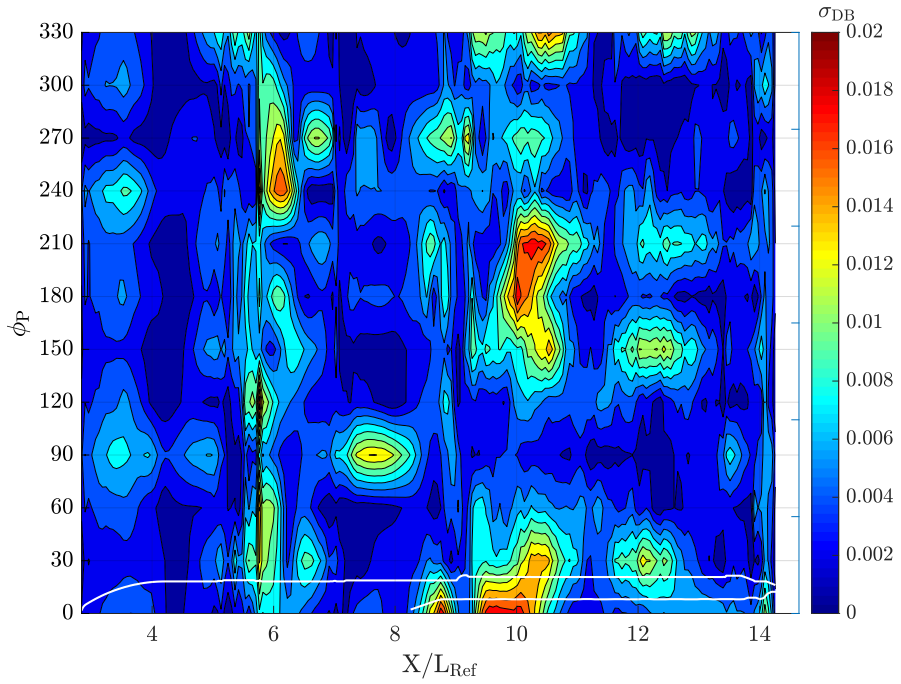


Figure 5.3 Three-dimensional contour plots of residuals computed from selected leave-one-out cases.



(a) $\bar{\sigma}_{\text{local}}$



(b) σ_{DB}

Figure 5.4 Contour plots of individual model form error terms.

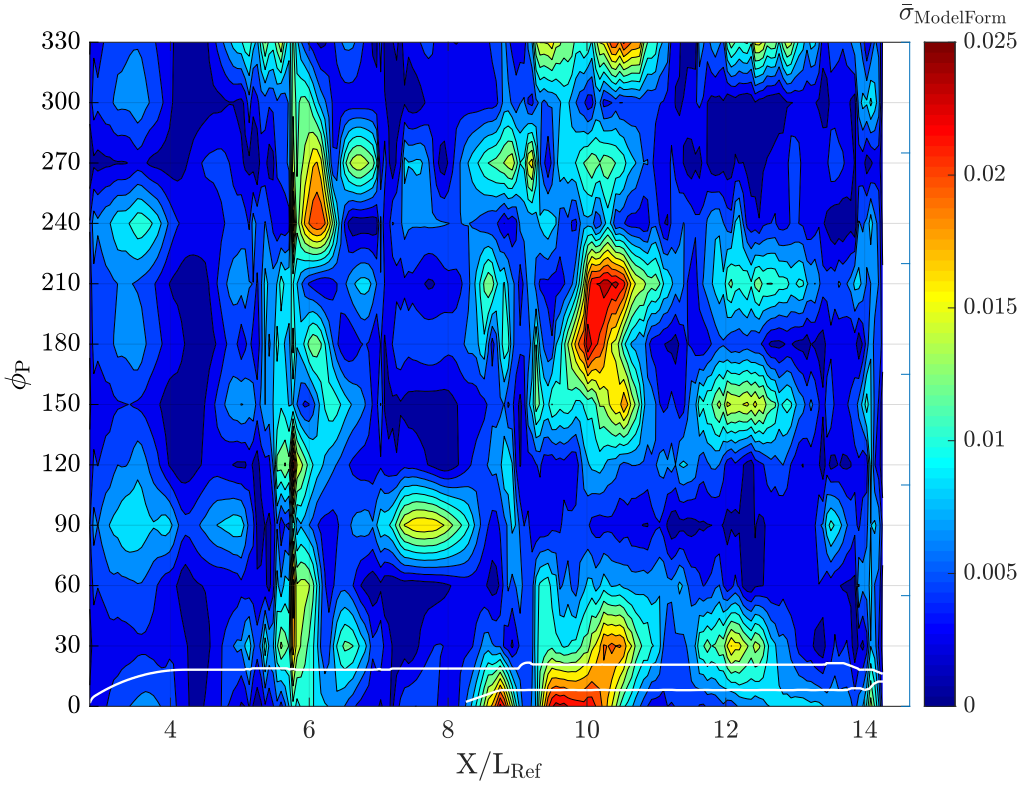


Figure 5.5 Contour plot of model form errors ($\sigma_{\mathrm{ModelForm}}$) computed as a function of body station and vehicle roll angle.

5.2 Incorporating Uncertainty from the Force and Moment Database

As explained in Chapter 4 the model used to predict lineloads can be adjusted to any force, moment or combination thereof. Since the trusted force and moment database has an uncertainty associated with each value, the forces and moments used in the data fusion process can be dispersed based on the uncertainty in that database. The lineloads that are calculated from this variation in source data can be analyzed to determine how the dispersion of the integrated values affect the predicted lineloads.

The uncertainty bounds in the force and moment database used in the data fusion model were quantified from the experimental data used to construct that database. In an effort to characterize lineload uncertainties, data fusion was also performed to match the ROM lineloads to dispersed forces. For this investigation, 5000 randomly selected combinations of normal force and pitching moment uncertainty levels (varying between -3σ and $+3\sigma$ on a normal distribution) were applied to the nominal database values, and 5000 alternate lineloads were computed for each nominal pitch and roll angle in the database. Figure 5.6 shows an example of some of the normal force lineloads generated from the Monte Carlo runs for $\alpha_P = 50°, \phi_P = 120°$, with each individual line being a different run. The maximum and minimum values seen across the Monte Carlo are plotted using dotted and dashed black lines, respectively. The nominal value is plotted in solid red. Along the bottom of the figure is a planform of the vehicle. Each blue line is associated with a Monte Carlo run (~200 are plotted). As shown, there are more lines close to the nominal and the frequency drops off further away from the nominal.

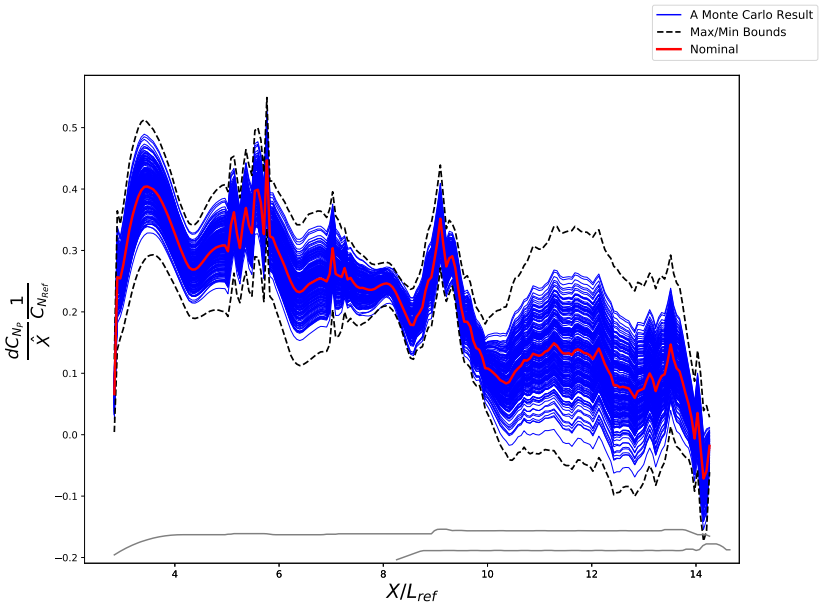


Figure 5.6 Variation in DFROM lineloads matched to dispersed normal forces for $\alpha_P = 50°, \phi_P = 120°$.

Once the Monte Carlo is finished, the analysis to get the uncertainty associated with the data fusion is almost identical to what is done to calculate the model form uncertainty. First, these 5000 alternate

lineloads were compared to the lineloads computed to match the nominal force and moment database values. As expected, the magnitude of the effect of including the force database uncertainties in the data fusion process varies across vehicle pitch and roll angles and for certain combinations of dispersion levels. At this point, plots of the difference as a function of X/L_{ref}, α_P, and ϕ_P can be generated to look for trends as seen in Fig. 5.7. This is plotted in the same way as the forces in Fig. 5.6. Again the dotted and dashed lines represent the maximum and minimum values observed. The nominal value (zero in this case) is again plotted in solid red. Again, each thin blue line represents one Monte Carlo run. The first item to note is the existence of larger values away from the center of the vehicle. This is due to the method of data fusion and the assumption that the normal force and pitching moment is uncorrelated. Because the moments are measured toward the rear of the vehicle, the data fusion tends to mostly adjust the nose of the vehicle when changes in pitching moment are needed. Normal force only dispersions tend to shift the entire lineload up and down. When these changes are in opposition, the geometric center of the vehicle changes the least. This is likely a consequence of the UQ model used for the force and moments, which are not treated as correlated.

The first data reduction done is converting the differences to a standard deviation at each X/L_{ref}, α_P, and ϕ_P. Figure 5.8 shows the standard deviation as a function of X/L_{ref} at a constant $\phi_P = 0^\circ$ at various α_P. The equivalent plot for constant α_P and varying ϕ_P is seen in Fig. 5.9. These plots show some interesting trends. The most notable is that with the constant ϕ_P, the standard deviation keeps the same general shape and magnitude as the angle of attack changes. This is opposed to the constant α_P standard deviations that show significant shape changes as the vehicle rolls, and is especially evident near the noses of the SRBs. Also of note, the standard deviations for opposite roll angles are similar. This confirms that similar to the model form error, it is best to pool the residuals along constant roll angles.

It is clear that at low angles of attack, there is no significant effect of dispersed forces on the potential variation in the sectional loads. As angle of attack increases from 25 to 60 degrees, the standard deviation increases slightly on the forward half of the vehicle and increases more dramatically on the aft half of the vehicle. The dispersed forces and moments have the largest effect on the sectional loads between 30 and 60 degrees on the aft half of the vehicle (aft of the forward SRB attach point). As the angle of attack increases beyond 60 degrees, the effect of matching the lineloads to dispersed forces decreases by roughly 50%. This is mostly a reflection of how the force and moment database uncertainty was developed with larger values in the harder to predict midrange α_P region with lower values of the uncertainties at the extremes.

Using the same methods in the model form section (Section 5.1), the total data fusion uncertainty was estimated by combining two levels of variations. Similar to Eqs. 5.2 and 5.3, a local and a database standard deviation can be calculated. Once calculated, they can be combined using Eq. 5.8

$$\sigma_{\text{DataFusion}} = \sqrt{\sigma_{\text{DB}}^2 + (1 - \frac{1}{\text{n}})\bar{\sigma}_{\text{local}}^2} \tag{5.8}$$

where *n* is equal to the number of Monte Carlo runs (5000 in this case).

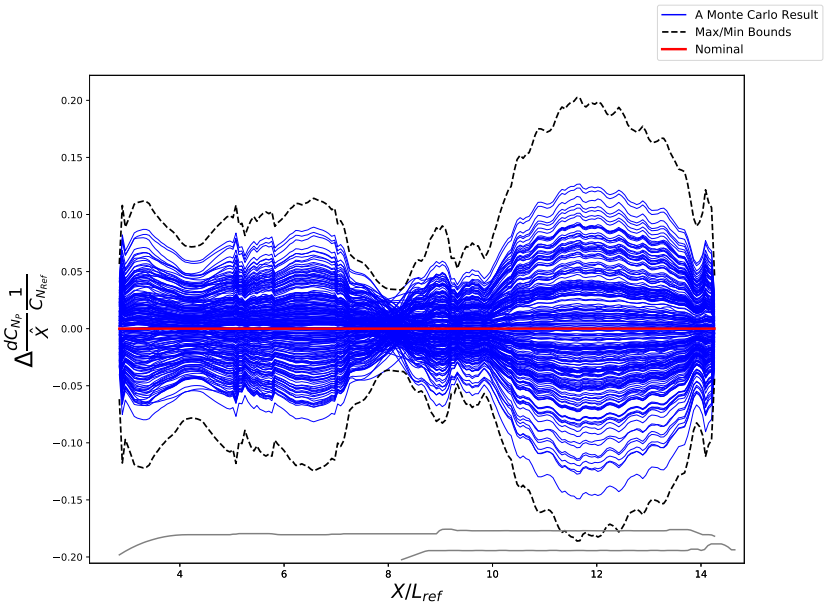


Figure 5.7 Difference between baseline DFROM lineload and lineloads matching dispersed values for $\alpha_P = 50°, \phi_P = 120°$.

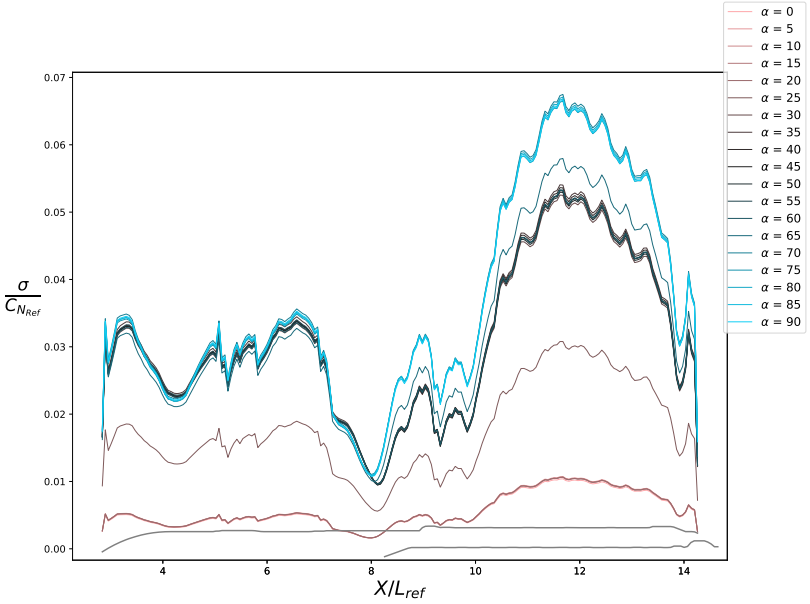


Figure 5.8 Standard Deviations from difference between baseline DFROM lineload and lineloads matching dispersed values at $\phi_P = 120°$.

Contour plots of the two data fusion error terms are presented in Fig. 5.10. The normalized distance along the body is shown on the horizontal axis, the nominal roll coordinates are represented on the vertical axis, and a mapping of the computed standard deviations is presented as color contours. The σ_{local} and σ_{DB} terms are similar, with σ_{local} being only slightly smaller in magnitude. Figure 5.11 presents a contour plot of the overall standard deviations associated with the data fusion process as a

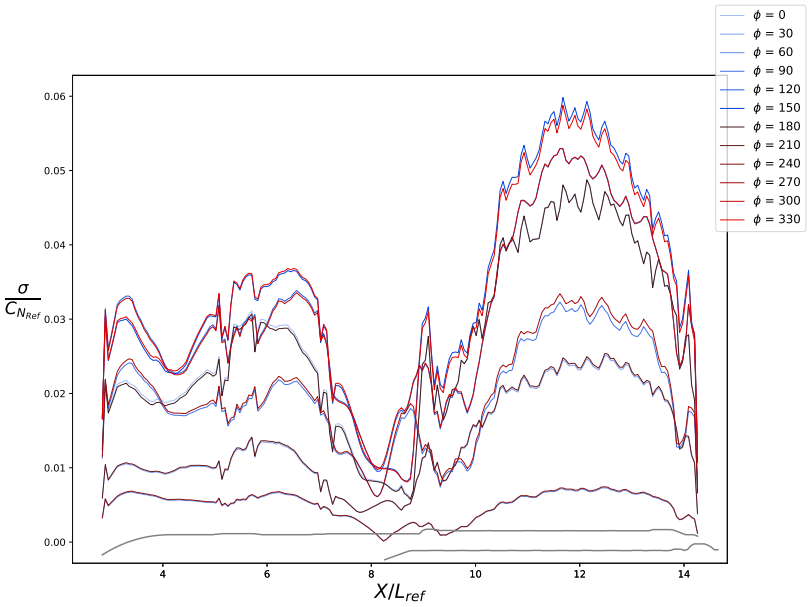
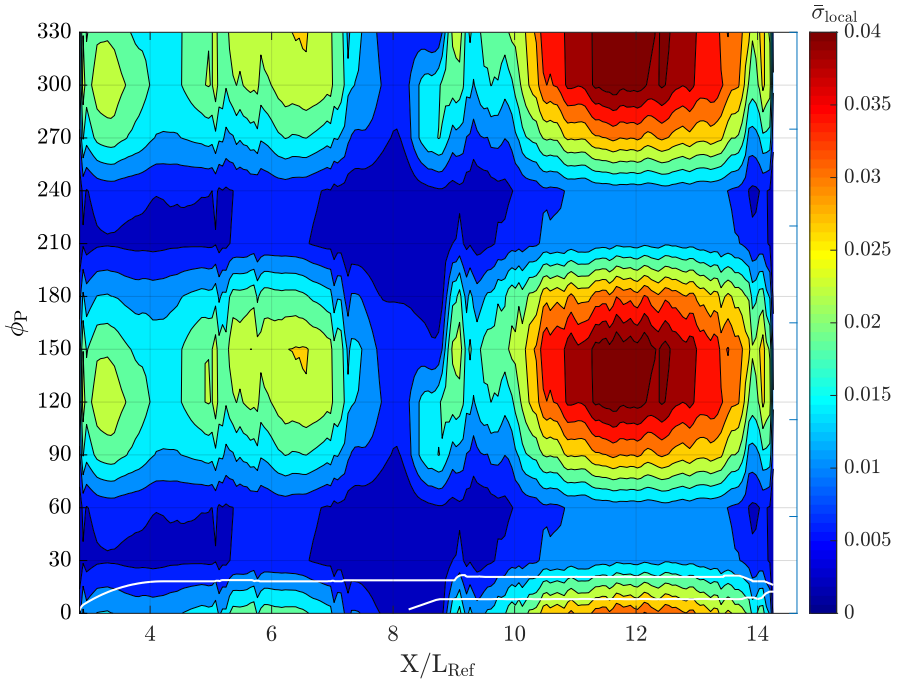
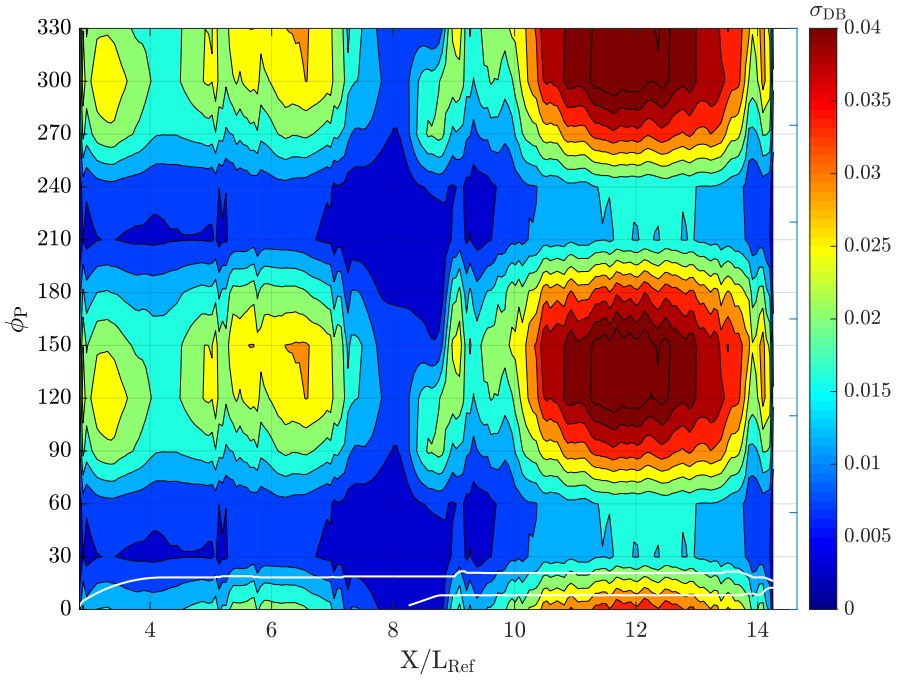


Figure 5.9 Standard Deviations from difference between the baseline DFROM lineload and lineloads matching dispersed values at $\alpha_P = 50°$.

function of vehicle location and roll angle.



(a) σ_{local}



(b) σ_{DB}

Figure 5.10 Contour plots of individual data fusion error terms.

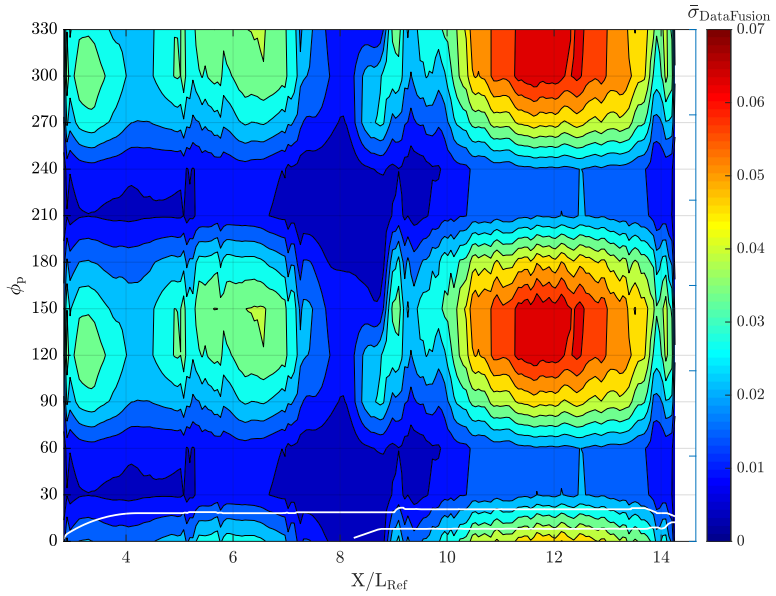


Figure 5.11 Contour plot of data fusion errors ($\sigma_{\mathrm{DataFusion}}$) computed as a function of body station and vehicle roll angle.

5.3 Combining Uncertainties

Once the uncertainty from multiple sources of error have been analyzed, they need to be combined. A fairly typical method, root sum square (RSS), is used to combine uncertainties here (Eq. 5.9).

$$\sigma_{LL} = \sqrt{\sigma^2_{ModelForm} + \sigma^2_{DataFusion}} \tag{5.9}$$

Figure 5.12, shows the RSS results at $\phi_P = 90°$. The light blue line is the standard deviation associated with the data fusion process while the red line is the standard deviation associated with the model form. The black line is the combined value. As is typical for this database, the combined value is dominated by the term related to data fusion. The model form term tends to have the greatest contribution in the region a short distance above the SRB noses. This is likely to be a result of the model form being more sensitive to the influence of the asymmetric vortices than the data fusion term. The aft region of the vehicle behind the SRB forward attach point is the region of the vehicle with the most variation because of the data fusion term. For the rest of the roll angles, a contour plot of the total standard deviations computed from Eq. 5.9 is presented in Fig. 5.13. This visualization of the total estimated errors shows that the trends at $\phi_P = 90°$ are similar across the whole domain.

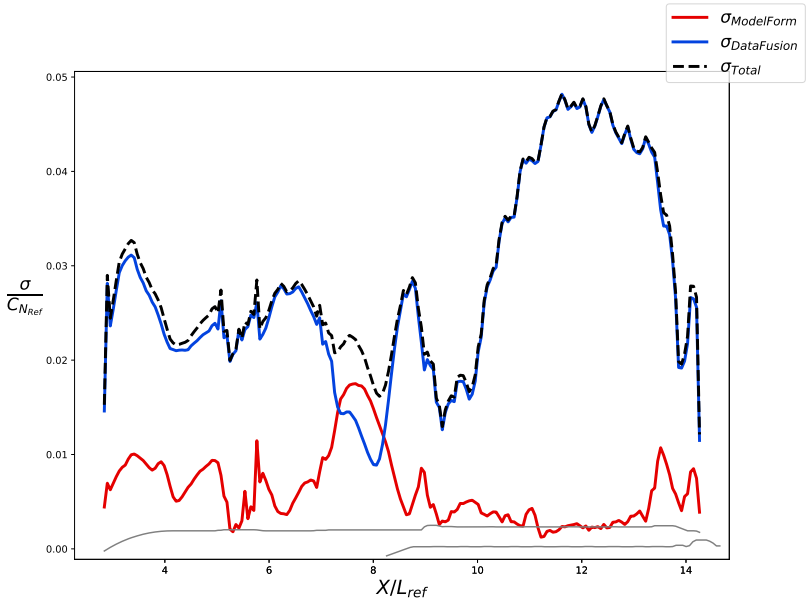


Figure 5.12 Results of the RSS of the sources of uncertainty $\phi_P = 90\,\mathrm{deg}$.

Once there is a total standard deviation that is a function of X and ϕ_P, it can be applied to the nominal database to get the dispersed lineloads. Typically SLS aerodynamic databases are delivered with 3σ uncertainties. To clarify, three times the combined standard deviation can be applied to the nominal to get a sense of the values that would be covered. Figure 5.14 shows the nominal at $\alpha_P = 50°, \phi_P = 90°$

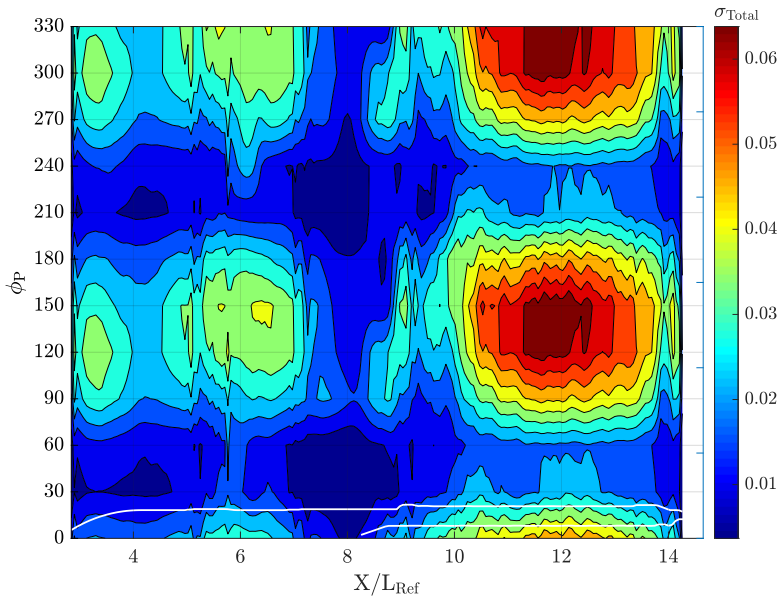


Figure 5.13 Contour plot of total standard deviations computed as a function of vehicle body station and roll angle.

with the 3σ values shaded. The blue shading represents what the data fusion term would cover, and the red shade would cover the model form term. The black is the combined final possible distributions used.

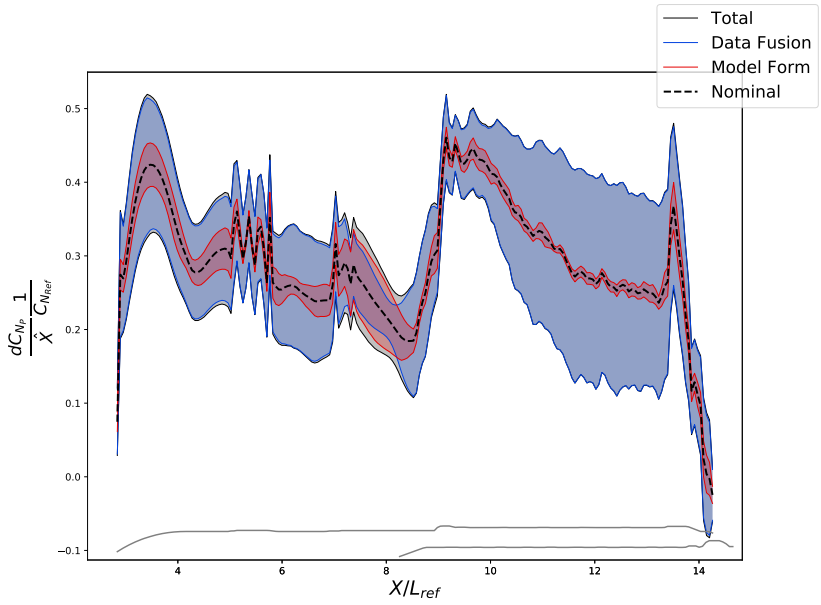


Figure 5.14 Nominal database value at $\alpha_P = 50°$, $\phi_P = 90°$ with range of dispersions.

The combined uncertainty model expands the developed DFROM to provide an increment-based

uncertainty model. This model is built up of two major terms, the model form error and variations in the source data used in the data fusion. The model form error is developed using leave-one-out cross validation techniques as opposed to the more typical test and validation points due to the expense of additional data. The uncertainty in the force and moment database is then used to explore the model's sensitivity to changes in the source data. This analysis resulted in a much larger variation in predictions than the model form error. These two error terms were then combined to create a unified error model that is a function of both α_P and ϕ_P.

As an alternative to this increment based approach, exploration in using variations in the ROM coefficients were also explored.

5.4 Distribution on ROM Coefficients

To better take advantage of the model structure, a secondary approach to uncertainty quantification was explored focusing on the PCA process. By characterizing the uncertainty in the PCA process, variation can be introduced in the recombination step. While either PCs or the coefficients could be varied, focus here is on finding a distribution for the ROM coefficients. This choice makes it easier to propagate during the data fusion step.

Based on work in the 1970s and 80s and by using asymptotic theory and assuming random variables, it is possible to calculate an asymptotic standard error for the coefficients [Jol02]. These assumptions are not valid in the current work; however, use of a standard error provides a simple and quick way to explore how changing the initial value for a can change the ROM prediction and where those assumptions fail.

To help address the shortcomings of assuming a normal distribution, bootstrapping is commonly used. In particular bootstrapping can help find confidence intervals on statistics of interest and has been applied to the transformation matrix [Efr82]. This method has been successfully used to characterize PCA confidence intervals in a variety of datasets [JD88; AF16]. However, unlike most applications in PCA instead of removing "samples", "variables" will be removed. The problem then becomes matching the reduced model set with the original ROM coefficients, a. By reducing the number of snapshots, the number and possibly the characteristics of the PCs change and so a direct mapping is impossible. To generate an indirect mapping, a least squares problem is set up and used to match the original PCs with predictions of the reduced set.

The resulting distributions on the least squares analysis is used to build a Laplace distribution, which allows sampling as necessary. Sampled coefficients are then used to reconstruct integrated forces and moments as well as lineloads and compared to the previous model form error calculated in Sec. 5.1. They are then used in the full data fusion process to explore changes in the final product.

5.4.1 Standard Error

To begin our evaluation of applying a stochastic model to the ROM coefficients, a normal distribution of our samples is assumed and the standard error is calculated. From Ref. [Jol02], the standard error for the coefficients can be written as:

$$\sigma_a(a_{m,h}) = \left[1/n\lambda_h \sum_{j \neq h}^{P} \frac{\lambda_j}{(\lambda_j - \lambda_h)^2} a_{m,j}^2 \right]^{1/2} \tag{5.10}$$

where a is a ROM coefficient and λ is an eigenvalue of the system. While covariance can also be calculated, it is neglected for this work.

This standard error can then be used as the standard deviation in a random draw on the coefficients using a normal distribution. However, the standard error only exists at points where CFD source data are available. As discussed previously, $a_{m,h}$ can be converted to a function of α_P and ϕ_P by mapping where the original CFD data is. In order to expand to the full parameter space, a cubic interpolation is

used, which is consistent with the values of a themselves. During ROM evaluation, a random number from $\mathcal{N}(0, 1)$ and multiplied by σ_a can then be added to the value of a.

The initial examinations of the standard error look promising despite the assumption on a normal distribution of the surface data. The distribution of these terms are favorable, as seen in Fig. 5.15, which shows the values of $\sigma_{a,i}$ related to the pressure PCs as a function of α_P when $\phi_P = 0°$ before interpolation. Each marker represents a different PC. As desired, these value are relatively consistent across the domain, with a slight increase in the midrange α_P values, which corresponds to the area most sensitive to multimodal effects. In practice, however, these values are multiplied by each PCs contribution to a lineload or force, which are distributed very unevenly.

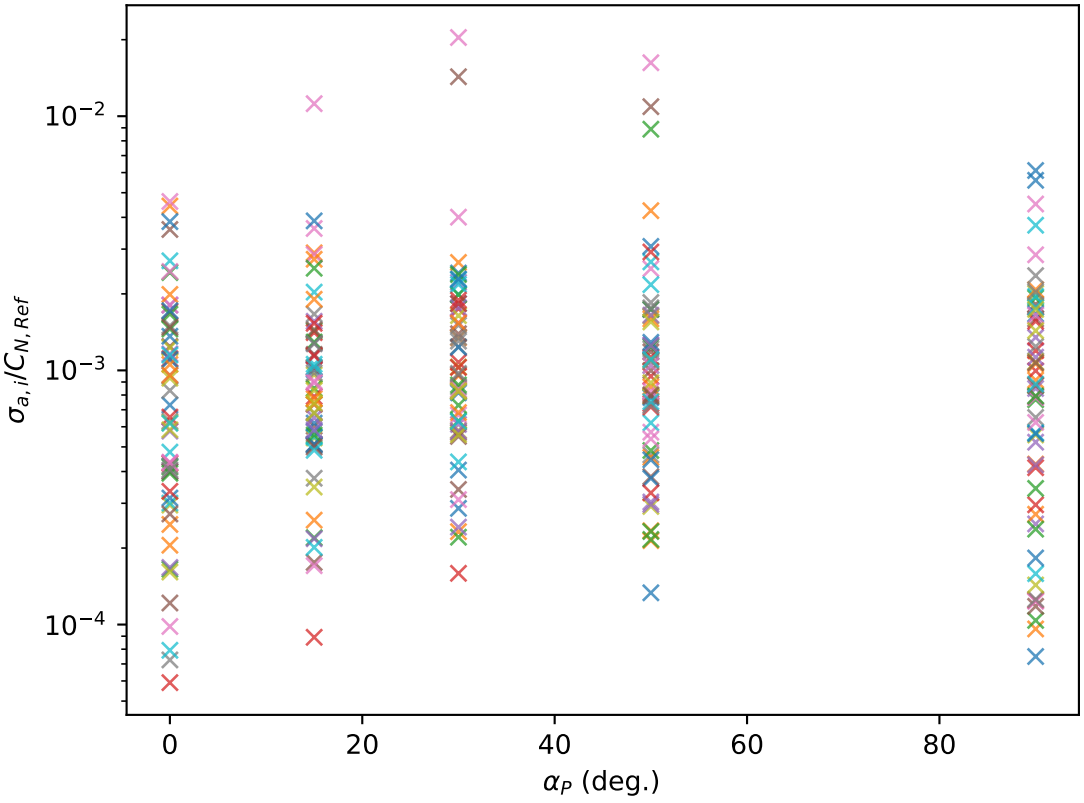


Figure 5.15 $\sigma_{a,i}$ as a function of α_P when $\phi_P = 0°$. Each marker represents a different pressure PC coefficient.

Remember that the force prediction at a given condition is

$$C_N = \sum a_i \Delta C_\psi, i \tag{5.11}$$

and by treating them as independent random variables, the variance of the prediction becomes:

$$\sigma_{C_N} = \sqrt{\sum (\sigma_{a,i} \Delta C_\psi, i)^2}. \tag{5.12}$$

The individual contributions and the total contribution are plotted in Fig. 5.16. The black dot represents the total standard deviation while the x symbols represent individual pressure PC contributions to the variance. Unlike before where the uncertainty looked relatively uniform across α_P, the uncertainty greatly increases as α_P decreases. This is inverted from what is expected because of the confidence in the underlying CFD data as well as the earlier LOO results. The lower the α_P, the easier the CFD problem is and the greater the confidence in the underlying answer. This is also in contradiction to the earlier LOO analysis, which saw less variations at the lower α_P values. Seeing such great variation at α_P =0º highlights the limitations of assuming the standard error could be used. While this method could be used to evaluate lineloads as well, the limitations seen here spurred development of alternative ways to look at the uncertainty in the coefficients.

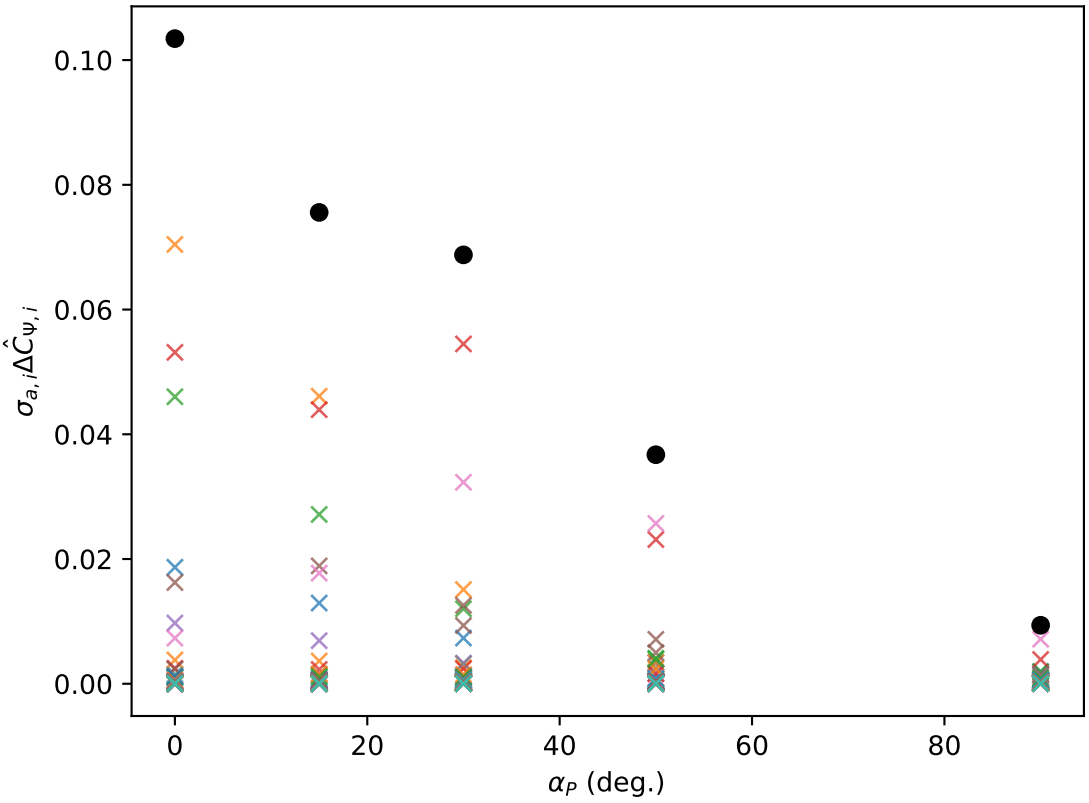


Figure 5.16 $\sigma_{a,i}\Delta C_\psi, i$ as a function of α_P when ϕ_P = 0°. Each marker represents a different pressure PC coefficient. The solid black dot is σ_{C_N}.

5.4.2 Bootstrapping with Least Squares Fitting

There are many situations in which the standard error is insufficient in PCA, often due to violation of the assumption of normally distributed data. In these cases, bootstrapping is a common tool to get an estimate of the confidence interval. Unlike most other cases, here instead of being able to drop samples, I would like to drop snapshots, which in traditional PCA framing are called variables or dimensions. Methods for doing this form of bootstrapping have not been explored and so a method is developed here.

The challenge with dropping snapshots is matching the nominal transformation matrix with the bootstrapped versions. For one thing, the dimension of the transformation matrix changes as variables are added or dropped. Filling the gap with a trivial case would not help with maintaining PC order even if it keeps the dimension constant. Also, if the trivial case is poorly constructed it could lead to numerical issues if the resulting system no longer is linearly independent. Even if a reduced PC space is used, the ordering of the PCs could radically change if a significantly influential snapshot is removed. Although much easier to handle, the eigenvectors used in PCA can be either positive or negative, which further makes mapping difficult. To overcome this, predictions from the bootstrapped models are compared to the full model to generate a residual. A least squares method is used to fit the nominal PCs to this residual to get a measure of how much the coefficients would need to change to match these bootstrap predictions. This bypasses the need for a direct mapping.

In other words, given that the nominal prediction is found via:

$$\Psi\overrightarrow{A} = S \tag{5.13}$$

and the bootstrap prediction is found via

$$\Psi_B\overrightarrow{A}_B = S_B \tag{5.14}$$

where Ψ_B represents the PCs generated by leaving snapshots out, $\overrightarrow{A}_B$ are the ROM coefficients generated using the reduced data, and finally S_B is the predicted surface. $\overrightarrow{A}^*$ can than be found by trying to solve

$$\Psi\overrightarrow{A}^* = S_B. \tag{5.15}$$

This is an underconstrained system, and so solving exactly is impossible. The simplest approach is treating it as a least squares problem, but some adjustments to the problem formulation are done first.

The first is solving for the residual instead of S_B directly. Since the goal is to determine the distribution on $\overrightarrow{A}$, this allows calculating variations around 0 as opposed to the nominal value. The resulting distribution is then easier to combine across the parameter space. To differentiate these "perturbed" values, $\overrightarrow{B}$ will be used.

$$\Psi\overrightarrow{B} = S - S_B \tag{5.16}$$

A simpler semispan plane geometry is used to explore this initial portion of the method as elaborated

in Appendix A. The results of this test case showed similar trends to the SLS geometry, which gives confidence in the new method. For the SLS geometries, this is an approximately 800,000 x 60 system and needs to be solved once for each variable and at each parameter point of interest. To save computational development time, values for $\vec{B}$ are found at every $\Delta\alpha_P$ = 10° starting at α_P= 5° instead of every $\Delta\alpha_P$ = 5°. One downside of this method is that at points where CFD data are available, all but one of the 46 residuals are 0, which makes a trivial system to solve and gives no information on the distribution. There is an open question of whether it is better to do some form of weighting to reflect the cascading variance and contributions of each PC; however, initial results do not seem to indicate that it is necessary. Now that the system is solved, the values of $\vec{B}$ are plotted after filtering out those that are 0, which represent the trivial system. Unlike previously where variations are explored as a function of α_P and ϕ_P, all the resultant values for B_i are combined. This is done to make any resulting distribution more apparent; however, future work should explore how dependent these distributions are to location in the parameter space. In general, the distributions across PC numbers are similar where most values are near zero, but there are a significant number of values in the tail that need to be accurately captured for proper uncertainty quantification.

Some summary statistics for the distributions can be found in Table 5.1, which shows how the mean (μ), variance (σ^2), 99.7th percentile, and the range change as a function of PC number. These are for the PCs associated with pressure, but the friction ones show similar trends. As expected, the mean values have positive and negative values showing the importance of a two sided distribution. While not monotonically increasing, the three measures of spread in the remaining columns increase as PC number increases. This shows this method predicts a smaller interval on the coefficients associated with the earlier PCs. One thing to also note is if the distributions were close to normal, the 99.7% value would be close to 3σ, which is not the case here. This points to the distribution having a long tail, which can often be difficult to capture.

The distribution for the first coefficient related to pressure, shown in Fig. 5.17, are similar to those

Table 5.1 Summary statistics on the distribution of B_i.

PC Number	μ	σ^2	99.7%	Range
1	-5.034e-05	1.190e-05	1.922e-02	1.138e-01
2	8.862e-06	6.284e-06	1.733e-02	7.227e-02
3	-3.171e-05	1.068e-05	1.730e-02	1.084e-01
4	2.908e-05	1.389e-05	3.229e-02	1.070e-01
5	7.540e-05	1.696e-05	2.310e-02	1.386e-01
10	4.612e-04	7.183e-05	5.536e-02	2.435e-01
20	2.062e-04	3.249e-04	1.012e-01	6.954e-01
30	2.827e-04	9.618e-04	2.364e-01	1.197e+00
40	-1.939e-04	3.094e-03	3.349e-01	2.094e+00
50	2.886e-04	5.575e-03	2.511e-01	3.444e+00
60	-2.473e-04	9.153e-04	1.950e-01	1.005e+00

seen in Appendix A, which lead to the exploration of looking for a canonical distribution to fit the data. The histograms show two possible distributions that could fit the data. On the left is a histogram of the absolute values as well as the probability density function of an exponential distribution. Of note, the data have a sharp decline in probability as values increase; however, even the exponential distribution fails to capture the probability of the larger values of B_1. This helps highlight the difficulty of capturing the long tails that are present in the distribution. On the right is the data transformed by the natural log and a lognormal distribution fit to the data. This does a better job of fitting observed data than the exponential distribution. However, the lognormal distribution does not go to 0, which is desired given the intended use of this distribution to generate perturbations from a nominal value. It is also theorized that due to the discrete nature of the residuals being used, there are missing values closer to zero. In other words, dropping information from the system during the bootstrap process is a large perturbation, which makes it impossible to capture small values of B. While not difficult, special care would also be needed make both positive and negative draws.

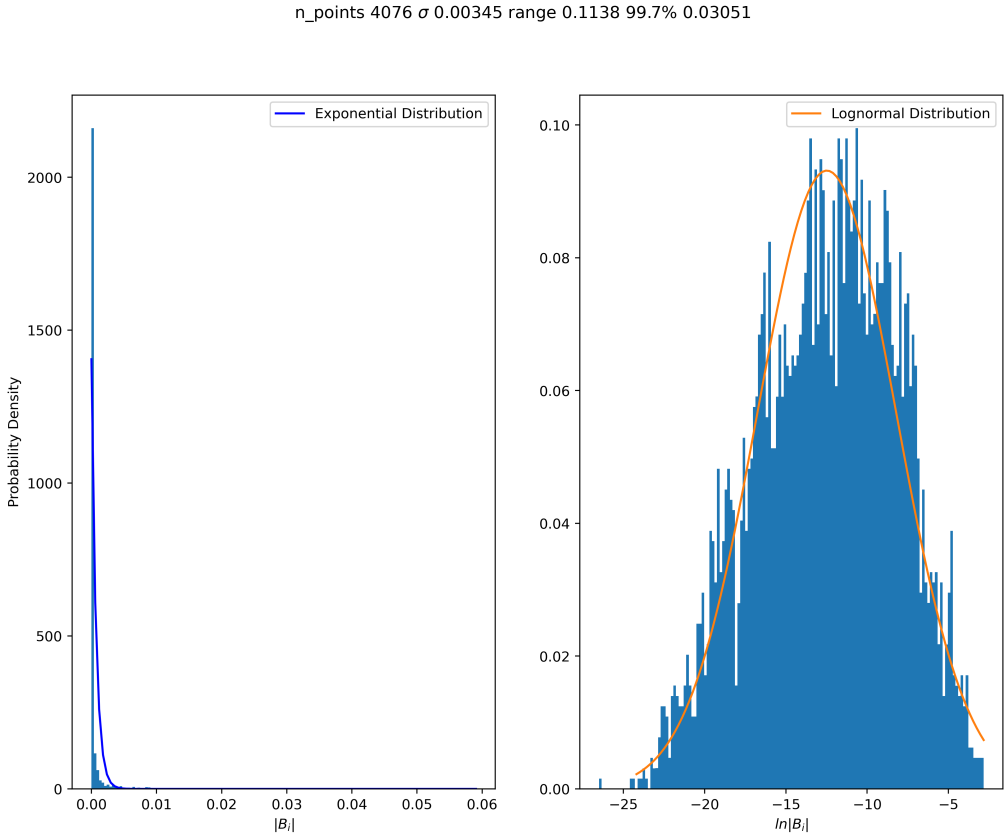


Figure 5.17 Histogram of $B_{C_P,1}$. (a) shows the absolute value and (b) shows the natural log of the absolute value.

Since the lognormal distribution does not naturally include the chance for 0, the exponential distribution was chosen. To capture the negative values as well, the exponential is doubled into a Laplace distribution. The Laplace distribution has the probability density function of

$$f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}. \tag{5.17}$$

With μ being the center of the distribution which would be 0 in this case, which leaves finding the scale parameter b. The maximum likelihood estimator for b is

$$b = 1/n \sum_{i=1}^{n} |x_i - \mu| \tag{5.18}$$

It is possible that estimating b could be improved. Further work could adjust the value to better match the observed long tails and account for the discrete nature of the residuals. One further limitation that could be addressed in future work is accounting for the error in the least squares fit itself. While it minimizes the remaining residual, it does not remove it completely. Unlike the standard error, which was relatively constant across the PCs, the Laplace distribution variance increases with PC number. Since the first few PCs represent the greatest contribution to integrated forces and lineloads, the variation in the integrated forces will be reduced compared to the standard error. With this in mind, the Laplace distribution is used to generate uncertainty for the ROM coefficients.

5.4.2.1 Bootstrap Results

Unlike the previous model form error term, these variations in the ROM coefficient are not a function of ϕ_P. This will have the tendency of smoothing out error across the space, however, because of the use of the missile axis, the contribution of each PC to the normal force will be a function of ϕ_P. The first result looks at the distribution on normal force as a function of ϕ_P before moving onto the lineloads. Finally, the random draws are integrated to the data fusion process to explore how this new uncertainty interacts with the force and moment uncertainty.

As before, the force is now the sum of a collection of random variables. The only difference is the underlying distribution and the value of $\sigma^2_{a,i}$ being a constant compared to the standard error. The resulting forces (and lineload stations to be described later) are the sum of many Laplace distributions and while not exactly a normal distribution, treating it as such is a minor assumption. Figure 5.18 shows the predicted distributions in $\hat{C_{N_P}}$. Each blue x symbol represents another evaluation using the developed distributions. Since these values are normalized to the maximum normal force observed, it is easy to see that this represents at maximum about a 1 percent variation in the force.

Much like the forces and moments, the lineloads are generated using the random perturbation to the ROM coefficient. However, since the lineload contributions are a function of x, these predicted variations will also be a function of x. The results of the lineloads are shown in Fig. 5.19, which shows the dispersed lineloads subtracted from the nominal value at a $\phi_P = 0°$. The black lineload is simply a representative case highlighted to help distinguish it from the 100 other lineloads. While difficult to pick

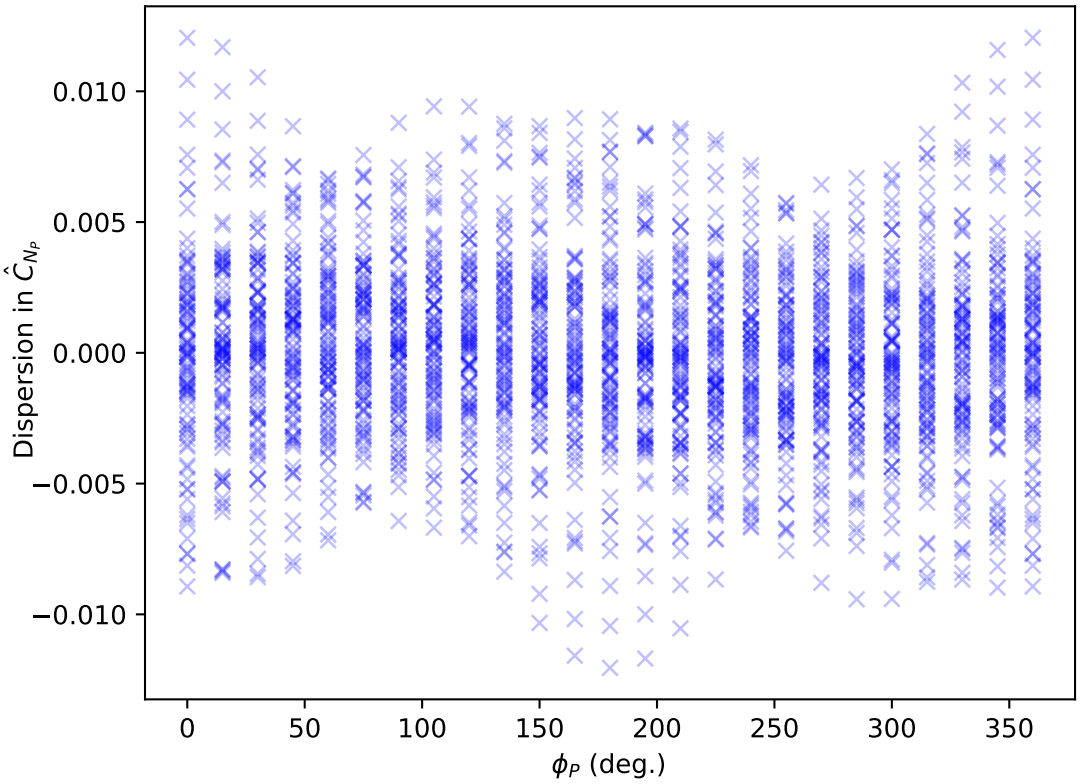


Figure 5.18 Predicted distributions on $\hat{C_{N_P}}$ as a function of ϕ_P.

out individual lineloads, some characteristics are evident. Overall, the variations are relatively constant across the vehicle with only a few regions of the rocket with a different variance. Unlike the previously derived method, variations in all lineloads are available at the same time.

To better match previous plots, Fig. 5.20 shows only the normal force lineload variations at ϕ_P = 90º. Unlike the previously derived increment method, these dispersed lineloads can cross the nominal allowing for variations in lineload shape and not just the magnitude. One thing to note is the slight asymmetry in the maximum and minimum bounds shown in the dotted lines. Although 5000 samples were used, there is no reason for the asymmetry since at each point positive and negative values are equally likely. This is likely due to the limitations of sampling from distributions with a long tail.

The next step is to see how the variations in the ROM coefficient affect the data fusion method. The Monte Carlo procedure is repeated but instead of drawing random values for the target forces and moments, the initial ROM coefficient varies. While all the differences from the nominal shown in Fig. 5.21 sum to zero, there are some interesting trends. These results are for α_P = 50° and ϕ_P =120°, which show that unique lineloads are predicted based on the uncertainty in the ROM coefficients. The data fusion method works by adjusting the ROM coefficients; and in some cases, it could be possible that this method overcomes small perturbations in the starting values. While the pure ROM lineloads are

approximately symmetric, these are not symmetric, at all, which can be seen by comparing the maximum and minimum dashed lines especially toward the aft end, which has a double peak structure on the minimum but not the maximum. The maximum and minimum variations at each x location are about the same size as those seen before the data fusion with it being about 1% of the maximum normal force.

To further the comparison with the previously derived model form error, the standard deviation of these dispersed lineloads are taken. The standard deviation of this bootstrap method σ_B can be seen in Fig. 5.22, which can be compared to Fig. 5.12. This figure shows the standard deviation of the values shown in Figs. 5.20 (labeled ROM) and 5.21 (labeled DFROM). Compared to $\sigma_{ModelForm}$, the ROM values are approximately half. Also, the general trend is different showing a local minimum where $\sigma_{ModelForm}$ peaked. However, there are two limitations when trying to directly compare this dispersion with the model form error. The first is that the model form error was derived on the LOO residuals after data fusion while the bootstrap draws are done before any data fusion. The second is that the model form error has a reliance on ϕ_P, which helps it maintain data from some of the larger residuals when doing the LOO analysis. The smaller range could also be a limitation in the fitting of the Laplace distributions, which often missed some of the values in the long tail. After introducing the data fusion step, there is a significant reduction in variation across the vehicle. This is an expected result since the data fusion was introduced to help overcome deficiencies in the ROM predictions, which this shows.

To finalize the evaluation of this method, it is used in conjunction with the uncertainty in the forces and moments. The Monte Carlo procedures are once again repeated with random values for the integrated force and starting ROM coefficients. An example of the resulting residuals of the dispersed lineloads are shown in Fig. 5.23, which shows the results for $\alpha_P = 50°$ and $\phi_P = 120°$. This figure is essentially a replica of Fig. 5.7, which is unsurprising. The variations seen in the ROM coefficients represent less change than the previously derived model form error, which was significantly dwarfed by the uncertainty in the target forces and moments. Unlike the increment-based method, this new method allows for predictions that cross the nominal, which allows for maintaining force and moment bounds.

By focusing on moving the model form uncertainty into the ROM coefficients, a more stochastic lineload model with uncertainty is possible. The new method predicts a lower level of model form error; however, the integrated lineloads after the data fusion process maintain the expected bounds of the forces and moments. This further helps alleviate the concern of increasing the spread of possible integrated forces by adding a new source of uncertainty. It further reinforces that the lineload errors are more strongly influenced by the uncertainty in the forces and moments compared to the ROM model form, and additional work should be done to better characterize them. While the long tails are not fully captured by the proposed exponential distribution, the bootstrapping method is successfully used to develop a prediction interval for the ROM coefficients. Further work should be done to incorporate information based on the remaining residual after the least squares fitting, which is currently ignored.

Significant work remains to incorporate uncertainties from the CFD used to generate the underlying ROM. Further work should also analyze the resulting uncertainty model to compared predicted integrated forces to the bounds of the force and moment database used. While the new model with uncertainty terms is expected to go outside the bounds of the force and moment database the frequency and magnitude of

excursions would be a helpful addition in evaluating this uncertainty model for use in production. While the DFROM presented and developed in earlier chapters is already in production, there are still significant improvements and new directions to expand all the components of this method.

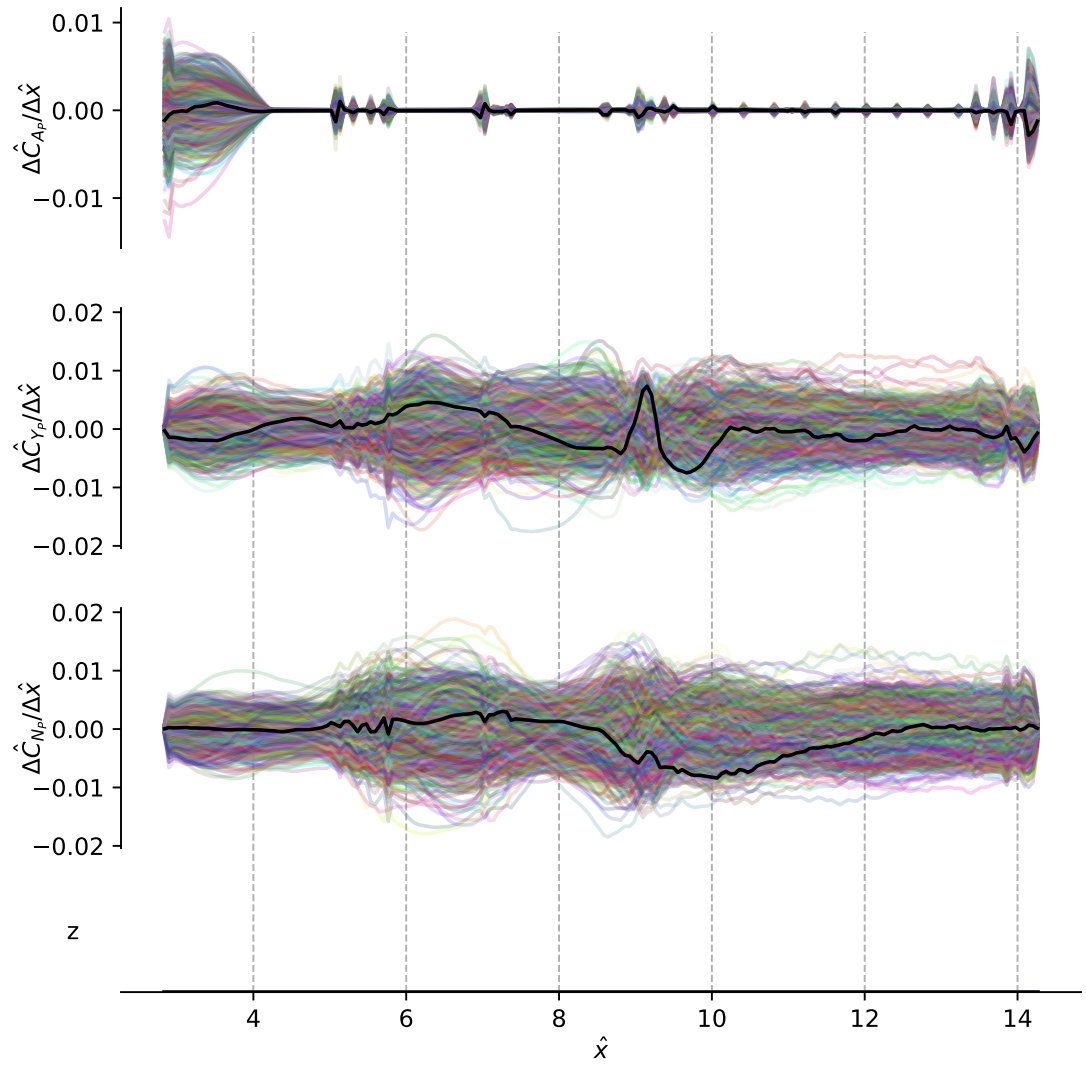


Figure 5.19 Predicted distributions on the lineloads at ϕ_P=0°.

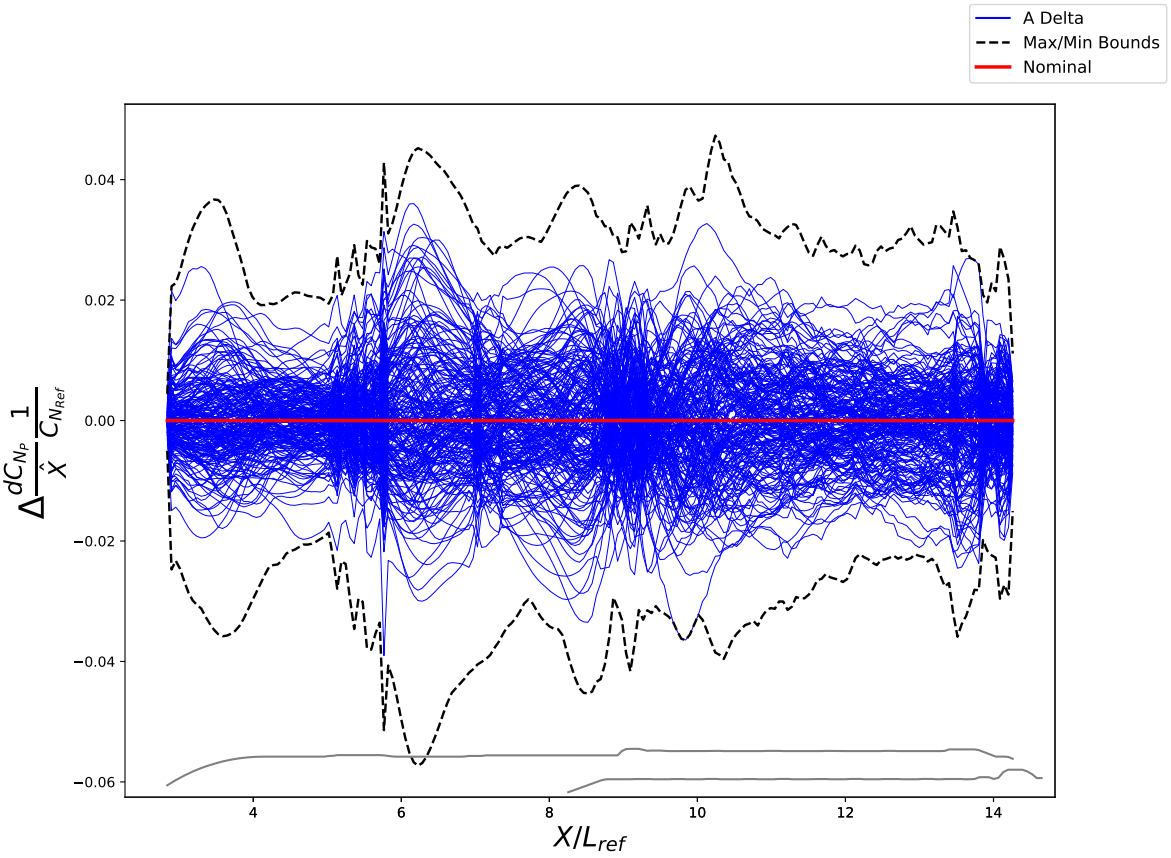


Figure 5.20 Difference between baseline ROM lineloads and the lineloads with dispersion on the ROM coefficient at ϕ_P=90°.

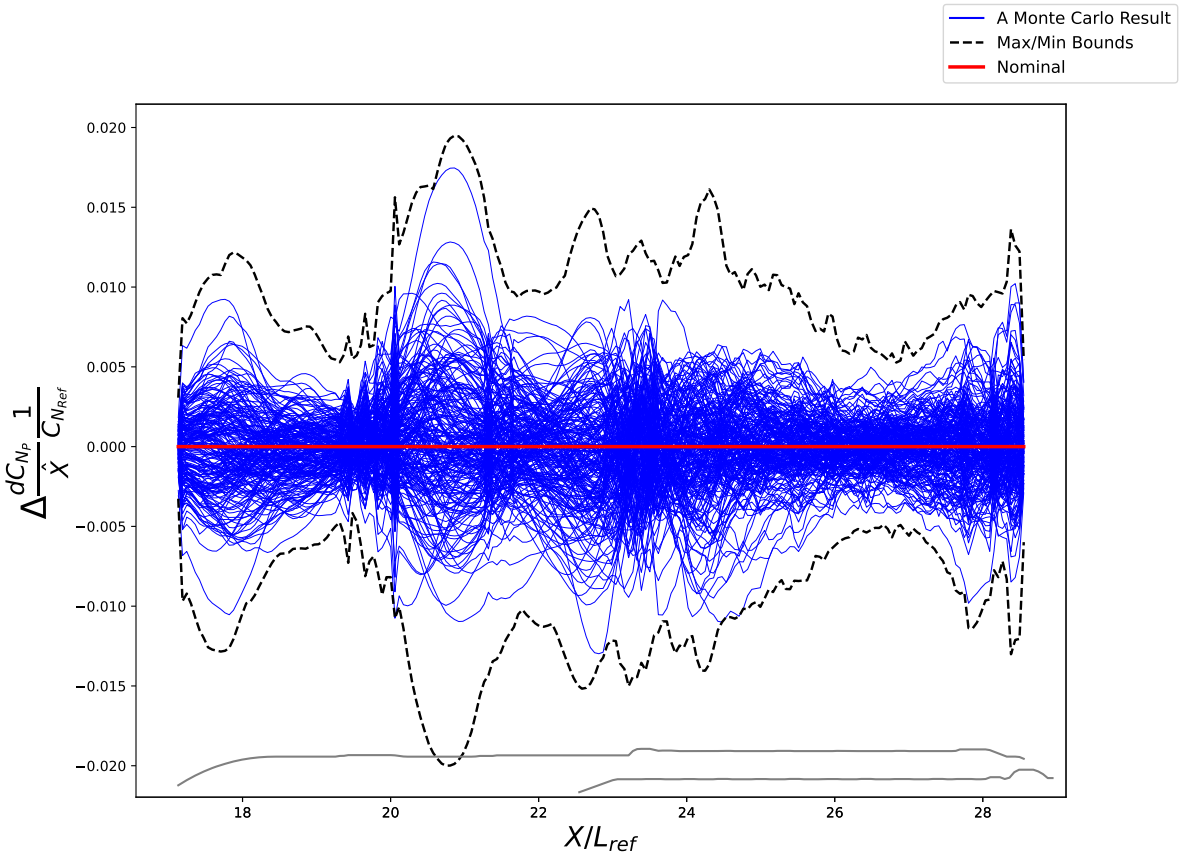


Figure 5.21 Difference between baseline DFROM lineloads and the lineloads with dispersion on the ROM coefficient at $\alpha_P = 50°$ and $\phi_P = 120°$.

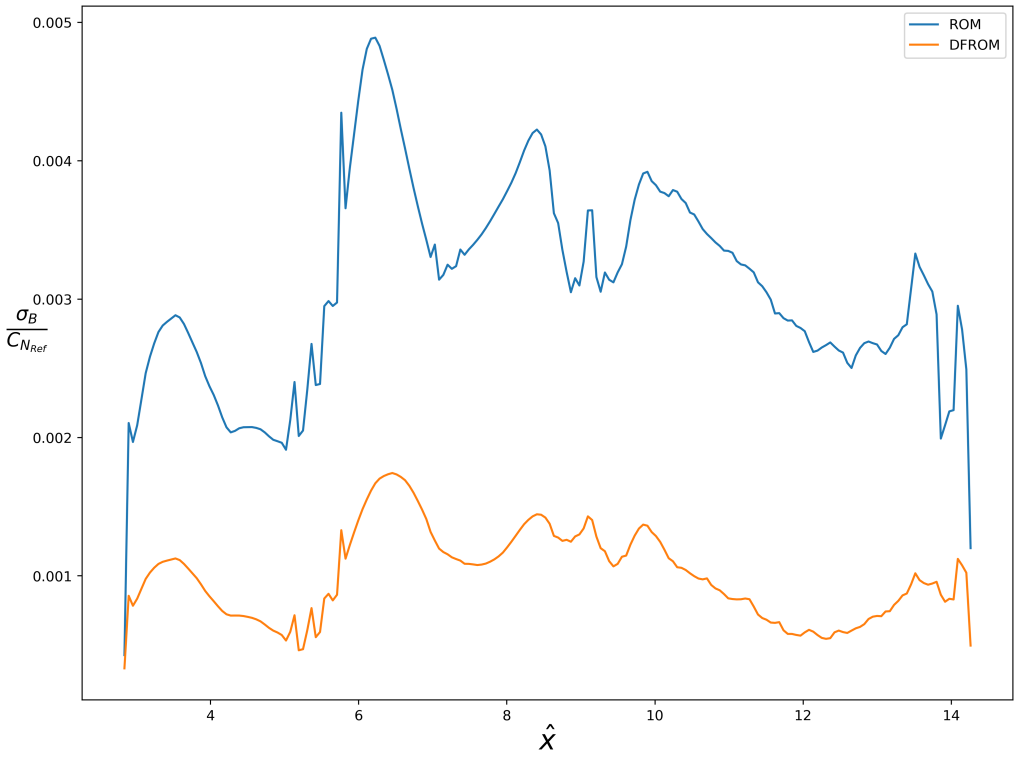


Figure 5.22 Standard deviation of the dispersed lineloads normalized by $C_{N,Ref}$ at $\alpha_P = 50°$, $\phi_P = 90°$.

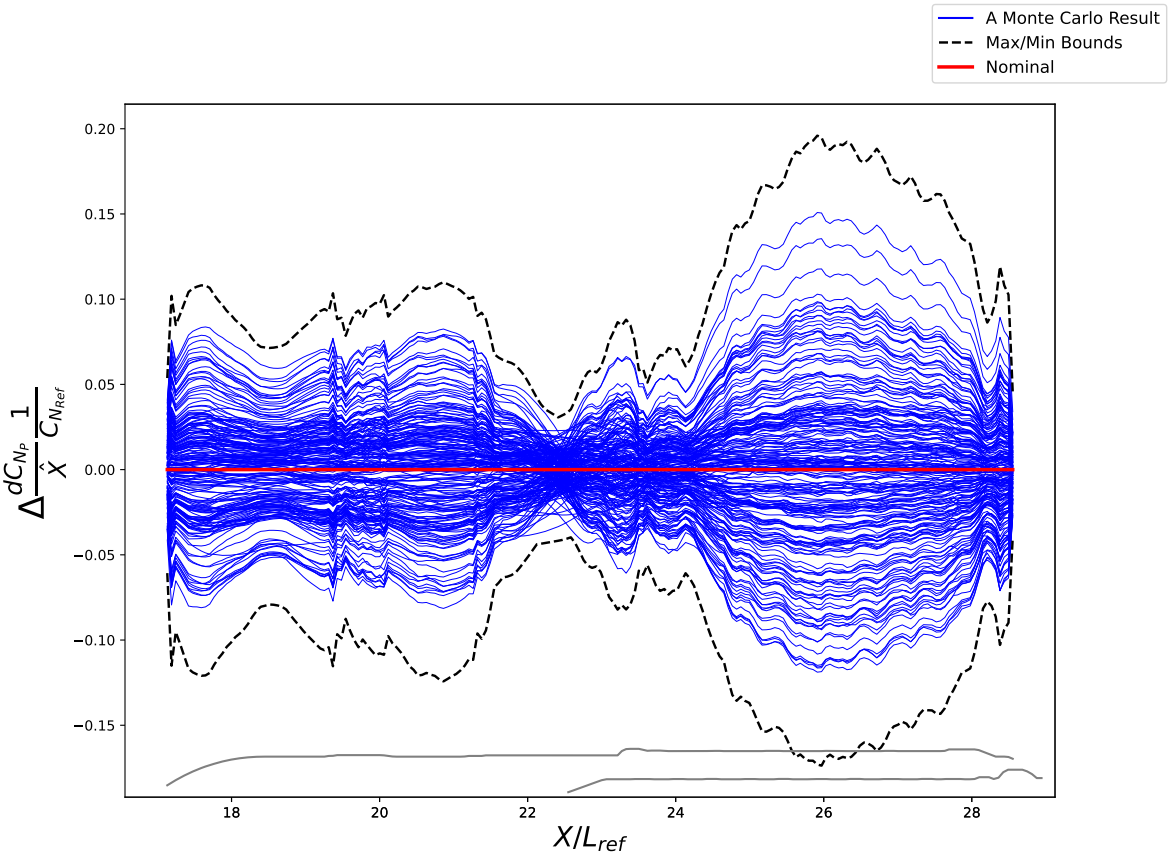


Figure 5.23 Difference between baseline DFROM lineloads and the lineloads with dispersion on both ROM coefficient and target forces and moments at α_P = 50° and ϕ_P =120°.

CHAPTER

6

CONCLUSIONS AND FUTURE WORK

A method for combining spatially resolved computational data and parametrically resolved wind tunnel data has been developed. This method allows for the incorporation of the strengths of both datasets in order to address their individual weaknesses. Using these methodologies, the SLS program has managed to reduced the computational cost for database development by half. There is yet a significant amount of further work needed to bring this new technique to maturity and address some shortcomings.

Limitations of the method have been discussed during the development, but a summary as well as possible approaches to address them are mentioned here. One of the first tasks would be to bring these methods to other configurations, flow regimes, and problems. A couple of candidates are considered and a path forward is outlined. This could be done simultaneously while working on improving the methods developed so far. The improvements undergoing active development are an expansion of the UQ methods to incorporate the uncertainty of the computational side of the work and incorporation of more physics into the constraints used.

6.1 Limitations of the Methods

One of the largest limitations of the developed methodology is the reliance on a linear transformation. While linear methods are very robust and are often capable of significant predictive power, they do run into issues when multiple modes are possible. The geometry used is plagued by many of these at the conditions being considered. The other limitation is making sure the sources of data are compatible. As flow conditions change, the dominant flow physics and features are expected to also change. At some point, the differences between sources of data are expected to be great enough that merging is no longer possible.

This methodology as outlined relies on a linear transformation, which makes it difficult for it to model nonlinear phenomena. While few configurations are plagued by slender body asymmetric vortices, fluid dynamics has many other examples of multimodal behavior. If sufficient data can be gathered, multiple models can be developed to account for these difficulties; however, the cost of collecting these extra data can be prohibitive. The cost increases significantly when the mechanism to control the different states are unknown and must be treated as stochastic. Improving the ROM techniques to account for nonlinearities or multimodal behavior would create significant improvements in the ROM as well at the expense of modeling these behaviors.

If sufficient data are gathered, categorization of each solution into the different possibilities and generating a ROM for each possibility is a mathematically simple approach. In practice, some multimodal behavior is extremely difficult to capture reliably. Recent work has focused on trying to fully characterize the Coandă effect that is present due to the gap flow between the centerbody and the SRBs [Wal; Wiga]. Despite the plethora of data collected by the SLS program, no means of repeatably generating a specific Coandă state has been found.

The path forward is most likely found in the feature extraction portion of the work. During the feature extraction step (the PCA in this work) different kernel methods can be explored with a focus on techniques that project the data into a higher dimensional space, which can cause the multimodality and nonlinearities to become linear in the different space. While a more complex reconstruction could also be a practical way to address this, it would require reconsidering the data fusion portion of the methodology.

One of the open questions regarding data fusion is how similar do different datasets need to be to allow fusion. For this work, the CFD and wind tunnel results have the same Reynolds number and other flow parameters, but often CFD is used to generate data at flight Reynolds numbers or other conditions not achievable in the wind tunnel. This becomes more important in configurations and flight conditions where Reynolds number variation causes significant changes in flow features. A possible method to overcome this could be creating a multistep data fusion process that uses an intermediate dataset, which would allow generation of a bridge function. For example, the data fusion could be developed using CFD that exactly matches the wind tunnel conditions and then applied to the flight conditions that can be simulated in CFD. These multiple datasets could also support improvements to the UQ process.

6.2 Expansion to Other Work

This DFROM technique is not only solver agnostic as mentioned during development, but there is nothing about it that is inherent to the current geometry or flow regime. The simplest expansion of this new technique developed in this dissertation could be applied to new configurations and new flow regimes. There are several candidate configurations and flows that have a wealth of both computational and experimental data that could serve as additional test cases to validate the method. To begin with, SLS goes through many different flow regimes during launch and could be a prime candidate for further development. NASA has several other open configurations that have both experimental and computational data available.

The SLS flies from the subsonic regime explored in this work through the transonic regime and into the supersonic regime. This gives an opportunity to explore these methods on the same geometry while exploring how these different regimes affect the DFROM process. Extensive testing has been performed in the NASA Ames Unitary Plan Wind Tunnel on the SLS and they have been accompanied by extensive computational simulations. While the focus of aerodynamics in these two other flow regimes is focused on navigation controller development and less on loading, accurate lineloads are still important. As explained in Section 2.6.2, the Ames SLS CFD team has already attempted this problem and provided inspiration for this work as well as some of the weighting terms explored in the data fusion process. By bringing these newly developed methods to the transonic and supersonic flight regime, the methods can be compared more directly.

Another candidate vehicle for expanding these methods is the NASA Orion spacecraft. This spacecraft is designed to carry humans on deep space exploration missions such as the upcoming Artemis launches to the Moon. A significant portion of the aerodynamic work related to the Orion spacecraft deals with entry, descent, and landing (EDL). Like the LOT domain of SLS, this regime is characterized by large wake flows that can be hard to capture in CFD, which makes it a prime candidate for this type of data fusion. For the hypersonic flow regime, expansions to the method to account for reacting flow will be necessary.

The most promising expansion to an open configuration is the NASA Common Research Model (CRM). This configuration, which is representative of a widebody commercial transport aircraft (similar to the Boeing 777), has been the subject of numerous wind tunnel tests throughout the world and is usually the configuration used in the AIAA Drag Prediction workshops. Primarily, these data are for the transonic flow regime, but a wide range of data are available at varying speeds, Reynolds numbers and angles of attack and sideslip. Much of this rich dataset is publicly available making it a prime testbed for this type of data fusion development work.

While this method was developed for building aerodynamic databases, there are other uses for it. Currently, ROMs are often used in optimization problems because of their ability to explore large parameter spaces cheaply. It would be a natural extension to use the DFROM presented here as a way to find optimums. A possible downside is reliance on experimental data, which tends to require significant planning making it less useful for early stage explorations as ROMs are currently used in practice. This can be overcome by using the ROM to inform the run matrix of the testing to best inform where the model is most sensitive to changes.

6.3 Expansion of the Methods

While this basic method has been employed to develop aerodynamic databases for the SLS program, there is no reason further improvements cannot be made. There are three major avenues to improvements. The first is modifying the feature extraction, which, as mentioned earlier, may be a path to improved handling of nonlinearities. The second is by improving the physical constraints used in the optimization problem used to apply the data fusion. The final is further improvements in the uncertainty quantification.

In the current work, the basis for the ROM was extracted from the source data using PCA. PCA already has been expanded in the literature using such methods as Kernel PCA and multilinear PCA. These methods tend to do better with nonlinearities which as discussed is a significant issue in the LOT problem because of the asymmetric vorticies. Other possibilities include using methods developed for use with computer vision to recognize and extract cohesive shapes from the computational data. While the features extracted are likely to have physical meaning, a methodology to recombine them is not readily available compared to PCA-like methods. Significant work would have to be done to develop a methodical way of recombining these computer extracted forms into a usable model.

6.3.1 Increase Physical Constraints

The current methods as developed do not incorporate significant physical knowledge or data of the system. By tweaking of the constraints used, in the data fusion step, knowledge of the underlying physics can be brought in. Since a goal of the model is to create accurate lineloads, the pressure and friction terms can be further coupled to make sure their relationship satisfy the Navier-Stokes Equations boundary conditions at the wall. Right now, there is no attempt to do so and as a result we see significant changes in the friction terms especially in the axial direction, which creates data that does not necessarily satisfy the expected physics of a fluid flow.

Another avenue of exploration worth following up is using the pressure measurement data collected during wind tunnel testing. Previous experiments with forcing the DFROM to match the surface pressure exactly proved to constrain the system in a way that gave unrealistic answers. For example, large pressure gradients were introduced too. It is possible to perhaps create some form of least squares weighting or constraint, which is similar to the ideas behind Ref. [Mif19]. To implement this constraint, a balancing in the resulting optimization problem would likely result in an ad-hoc weighting between the forced constrains on the integrated forces and moments and these discrete pressure measurements.

6.3.2 Expand Uncertainty Quantification

One of the potential avenues for follow on work is to further expand the uncertainty model. Modeling lineloads using a technique that can generate sample functions such as Gaussian process regressors could be of use. Sample functions can be generated that can be used in a Monte Carlo type analysis, which recovers the point statistics along the axial station of the vehicle while providing different shapes. The bootstrap UQ model used a Laplace distribution; however, improvements in categorizing that distribution is possible. Further work on the simplified fighter geometry or other simpler problems could lead to better understanding of the bootstrap method as implemented.

In this work, the confidence in the computational data are not explored. In general the characterization of computational results is an expanding field with significant research being put into it. A full characterization of the uncertainties for this flow problem is an enormous undertaking, but there are some relatively low hanging fruit that could be explored. An initial way to explore sensitivities to the quality of the CFD data is to examine how sensitive the derived ROM coefficients are to the input data. While the bootstrap UQ method showed one way to look at this sensitivity, another possible way to

perturb the CFD surfaces themselves. The difficulty here would be finding a meaningful way to perturb the solution across the domain. Another option is to make use of instantaneous snapshots instead of the time-averaged results that were used. However, if the coefficients remain constant despite significant changes in the input data, it would provide greater confidence in the ROM and subsequent DFROM. In general, incorporation of instantaneous data could provide further improvements of the end model with respect to quantification. The time-accurate data includes many lineload shapes and shows how a peak in one section usually results in a valley in another location as flow features travel downstream. If this is captured in the UQ model, a significant limitation of the derived methods would be addressed.

BIBLIOGRAPHY

[AF16] Aaron Fisher Brian Caffo, B. S. & Zipunnikov, V. "Fast, Exact Bootstrap Principal Component Analysis for p > 1 Million". *Journal of the American Statistical Association* **111**.514 (2016). PMID: 27616801, pp. 846–860. eprint: https://doi.org/10.1080/01621459.2015.1062383.

[Aru00] Arunajatesan, S. et al. "On The Application Of Hybrid Rans-Les And Proper Orthogonal Decomposition Techniques To Control Of Cavity Flows". *Proceedings of the Third AFOSR International Conference on DNS/LES*. 2000.

[Ber93] Berkooz, G. et al. "The proper orthogonal decomposition in the analysis of turbulent flows". *Annual review of fluid mechanics* **25**.1 (1993), pp. 539–575.

[Bib11a] Bibb, K. et al. "Development of the Orion Crew Module Static Aerodynamic Database, Part I: Hypersonic". *29th AIAA Applied Aerodynamics Conference*. AIAA, 2011.

[Bib11b] Bibb, K. et al. "Development of the Orion Crew Module Static Aerodynamic Database, Part II: Subsonic/Supersonic". *29th AIAA Applied Aerodynamics Conference*. AIAA, 2011.

[Ble14] Blevins, J. A. et al. "An Overview of the Characterization of the Space Launch System Aerodynamic Environments". *52nd Aerospace Sciences Meeting*. AIAA SciTech Forum 0. American Institute of Aeronautics and Astronautics, 2014.

[Bra11] Braconnier, T. et al. "Towards an adaptive POD/SVD surrogate model for aeronautic design". *Computers Fluids* **40**.1 (2011), pp. 195–209.

[Buc] Buchmeier, J. et al. "Parametric studies of subsonic and supersonic reentry phases of the Orion Crew Module through numerical simulations". *AIAA SCITECH 2023 Forum*. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2023-1375.

[BT] Bui-Thanh, T. et al. "Proper Orthogonal Decomposition Extensions for Parametric Applications in Compressible Aerodynamics". *21st AIAA Applied Aerodynamics Conference*. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2003-4213.

[Car04] Caraballo, E. et al. "A Study of Subsonic Cavity Flows - Low Dimensional Modeling". *2nd AIAA Flow Control Conference*. American Institute of Aeronautics and Astronautics, 2004.

[Car15] Carlson, H. et al. "Reduced-order Model for NASA Space Launch System Liftoff Aerodynamics". *AIAA 2015-0777* (2015).

[Cas13] Castanedo, F. "A Review of Data Fusion Techniques". *The Scientific World Journal* **2013** (2013), p. 19.

[Cha19] Chan, D. T. et al. "Aerodynamic Characterization and Improved Testing Methods for the Space Launch System Liftoff and Transition Environment". *AIAA Aviation 2019 Forum*. 2019. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2019-3398.

[Cha18] Chan, W. et al. "Chimera Grid Tools User's Manual, Version 2.2". *NASA Ames Research Center* (2018).

[Dal18] Dalle, D. J. et al. “Adjustments and Uncertainty Quantification for SLS Aerodynamic Sectional Loads”. *2018 Applied Aerodynamics Conference*. American Institute of Aeronautics and Astronautics, 2018.

[Dav] Davis, S. & Askins, B. “Ares I-X: First Flight of a New Generation”. *46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2010-6910.

[DW88] Durrant-Whyte, H. F. “Sensor Models and Multisensor Integration”. *The International Journal of Robotics Research* **7**.6 (1988), pp. 97–113. eprint: https://doi.org/10.1177/027836498800700608.

[Efr82] Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, 1982. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611970319.

[Eri20] Erichson, N. B. et al. “Shallow neural networks for fluid flow reconstruction with limited sensors”. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **476**.2238 (2020), p. 20200097.

[Fav16] Favaregh, A. L. et al. “Space Launch System Aerodynamic Database Uncertainty Quantification Methodologies”. *54th AIAA Aerospace Sciences Meeting*. AIAA SciTech Forum 0. American Institute of Aeronautics and Astronautics, 2016.

[Fid75] Fidler, J. & Bateman., M. “Asymmetric Vortex Effects on Missile Configurations”. *Journal of Spacecraft and Rockets* **12**.11 (1975), pp. 674–681.

[Gan13] Ganguli, R. et al. “Interpolation of Transonic Flows Using a Proper Orthogonal Decomposition Method”. *International Journal of Aerospace Engineering* **2013** (2013), p. 928904.

[Har90] Hartwich, P. et al. “Navier-Stokes Computations of Vortex Asymmetries Controlled by Small Surface Imperfections”. *AIAA 90-0395* (1990).

[Hem] Hemsch, M. et al. “Detailed Uncertainty Analysis for Ares I Ascent Aerodynamics Wind Tunnel Database”. *26th AIAA Aerodynamic Measurement Technology and Ground Testing Conference*. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2008-4259.

[Hemaa] Hemsch, M. J. & Houlden, H. P. “Repeatability Modeling for Wind-Tunnel Measurements: Results for Three Langley Facilities”. *52nd Aerospace Sciences Meeting*. AIAA 2014-0096. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2014-0096.

[Hot33] Hotelling, H. “Analysis of a complex of statistical variables into principal components”. *Journal of Educational Psychology* **24**.6 (1933), pp. 417–441.

[Hou] Houlden, H. et al. “Quantification of the Uncertainties for the Ares I A106 Ascent Aerodynamic Database”. *27th AIAA Aerodynamic Measurement Technology and Ground Testing Conference*. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2010-4926.

[Hun] Hunter, C. et al. “Advanced aerodynamic design of passive porosity control effectors”. *39th Aerospace Sciences Meeting and Exhibit*. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2001-249.

[JD88] J.J. Daudin, C. D. & Trecourt, P. “Stability of principal component analysis studied by the bootstrap method”. *Statistics* **19**.2 (1988), pp. 241–258. eprint: https://doi.org/10.1080/02331888808802095.

[Jol02] Jolliffe, I. *Principal Component Analysis*. 2nd ed. Springer, 2002.

[Koh95] Kohavi, R. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”. Morgan Kaufmann, 1995, pp. 1137–1143.

[Kun02] Kunisch, K. & Volkwein, S. “Galerkin Proper Orthogonal Decomposition Methods for a General Equation in Fluid Dynamics”. *SIAM Journal on Numerical Analysis* **40**.2 (2002), pp. 492–515. eprint: https://doi.org/10.1137/S0036142900382612.

[Leb] Lebedev, Y. et al. “Construction of a Fluid Flow Field from Discrete Point Data using Machine Learning”. *AIAA SCITECH 2023 Forum*. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2023-1186.

[Lee10] Lee, K. “Investigation Of Probabilistic Principal Component Analysis Compared To Proper Orthogonal Decomposition Methods For Basis Extraction And Missing Data Estimation”. PhD thesis. Georigia Institute of Technology, 2010.

[Lee] Lee, M. W. & Dowell, E. H. “On the Importance of Numerical Error in Constructing POD-based Reduced-Order Models of Nonlinear Fluid Flows”. *AIAA Scitech 2020 Forum*. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2020-1067.

[Lee22] Lee, M. W. & Wignall, T. “PCLAM: a Python Module for Computing Surface Lineloads and Moments”. Version 20225189. *NASA Technical Memorandum* (2022).

[Lum67] Lumley, J. L. *The Structure of Inhomogenous Turbulence*. Atmospheric Turbulence and Wave Propagation, Yaglom AM, Tatarski VI (eds). Nauka: Moscow. 1967, pp. 166–178.

[Mal22] Malik, M. R. et al. “Principal component analysis based combustion model in the context of a lifted methane/air flame: Sensitivity to the manifold parameters and subgrid closure”. *Combustion and Flame* **244** (2022), p. 112134.

[Mif19] Mifsud, M. et al. “Fusing wind-tunnel measurements and CFD data using constrained gappy proper orthogonal decomposition”. *Aerospace Science and Technology* **86** (2019), pp. 312–326.

[Mor09] Morton, S. A. et al. “Kestrel – A Fixed Wing Virtual Aircraft Product of the CREATE Program”. *AIAA 2009-338* (2009).

[Nasa] *NASA High-End Computing Capability*.

[Nasb] *NASA Space Launch System*. Available at https://www.nasa.gov/exploration/systems/sls/index.html. Accessed: 10/15/2021.

[O'N94] O'Neil, P. J. et al. *Aero Configuration / Weapons Fighter Technology (ACWFT) - Summary Technical Report*. Tech. rep. WL-TR-95-3002. Wright Labs, 1994.

[Pana] Pandya, M. J. et al. "Accuracy, Scalability, and Efficiency of Mixed-Element USM3D for Benchmark Three-Dimensional Flows". *AIAA Scitech 2019 Forum*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2019-2333`.

[Panb] Pandya, M. J. et al. "Toward Verification of USM3D Extensions for Mixed Element Grids". *31st AIAA Applied Aerodynamics Conference*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2013-2541`.

[Pan17] Pandya, M. J. et al. "Assessment of USM3D Hierarchical Adaptive Nonlinear Method Preconditioners for Three-Dimensional Cases". *AIAA Journal* **55**.10 (2017), pp. 3409–3424. eprint: `https://doi.org/10.2514/1.J055823`.

[Pan11] Pandya, S. & Chan, W. "Computation of Sectional Loads from Surface Triangulation and Flow Data". *20th AIAA Computational Fluid Dynamics Conference*. American Institute of Aeronautics and Astronautics, 2011.

[Pea01] Pearson, K. "LIII. On lines and planes of closest fit to systems of points in space". *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**.11 (1901), pp. 559–572. eprint: `https://doi.org/10.1080/14786440109462720`.

[Pin12] Pinier, J. T. "New Aerodynamic Data Dispersion Method with Application to Launch Vehicle Design". *Journal of Spacecraft and Rockets* **49**.5 (2012), pp. 834–841. eprint: `https://doi.org/10.2514/1.A32219`.

[Pin19] Pinier, J. T. et al. "Advances in the Characterization of NASA's Space Launch System Aerodynamic Environments". *AIAA Aviation 2019 Forum*. 2019. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2019-3397`.

[Pom] Pomeroy, B. W. "Validation of Kestrel IDDES Simulations for SLS Transition Analysis". *AIAA SCITECH 2022 Forum*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2022-1335`.

[Pom23] Pomeroy, B. W. et al. "Sensitivity of Space Launch System Low-Speed Aerodynamic CFD Results to Kestrel Flow Solver Inputs". *NASA/TP–20230003697* (2023).

[Ras17] Rashinkar, P. & Krushnasamy, V. S. "An overview of data fusion techniques". *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*. 2017, pp. 694–697.

[Rataa] Ratnayake, N. A. et al. "Selection of Computational Fluid Dynamics Tools Used in Development of the Space Launch System Liftoff and Transition Lineloads Databases". *57th AIAA Aerospace Sciences Meeting*. AIAA, 2019.

[Rei06] Reisenthel, P. H. et al. "Innovative fusion of experiment and analysis for missile design and flight simulation". *Proceedings of the RTO AVT-135 Symposium on Innovative Missile Systems*. 2006.

[Rey20] Reyes, R. & Codina, R. “Projection-based reduced order models for flow problems: A variational multiscale approach”. *Computer Methods in Applied Mechanics and Engineering* **363** (2020), p. 112844.

[Ril18] Riley, L. P. et al. “Transient Feature Extraction in Simulated Scramjet Unstart via Model Order Reduction Methods”. *2018 Joint Propulsion Conference*. 2018, p. 4535.

[Sco] Scoggins, J. B. et al. “Multihierarchy Gaussian Process Models for Probabilistic Aerodynamic Databases using Uncertain Nominal and Off-Nominal Configuration Data”. *AIAA SCITECH 2023 Forum*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2023-1185`.

[She19] Shea, P. R. et al. “Ascent Aerodynamic Force and Moment Database Development for the Space Launch System”. *AIAA Aviation 2019 Forum*. AIAA AVIATION Forum 0. American Institute of Aeronautics and Astronautics, 2019.

[Sil18] Silva, W. A. “AEROM: NASA’s Unsteady Aerodynamic and Aeroelastic Reduced-Order Modeling Software”. *Aerospace* **5**.2 (2018).

[Spe] *SLS-SPEC-048: Cross-Program Integrated Coordinate Systems*. Revision F.1, 2018.

[Smi13] Smith, R. C. *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM Computational Science and Engineering, 2013.

[Sny23] Snyder, S. et al. “AeroFusion: Data Fusion and Uncertainty Quantification for Entry Vehicles”. *AIAA SCITECH 2023 Forum*. AIAA SciTech Forum 0. American Institute of Aeronautics and Astronautics, 2023.

[Sut09] Sutherland, J. C. & Parente, A. “Combustion modeling using principal component analysis”. *Proceedings of the Combustion Institute* **32**.1 (2009), pp. 1563–1570.

[Tai17] Taira, K. et al. “Modal Analysis of Fluid Flows: An Overview”. *AIAA Journal* **55**.12 (2017), pp. 4013–4041.

[Vir20] Virtanen, P. et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. *Nature Methods* **17** (2020), pp. 261–272.

[Wal] Walker, M. A. et al. “Experimental Identification of Bistable Flow States on the Space Launch System at Liftoff Conditions”. *AIAA AVIATION 2022 Forum*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2022-3665`.

[Wiga] Wignall, T. J. et al. “Experimental and Computational Examination of the Coandă Effect on the Space Launch System at Liftoff Conditions”. *AIAA SCITECH 2023 Forum*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2023-0648`.

[Wigaa] Wignall, T. “Liftoff and Transition Database Generation for Launch Vehicles Using Data-Fusion-Based Modeling”. *AIAA Aviation Forum*. AIAA, 2019.

[Wigb] Wignall, T. et al. “Predicting and Accelerating Chemistry in High Speed Reacting Flows”. *54th AIAA Aerospace Sciences Meeting*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2016-1689`.

[Xu20] Xu, J. & Duraisamy, K. “Multi-level convolutional autoencoder networks for parametric prediction of spatio-temporal dynamics”. *Comput. Methods Appl. Mech. Eng.* **372** (2020).

[Zil16] Zilberter, I. “Hybrid Large Eddy Simulation / Reynolds Averaged Navier-Stokes Modeling in Directed Energy Applications”. PhD thesis. NCSU, 2016.

[Zim14] Zimmermann, R. et al. “Reduced-Order Modeling of Steady Flows Subject to Aerodynamic Constraints”. *AIAA Journal* **52**.2 (2014), pp. 255–266.

APPENDIX

APPENDIX

A

DEVELOPMENT OF BOOTSTRAP ON SIMPLE CASE

A test case is used to explore this methodology for a more benign problem not plagued by multimodal behavior. The geometry used is a simplified tactical fighter (STF) configuration developed under the Air Force Wright Lab [O'N94]. It is a simplified semispan planform used in studies for fighter craft-like configurations [Hun]. Unlike the rest of the work in the dissertation, the results were generated using the NASA CFD solver USM3D. USM3D is a finite volume, cell centered, unstructured grid solver developed at NASA Langley [Pan17; Panb; Pana]. The problem is a standard test case provided by the developers of USM3D and only inputs that were changed from the example were α and M. The grid is made up of 1,511,114 mixed elements and were run on the NASA Langley K-Cluster using 80 cores and took about 15 minutes each. Simulations used the SA turbulence model and evaluated for 1000 iterations, which dropped the initial residuals by at least 10 orders of magnitude. Figure A.1 shows the pressure on the STF at M =0.75, α = 3°.

Like the SLS problem, two parameters were varied to collect data. The first being α, which was varied between -7° and 7°, and the second being M, which varied from 0.3 to 0.9. Data was generated at ΔM =0.1 and at α = -7°, -5°, -3°, -1°, 0°, 1°, 3°, 5°, and 7°, which results in 63 solutions. Originally more solutions were used, but eventually the system was no longer linearly independent and so approximately

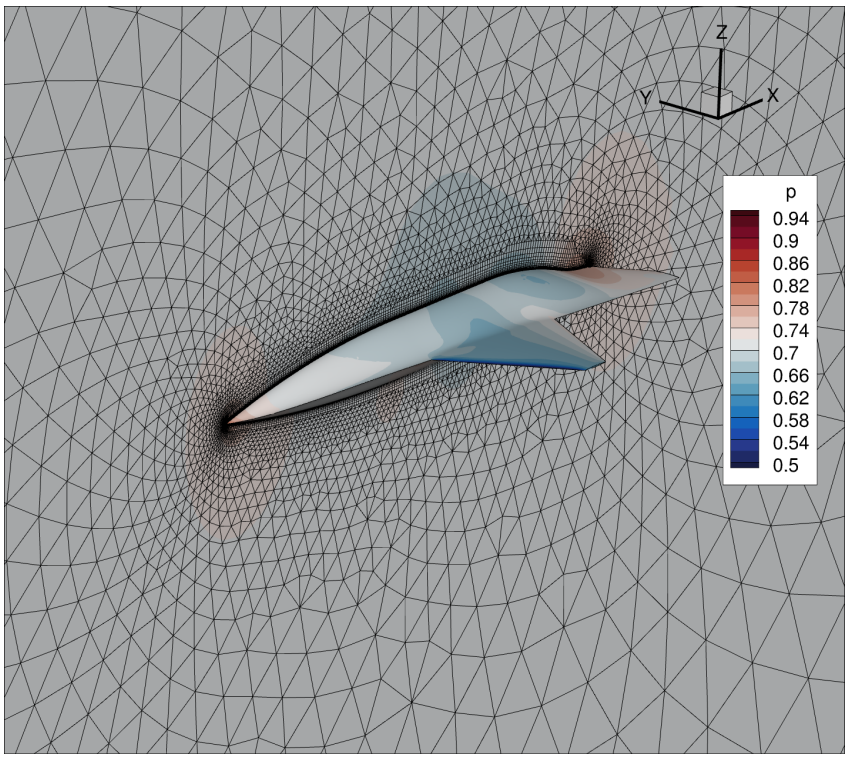


Figure A.1 Surface pressures on the STF at M =0.75, α = 3°

half the solutions were discarded. Instead of halving the number of data points it may have also been possible to do some form of regularization. Unlike the SLS, this problem does not suffer from multimodal behavior, which helps eliminate interacting effects in the analysis.

The ROM was developed using the same code and techniques as used on the SLS geometry. The reduced size in the surface grid (14,753 nodes) makes the problem much cheaper to solve and explore new methods. Also due to the simpler aerodynamics, instead of removing one snapshot, multiple snapshots are removed. The bootstrapping processes was ran four times with 10, 20, 30, and 40 snapshots removed and with 100 different permutations each. Model evaluations were at ΔM =0.025 and $\Delta\alpha$ = 0.5°. Any values for B that were less than 1e-14 were excluded from the analysis. This filters out values associated with the trivial solution where CFD data is available. The focus of this is to explore the distributions on the ROM coefficients and the results on quantities of interest are not explored.

The results of the least squares problem as outlined in Sec. 5.4.2 are plotted in the next couple of figures. These values are pooled across the parameter space. Figure A.2 shows the distribution of $B_{C_P,1}$ which is the variations of the coefficient associated with the first PC and pressure. The different colors represent the different number of dropped snapshots and the values are stacked to give information about the combined distribution. On the left is the absolute values fitted with an exponential distribution while the right shows the natural log of the absolute values fitted by a normal distribution, which is equivalent to a lognormal distribution. As the number of dropped snapshots increases, the distribution shifts to the right. As more information is removed from the system, the residuals increase which means the ROM coefficients perturbations must also increase to accommodate the larger errors. While the 20, 30, and 40 sets are well represented by a lognormal distribution, the Bootstrap 10 results have a long tail on the left. Ideally when doing sensitivity analysis, perturbations are of controlled sizes and done with a continuous distribution of possible values. However, removing snapshots is a discrete perturbation that can only be made so small. These tailing values to the left are representative of smaller perturbations to the system. This evidence points to the lognormal distribution being a mirage.

Further evidence for this is seen at the higher PC numbers. Figure A.3, which shows the distribution of $B_{C_P,30}$, begins to look more bimodal with a group of values around $\ln|B_{C_P,30}| < -30$. The sharp cut off is a side effect of the filtering done to remove trivial values. As PC number increases, this secondary distribution continues to slide right as shown in Fig. A.4, which shows $B_{C_P,63}$. There is some behavior at the smaller B_i values that are not being captured by the bootstrap method.

While the exponential fit has problems, in particular with capturing the long tails, it has other beneficial quantities. It is defined for 0 and can be double sided by converting it to a Laplace distribution. It is also relatively easy to sample from, which makes doing Monte Carlo draws easier. The canonical maximum likelihood estimator for the scale factor is used, which could account for some of the mismatch seen. Adjustments to the scale factor could likely lead to a better match of the data.

These results in general show the applicability to this method on a test problem and have similar behavior to the SLS geometry. Unlike with the SLS where the number of dropped snapshots was limited to 1, a number of values were used here. While dropping more information from the system tended to spread the data, it kept many of the same characteristics pointing to some fundamental underlying distribution; however, there is likely to be some behavior missing due to the limitations of the bootstrap method applied to snapshots.

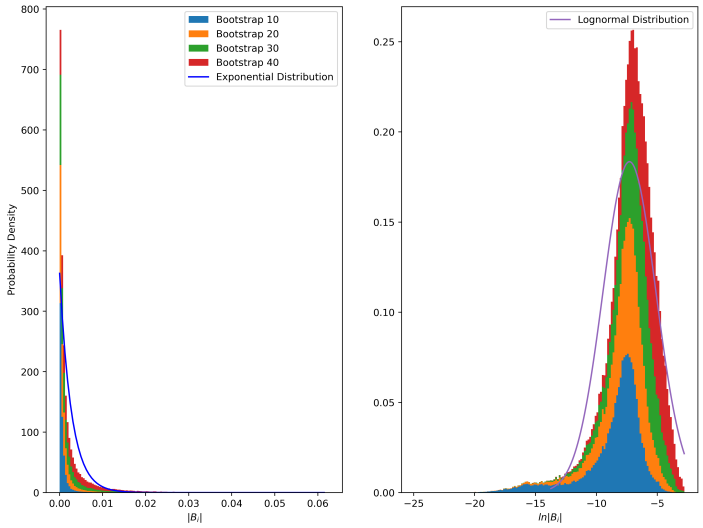


Figure A.2 Histogram of $B_{C_P,1}$. (a) shows the absolute value and (b) shows the natural log of the absolute value.

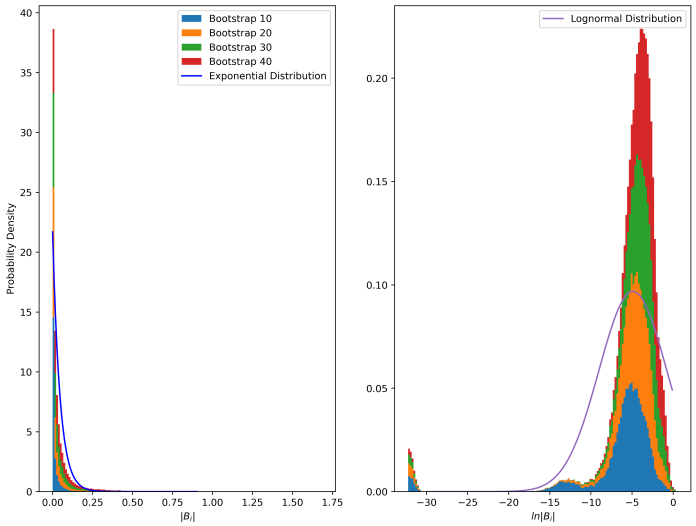


Figure A.3 Histogram of $B_{CP,30}$. (a) shows the absolute value and (b) shows the natural log of the absolute value.

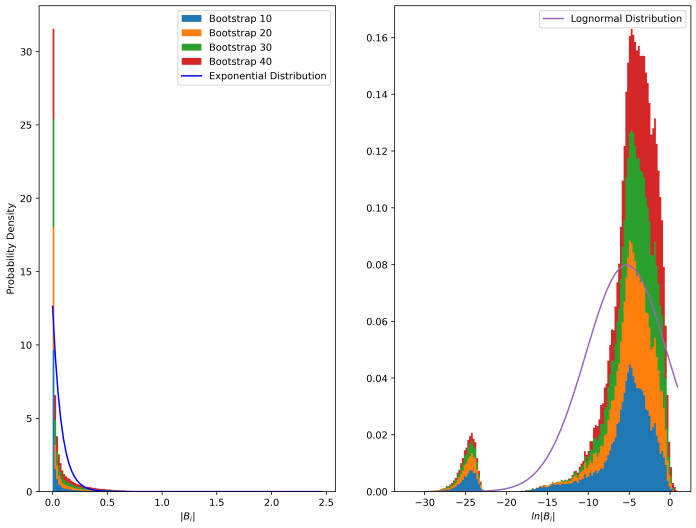


Figure A.4 Histogram of $B_{CP,63}$. (a) shows the absolute value and (b) shows the natural log of the absolute value.