Modeling of the Nonlinear Unsteady Pitching Moment Stall Characteristics from Cessna Citation II Flight Test Data

MSc Thesis

C.J. van Wezel

TU Delft

Modeling of the Nonlinear Unsteady Pitching Moment Stall Characteristics from Cessna Citation II Flight Test Data

MSc Thesis

by

C.J. van Wezel

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on September 27, 2024 at 13:00

Thesis committee:	
Chair:	Dr.Ir. D.M. Pool
Supervisors:	Dr.Ir. C.C. de Visser
	Dr.Ir. D.M. Pool
External examiner:	Dr. S.J. Hulshoff
Place:	Faculty of Aerospace Engineering, Delft

Project Duration:	September, 2023 - September, 2024
Student number:	4863755

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Faculty of Aerospace Engineering · Delft University of Technology

Preface

Six years ago I took off on my journey to become an aerospace engineer at TU Delft. However, only shortly before that moment I was not sure whether that was going to be the case. I had scored so low on a multiple choice motivation test in the admission process of the bachelor's program, that I was initially not offered a spot. A long story short, I was admitted in the end, and three years later I celebrated my bachelor graduation. The content of this work, embodies the apotheosis of another 2.5 years at TU Delft, and a full year that went into this thesis. A year where I learnt something new every day, a year of hard work, and a year of unforgettable experiences. I can only be thankful for the opportunity to work on this thesis.

This thesis was characterized by ups and downs. After completing my literature study, the course of my thesis took a different direction due to a new discovery. The ideas I had been working on for months, now had become redundant for a large part. After turning the page, I went back to the literature and discovered lots of fascinating research that I was able to apply to my own work. Furthermore, two highlights of my thesis work were surely the flight tests I was able to design and participate in. I have the PH-LAB pilots Alexander, Hans and Bernard to thank for their helpfulness and dedication to realizing these flight.

I would also like to express my gratitude towards my supervisors Daan and Coen. Every stall meeting I found myself wondering whether I had enough results to not waste your time. It never showed, and the stall meetings have been an indispensable link in the realization of this thesis. The same goes for the meetings with my daily supervisor Coen, your insights and feedback have helped enormously, so thank you. Furthermore, my gratitude goes to friends, family and anyone else who showed willingness to listen to me talk about splines, stalls and vortices. Lastly, I would like to thank Dr. Steven Hulshoff for completing my graduation committee.

Now the only thing that rests me to say, is that I wish you a lot of pleasure reading this thesis. I hope you enjoy reading it as much as I have had working on it!

Contents

Nomenclature

List of Abbreviations

BEC	Basic Empty Weight
AIC	Akaike Information Criterion
CFD	Computation Fluid Dynamics
EKF	Extended Kalman Filter
FAA	Federal Aviation Administration
FL	Flight Level
FPR	Flight Path Reconstruction
FSTD	Flight Simulator Training Devices
FTIS	Flight-Test Instrumentation System
GCV	Generalized Cross Validation
IEKF	Iterated Extended Kalman Filter
KF	Kalman Filter
LOC-I	Loss of Control in Flight
MDF	Minimum Distance Function
MIS	Matrix Iterative Solver
MOF	Multivariate Orthogonal Function
MSE	Mean Squared Error
NLR	Netherlands Aerospace Centre
OLS	Ordinary Least Squares
QR	Quasi-Random
RLS	Regularized Least Squares
SVD	Singular Value Decomposition
TSVD	Truncated Singular Value Decomposition
UKF	Unscented Kalman Filter
UPRT	Upset Prevention and Recovery Training
VIF	Variance Inflation Factor

List of Symbols

A_*	specific force in * direction [m/s^2]
A_t	simplex volume
$\boldsymbol{a}$	directional coordinate
a_1	Kirchhoff's flow separation shaping parameter [-]
$\boldsymbol{B}_t^d(\boldsymbol{b})$	vector of basis polynomials on simplex t
$B_\kappa^d(\boldsymbol{b})$	single basis polynomial of degree d
$\overline{b}$	Wing span [m]
$b(\boldsymbol{x})$	Barycentric coordinates of the point x
b_i	single barycentric coordinate component
C_*	aerodynamic force/moment coefficient [-]
$\hat{C}_*$	aerodynamic force/moment coefficient model estimate [-]
C_p	Mallow's C_p
$\boldsymbol{c}$	global B-coefficient vector; model parameter vector
$\hat{c}$	global B-coefficient vector estimate
$\overline{c}$	average chord length [m]
c_κ	single B coefficient
c_κ^t	single B coefficient on simplex t
$\boldsymbol{c}_\mu$	regularized model parameter estimate
$\boldsymbol{c}_b$	global B-coefficient in Barycentric space
$\boldsymbol{c}_p$	global B-coefficient in physical space
$\boldsymbol{c}_{ridge}$	ridge regression model parameter estimate
$\boldsymbol{c}_t$	B-coefficient vector on simplex t

$\boldsymbol{D}_u^k$	difference matrix in direction u with order k
$\boldsymbol{D}_c$	difference matrix B-coefficient c
$\boldsymbol{D}_g$	global difference matrix
$\boldsymbol{D}_k$	discrete differential operator order k
$\boldsymbol{D}_t$	difference matrix simplex t
d	polynomial degree
$\hat{d}$	number of basis functions per simplex
f	filter factor
$\boldsymbol{G}$	global B-coefficient matrix
g_f	gradient of function f
$\boldsymbol{H}$	global smoothness matrix; hat-matrix
$\boldsymbol{H}_f$	Hessian of function f
h	altitude [m]
I_*	angular moment of inertia around * axis [kg m^2]
k	difference order
$\boldsymbol{L}$	semi-norm
L_1	distance true and estimated parameter vector
M	Mach [-]
m	mass [kg]; number of model parameters
N	number of data points
n	number of dimensions
$\boldsymbol{P}$	penalty matrix
$\boldsymbol{P}^{d,d-m}$	degree m Casteljau matrix
p	roll rate [rad/s]
$\boldsymbol{Q}$	dispersion matrix
q	pitch rate [rad/s]; univariate B-spline degree
$\overline{q}$	dynamic pressure [kg/m^3]
R	coefficient of determination
r	yaw rate [rad/s]; continuity order
S	wing surface [m^2]; Secondary global penalty matrix
T	total number of simplices
$\mathcal{T}$	Triangulation
t	simplex
$\tilde{t}$	edge simplex
$\boldsymbol{U}$	left singular vector matrix
$\boldsymbol{V}$	airspeed [m/s]; right singular vector matrix
V_g	GCV score
V_{stall}	stall speed [m/s]
V_{TAS}	True air speed [m/s]
v	simplex vertex
$\boldsymbol{W}$	B-coefficient constraint matrix
$\boldsymbol{X}$	point of flow separation variable [-]; regression matrix
x	aircraft state vector; univariate spline knot
X_0	steady point of flow separation [-]
x_e, y_e, z_e	Cartesian position in F_E [m]
$\dot{x}_e$, $\dot{y}_e$, $\dot{z}_e$	velocity components in F_E [m/s]
$\boldsymbol{Y}$	observation vector
$\hat{\boldsymbol{Y}}$	model prediction vector
y_w	moment arm of single-wing lift vector
α	angle of attack [rad]
α^*	Kirchhoff's flow separation delay parameter [rad]
β	angle of side slip [rad]
Γ	Tikhonov matrix
γ	multi-index
γ_1	total parameter variance

γ_2	total parameter squared bias
Δ	difference operator
δ_a	aileron deflection [rad]
δ_e	elevator deflection [rad]
δ_r	rudder deflection [rad]
ζ	multi-index difference offset
θ	model parameter vector; pitch angle [rad]
κ	multi-index
$\mathbf{\Lambda}$	global Barycentric to Cartesian transformation matrix
$\mathbf{\Lambda}_t$	Barycentric to Cartesian transformation matrix on simplex t
λ	Lagrange multiplier; eigenvalue
μ	regularization parameter
ν	multi-index
ϕ	roll angle [rad]
ψ	heading angle [rad]
$\mathbf{\Sigma}$	singular value matrix
σ	model residual standard deviation; singular value
τ_1	Kirchhoff's time constant of lag [s]
τ_2	Kirchhoff's time constant of aerodynamic hysteresis [s]

List of Figures

List of Tables

Part I

Thesis Introduction

Introduction

Colgan Air Flight 3407 crashed on the 12th of February in 2009 with 4 crew members and 45 passengers on board. During final approach, the pilot in command was being warned by the stick shaker that he was near stall entry. The pilot reacted by abruptly pulling back on the control column and increasing thrust to 75% power. Consequently, the aircraft pitched up even further. To make matters worse, the pilot overrode the stick pusher, that was meant to aid recovering stall, and the first officer retracted the flaps making recovery even more challenging. The pilots lost control of the aircraft and moments after crashed into a home near the airport. No one inside the aircraft survived. [1].

The pilots did not follow the right protocol for stall recovery. This example is not an incident; loss of Control in Flight (LOC-I) bears the most common cause of fatal accidents, and the main contributor is aerodynamic stall [2]. After the Colgan Air crash, the Airline Safety and Federal Aviation Administration Extension Act was adopted in the United States. Part of this act focuses on realizing improved stall and upset recovery training [3]. Later, in 2019, flight simulator-based stall prevention and recovery training became obligatory in the US [4] and similar legislation was passed in the European Union [5]. Simulator based training requires stall models [6]. This gives rise to an increased demand in accurate stall models for aircraft.

In the pursuit of developing a good stall model from flight-test data, a stall task force was founded at the section Control & Simulation of Delft University of Technology. The task force has developed a number of stall models in recent years for the Cessna Citation II PH-LAB. This laboratory aircraft has been employed numerous times for the purpose of collecting stall model identification data. Van Horssen et al. [7] presented a buffet model and a first longitudinal stall model. Van Ingen et al. [8] further improved this longitudinal stall model and de Fuijk et al. [9] added a lateral stall model. The models have a high accuracy in the stall regime, but are lacking in other flight conditions. These stall models are local models that have been developed for that specific purpose. This leaves the need for a global stall model, so that pilots can be exposed to a fully continuous experience from nominal flight to pre-stall, stall and post-stall. Therefore, the main focus of this thesis is to develop a global stall model.

Research Formulation

During the course of this thesis it was discovered from experiments in the SIMONA research simulator that the stall model by van Ingen et al. [8] lacked sufficient aerodynamic damping. It was then found that the method used in that research is not suitable to both identify a Kirchoff and aerodynamic pitch damping contribution. A pitch damping model contribution, complementary to a Kirchoff based one, often comes out as zero or even positive, meaning unstable damping. This finding gave rise to a new gap in literature, largely focused around the pitching moment and aerodynamic damping in stall.

For this reason, the main and first four research questions from the initial literature study were slightly altered to concur better with this new research interest. In addition, research question 5 was revised entirely to directly represent this new aim of identifying the aerodynamic damping in stall.

Now, the primary research goal of this thesis defines the main focus of the work. It is posed as follows:

Research Objective

To develop a global aerodynamic pitching moment model that can be used to describe both nominal flight and all phases of stall and captures aerodynamic damping, by using system identification techniques applied to flight test data.

in similar fashion, the main research question captures the scope of this research and sheds light on the research gap of interest:

Main Research Question

How can a full flight envelope pitching moment model including both steady and unsteady aerodynamics of a Cessna Citation II aircraft best be identified from flight test data using a global modeling method?

Alongside the main research question, five sub-questions are formulated to break down the main research question in better researchable ones.

Research Question 1

What are the advantages and disadvantages of using Multivariate Simplex B-splines for identifying and validating a full flight envelope pitching moment model from sparse flight-test data of the Cessna Citation II aircraft?

Research Question 2

What flight test maneuvers are suitable for identifying both steady and unsteady pitching moment aerodynamics?

Research Question 3

Which model structure can best be used for steady and unsteady aerodynamic modeling of the pitching moment?

Research Question 4

How does a global model compare to the local stall model by van Ingen et al.[8]?

Research Question 5

What modifications to Kirchoff's flow separation based model are necessary to enable the effective modeling of aerodynamic damping?

Part II

Scientific Article

A Nonlinear Unsteady Pitching Moment Stall Model from Cessna Citation II Flight Test Data

C.J. van Wezel*, D.M. Pool † and C.C. de Visser‡
Delft University of Technology, 2600 GB Delft, The Netherlands

One of the most widely applied identification methods for stall modelling from flight test data is based on Kirchoff's method of flow separation. This approach has not lead to a satisfactory aerodynamic pitching moment model. The introduction of the so-called X-variable, representing the point of flow separation on the wing, interferes with identification of a pitch damping term, that is needed for dynamic stability. Moreover, flow separation is only a small contributor to the pitching moment, leading to a lack of physical interpretability. In general, Kirchoff methods lead to models incompatible with the nominal flight envelope. This paper presents a nonlinear unsteady model of the pitching moment using lag states of the angle of attack measurements, identified from flight test data collected with a Cessna Citation II. The model is formulated in terms of well-known stability derivatives and is a one-on-one extension of the nominal envelope model. Model regressors are selected from a large pool of candidates using Multivariate Orthogonal Function Modeling. The candidate pool is based on a newly formulated mathematical model, such that each model contribution has a clear physical interpretation. This has lead to a C_{m_α} contribution depending on various lag states of the angle of attack, and C_{m_q} and $C_{m_{\dot{\alpha}}}$ contributions as univariate splines. The model has good predictive abilities and can report a reduction of 55.9% in validation MSE compared to a Kirchoff based pitching moment model by van Ingen et al..

I. Nomenclature

Roman

A_*	Specific force in -direction [m/s^2]	S	Wing surface [m^2]
$\overline{b}$	Wing span [-]	$\overline{S}_T$	Tail to Wing surface ratio [-]
$\overline{c}$	Mean aerodynamic chord [-]	t	Time [s]
C_*	Aerodynamic force/moment coefficient [-]	u	Step function
C_{m_*}	Aerodynamic pitching moment derivative w.r.t. *	$\overline{V}_T$	Tail to wing airspeed ratio [-]
C_{L_*}	Aerodynamic lift derivative w.r.t. *	V_∞	Free stream air speed [m/s]
F_e	Earth-fixed reference frame	V_{TAS}	True air speed [m/s]
h	Altitude [m]	V_{stall}	Stall speed
I_*	Mass moment of inertia about *-axis [kgm^2]	X	Point of flow separation variable [-]
$\overline{L}_T$	Tail moment arm [-]	$\mathbf{x}$	Physical regressor vector
m	Mass [kg]	x	Spline knot
N	Number of samples	x_e, y_e, z_e	Cartesian position in F_e [m]
n	Number of model terms	x^W	Wing lift moment arm [-]
$\mathbf{p}$	Orthogonalized regressor vector	$\mathbf{y}$	Measurement vector
p	Roll rate [rad]	$\hat{\boldsymbol{y}}$	Model output vector
q	Pitch rate [rad]	y	Spline knot
r	Yaw rate [rad]		

*M.Sc. student, Control and Simulation Division, Faculty of Aerospace Engineering, P.O. Box 5058, The Netherlands; c.j.vanwezel@tudelft.nl
†Assistant Professor, Control & Simulation Section, Faculty of Aerospace Engineering, P.O. Box 5058; d.m.pool@tudelft.nl. Associate Fellow AIAA.
‡Associate Professor, Control & Simulation Section, Faculty of Aerospace Engineering, P.O. Box 5058; c.c.devisser@tudelft.nl. Member AIAA.

Greek

α	Angle of attack [rad]	θ	Pitch angle [rad]
α_{t-k}	kth sample angle of attack lag state	σ	Standard deviation
Γ	Transformation matrix	τ_*	Time lag with index *
$\gamma_{k,j}$	Gram-Schmidt scaling parameter	$\boldsymbol{\phi}$	Orthogonal parameter vector
δ	Control surface deflection [rad]	ϕ	Bank angle [rad]
ε	Downwash at tail [rad]	ψ	Heading angle [rad]
$\boldsymbol{\theta}$	Physical parameter vector		

Subscripts

0	Constant contribution/steady state	e	Elevator
1	Nominal flight contribution/amplitude	L	Lift force
2	Stall contribution	m	Pitching moment
+	Spline	T	Thrust force
d	Dynamic		

Superscripts

T	Tail	$\circ$	Degrees
W	Wing	$\intercal$	Transpose
B	Body		

II. Introduction

Loss of control in flight (LOC-I) bears the most common cause of fatal accidents in civil and general aviation, and the main contributor is aerodynamic stall [1]. Therefore, flight simulator-based stall prevention and recovery training has become obligatory in the US [2] per 2019. Similar legislation was passed in the European Union [3]. Simulator based training requires stall models [4]. This gives rise to an increased demand in accurate stall models for aircraft.

Stall is a highly nonlinear and unsteady flight condition [5]. Conventional aircraft models, that are based on the stability derivative concept, are incapable of modeling unsteady nonlinear aerodynamics [6–8]. Therefore, Goman et al. [8, 9] and Klein et al. [10, 11] model the aerodynamic parameters as function of oscillation frequency and amplitude using forced-oscillation data from wind tunnel experiments. This approach leads to parameterized look up tables that are only applicable to forced oscillations. Moreover this experimental data lacks control inputs. A model fit for simulators should only include universal aircraft states, including control variables, and be valid for arbitrary motion. This is what models identified from flight test data can provide.

In the pursuit of developing a good stall model from flight-test data, a stall task force was founded at the section of Control & Simulation at Delft University of Technology. The task force has some unique resources to its availability, including a laboratory aircraft (Cessna Citation II PH-LAB) and six-degree-of-freedom motion research simulator SIMONA. In a previous research effort, a longitudinal model was presented by van Ingen et al. [12] based on Kirchoff's theory of flow separation. These type of models make use of the internal state variable X, that represents the point of flow separation on the wing [13]. By means of this same method, Wang et al. presented a longitudinal model with higher order X-terms to resolve issues with, amongst other things, the pitching moment model. Although it was found that such models provides a close model fit, it comes with a number of drawbacks. First of all, the X-variable cannot be directly measured but needs to be computed post-flight through a nonlinear optimization that is based on a phenomenological relationship, making the accuracy of the modelling technique questionable in the first place. Secondly, this type of model is generally incompatible with nominal flight models, that consist of the familiar control and stability derivatives. In other words, these models are local models, and cannot be generalized easily to the global flight envelope. Furthermore, obtaining a physically accurate pitch damping turns out to be problematic. The inclusion of the X-variable interferes with the identification of a pitch damping coefficient that is required for dynamic stability. Moreover, the X-variable has limited physical meaning with respect to the pitching moment because flow separation only partly explains its dynamic behavior. In reality it is a combination of flow separation, moving centre of pressure (COP) on the wing [14–16] and downwash lag at the tail [17–19]. Lastly, the X-variable runs between 0 and 1, 0 meaning full flow separation. As a result, after the point of full flow separation, this regressor is reduced to a constant and can not represent the aerodynamics that continue to develop even after flow separation.

This paper aims to provide an alternative to the Kirchoff based modelling of the pitching moment. Flight data is used from stall manoeuvres excited with quasi-random inputs, obtained from flight tests with the PH-LAB. A newly formulated mathematical pitching moment model is derived that includes all aspect of the pitching moment. Based on this mathematical model, a large set of circa 100 candidate regressors is generated. These include lag states of the angle of attack to model flow separation and the movement of the COP, and univariate splines to model nonlinear pitch damping and downwash lags. Multivariate Orthogonal Function (MOF) [20, 21] modeling is then used to identify the most relevant model contributions, while not overfitting the information content of the data. The candidates include lag states of the angle of attack to model the combined effect of flow separation and moving COP, and univariate splines in α to model nonlinear pitch damping and downwash lag.

The structure of this paper is as follows: section III provides an overview of various unsteady aspects that need to be taken into account in the pitching moment model. In section IV a new mathematical model formulation is presented. section V discusses the flight test data and section VI the methodology. Then the results are presented in section VII and finally a conclusion is found in section VIII.

III. Dynamic Stall Characteristics

Dynamic stall refers to the delay of steady stall beyond the static stall angle of attack on airfoils and wings under unsteady motion [22]. Such unsteady motions include pitching, plunging and heaving. In the subsequent only pitching motions are considered. During dynamic stall, a series of aerodynamic phenomena can be observed, that each have a pronounced effect on the aerodynamic forces and moments [16, 23]. This is discussed in subsection III.A. Furthermore, subsection III.B talks about two different stall types. Lastly, subsection III.C explains how the interaction between wing and tail contributes to dynamic stall.

A. Stages of Dynamic Stall

Figure 1 shows an illustration of what an unsteady lift curve can look like for an airfoil undergoing a pitching motion with a constant angle of attack rate. This illustration is of the hand of Choudhry et al., and they identify 6 stages of dynamic stall:

1) On the onset of rotation, the lift curve closely follows the steady-state curve, although a slight offset can be observed that indicates the first effects of dynamic stall.
2) A regime follows that is characterized by a near linear lift slope past the point of steady stall. The reach of this regime is related to the angle of attack rate. This phenomenon can be explained by the formation of a vortex at the trailing edge, providing additional suction pressure, keeping the flow attached. Another explanation is the so called *leading edge jet effect*.
3) Next, a flat portion appears, referred to as the plateau. At this stage, separation has initiated at the trailing edge and moves towards the leading edge. Meanwhile, a separation bubble forms at the leading edge.
4) The plateau is an indication for the formation of a large vortex at the leading edge (LE), the dynamic stall vortex (DSV). Immediately following the plateau, a sharp rise in lift is observed. This increase is the direct result of the formation and shedding of the DSV. The DSV rolls over the wing, increasing lift momentarily. When the DSV 'spills', it moves aft, resulting in a displacement of the center of pressure (COP) in the same direction.
5) The DSV has convected to mid-chord, where the maximum lift is achieved. At this point the DSV leaves the airfoil surface.
6) Now the post-stall domain is entered, where the wing experiences a dramatic drop in lift.

Especially the formation and the subsequent shedding of the leading edge vortex has dramatic effects on the lift, drag and pitching moment. In its final act, the DSV causes a sharp peak in the lift force. During this event, the COP of the wing moves aft. The combination of the increased lift force and moment arm leads to a strong increase in the negative pitching moment. This phenomenon is referred to as the *moment stall* [14, 16, 23]. The instant where the DSV has separated and the lift finally decreases is referred to as the *lift stall*. The moment stall precedes the lift stall and the delay between the two can be related to the velocity at which the DSV convects over the wing [24]. Although not shown in Figure 1, additional secondary and even tertiary vortices can appear on the wing, leading to further lift, drag and pitching moment peaks [14].

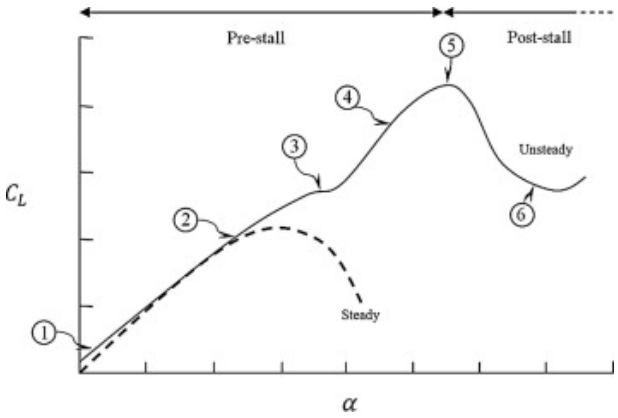


Fig. 1 Comparison of unsteady and steady-state primary lift-curve undergoing unsteady ramp-type dynamic stall. This illustration was taken from Choudhry et al. [23].

B. Stall Degree

Two types of stall can be distinguished in pitching airfoils, being light and deep stall. In light stall, the maximum angle of attack and lift stall more or less coincide. In deep stall, the maximum angle of attack is increased beyond the point of lift stall. Also, in light stall the DSV is not fully developed, while it is in deep stall [25]. As a result the load peaks in deep stall reach significantly larger values. The type of stall can be clearly recognized from the moment hysteresis curves. Figure 2a and 2b illustrate the moment hysteresis for light and deep stall respectively.

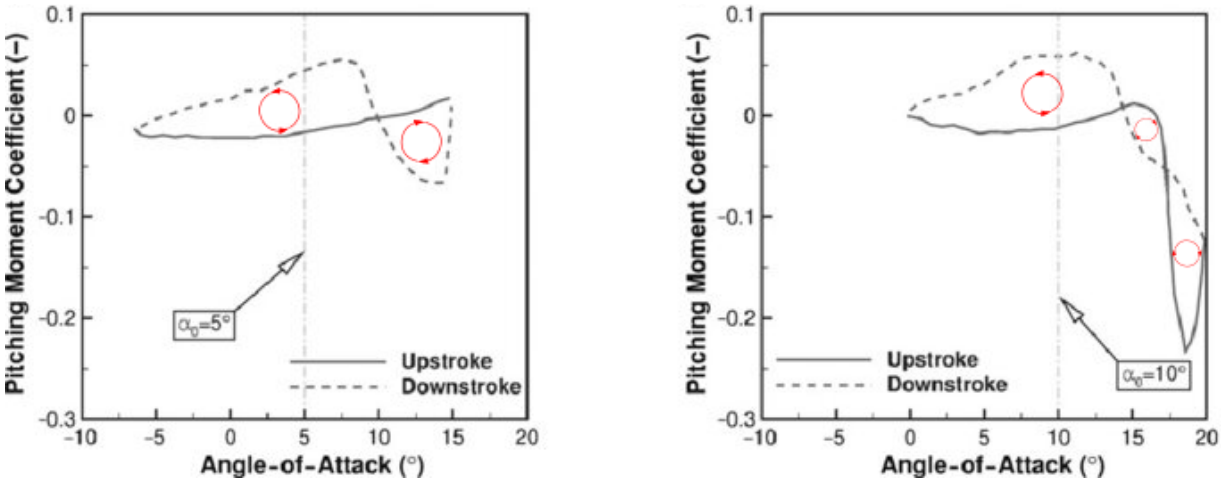


Fig. 2 Typical pitch moment coefficient hysteresis curves for (a) light stall and (b) deep stall for a NACA0012 airfoil in pitch oscillation, each with a different mean angle of attack (α_0), but the same amplitude and frequency. These illustrations were taken from McCroskey et al. [25].

As it turns out, the aerodynamic damping in stall can become unstable, meaning the pitching motion amplifies itself, instead of dampening out [26, 27]. This stall induced unstable damping is found in clockwise moment hysteresis loops, i.e. where the pitching moment is higher on the upstroke than on the downstroke [16, 24]. According to Corke et al. the amount of damping, whether positive or negative, is related to the size of the moment hysteresis loops. McCroskey et al. noted that light stall is more aerodynamically unstable than deep stall [25]. And indeed, Figure 2a and b reveal that the light stalls have a larger clockwise loop.

Since this literature only considers pitching airfoils without a tail, the negative moment peak corresponds exactly to the point of lift stall. After that point the moment recovers due to the stalled wing lift. The magnitude of the negative

moment peak is related to the stall degree: the deeper the stall the larger the moment excursion. That means that the negative moment peak can be related to the angle of attack rate or reduced frequency of the motion.

C. Wing Tail Interaction

The literature discussed in subsection III.A and III.B exclusively looked at airfoils in unsteady motion. An aircraft with a wing-tail configuration has an additional source of unsteadiness. This stems from the interaction between wing and tail [17, 18, 28]. The wing causes the free stream flow to deflect downwards, which results in a downwash ε at the tail. The downwash is a function of the wing angle of attack and assumed to be linear in attached flow. An instantaneous change in wing angle of attack will not be instantly reflected in the tail downwash. Instead, the modified flow resulting from this change in angle of attack reaches the tail with a time lag. This lag approximately amounts to the time needed for the flow to travel from the trailing edge of the wing to the tail surface. The downwash lag can be proportionally related to the angle of attack rate in a first order approximation, effectively resulting in a quasi-steady effect rather than an unsteady effect.

IV. Mathematical Modeling of Pitching Moment

Since the early days of aviation, the mathematical formulation of aircraft models is given by a Taylor series with only first order linear terms retained [28]. This is the familiar control and stability derivative formulation. However, the first order model is only valid in the nominal flight envelope where nonlinear and unsteady effects are negligible. In a nonlinear unsteady regime, such a model is inadequate. In the subsequent, the classical linear steady mathematical formulation and a nonlinear unsteady mathematical model by Khrabov et al. [15] for the pitching moment are discussed in subsection IV.A and subsection IV.B, respectively. Then, a newly modified model formulation for the pitching moment is presented in subsection IV.C that encompasses both linear steady and nonlinear unsteady aerodynamics.

A. Mathematical Modeling of Linear Steady Pitching Moment

The classical first order Taylor expansion in terms of the aircraft states and control variables is given by Equation 1.

$$C_m = C_{m_0} + C_{m_\alpha}\alpha + C_{m_q} q + C_{m_{\dot\alpha}}\dot\alpha + C_{m_{\delta_e}}\delta_e \tag{1}$$

Where only the angle of attack α, pitch rate q, $\dot\alpha$ and elevator deflection δ_e are considered to effect the pitching moment. The variation with altitude and velocity are excluded assuming a single flight condition. The stability derivatives C_{m_α}, C_{m_q} and $C_{m_{\dot\alpha}}$ can be broken down into equations Equation 2-4. A derivation is provided by for example Etkin [28].

$$C_{m_\alpha} = -C_{L_\alpha} x^W - C^T_{L_\alpha}(1 - \frac{\partial\varepsilon}{\partial\alpha})\overline{V}_T^2\overline{S}_T\overline{L}_T \tag{2}$$

$$C_{m_q} = \frac{\partial C^T_L}{\partial\alpha}\overline{V}_T^2\overline{S}_T\overline{L}_T^2 \tag{3}$$

$$C_{m_{\dot\alpha}} = \frac{\partial C^T_L}{\partial\alpha}\frac{\partial\varepsilon}{\partial\alpha}\overline{V}_T^2\overline{S}_T\overline{L}_T^2 \tag{4}$$

In Equation 2-4 $\overline{V}_T$ is the ratio of the airspeed at the tail and wing and is assumed to be constant. $\overline{S}_T$ is the tail to wing surface ratio, and $\overline{L}_T$ the tail arm normalized by the mean aerodynamic chord (MAC) $\overline{c}$. The superscripts W and T indicate the tail and wing. C_{L_α} is the lift gradient and $\frac{\partial\varepsilon}{\partial\alpha}$ the downwash gradient. C_{m_α} represents the moment generated by the wing and tail. C_{m_q} is the additional pitching moment in a pure pitch motion generated at the tail due to curvature of the air around the center of gravity (CG). Here the contribution by the wing is neglected. Furthermore, $C_{m_{\dot\alpha}}$ is the downwash lag coefficient and it represents the quasi-steady effect described in subsection III.C.

Due to the tight kinematic coupling between q and $\dot\alpha$, the pitch damping and downwash lag coefficient are oftentimes combined into one parameter. Since $q \simeq \dot\alpha$ the following can be derived:

$$C^*_{m_q} = C_{m_q} + C_{m_{\dot\alpha}} = \frac{\partial C^T_L}{\partial\alpha}(1 + \frac{\partial\varepsilon}{\partial\alpha})\overline{V}_T^2\overline{S}_T\overline{L}_T^2 \tag{5}$$

resulting in the simplified model in Equation 6.

$$C_m = C_{m_0} + C_{m_\alpha}\alpha + C^*_{m_q} q + C_{m_{\delta_e}}\delta_e \tag{6}$$

B. TsAGI Model

Khrabov et al. proposed a mathematical model formulation that can describe the nonlinear unsteady aerodynamics of the pitching moment of a conventional wing-tail-body configured aircraft, commonly referred to as the *TsAGI model* [15]:

$$\begin{aligned} &C_m = C^B_m(\alpha) + C^W_{m_0} + C^W_L x^W - \overline{L}_T \overline{S}_T C^T_L(\alpha + \alpha_d - \varepsilon) + C_{m_q}(\alpha) q \\ &\text{where:} \\ &\alpha_d = -\frac{\overline{L}_T \overline{c} q}{V_\infty} \end{aligned} \tag{7}$$

Where C^W_L and C^T_L are the wing and tail lift respectively, as indicated by the superscripts W and T. C^B_m is a moment generated by the fuselage, x^W is the distance from the COP of the wing to the CG of the aircraft. Furthermore, α_d is an additional contribution to the angle of attack as seen by the tail caused by the aircraft rotating around its CG. $C_{m_q}(\alpha)$ is an additional nonlinear pitch damping derivative. C^W_L is partitioned into two components:

$$C^W_L = C_{L1} + C_{L2} \tag{8}$$

Here C_{L1} represents the linear part of the wings' lift, associated with attached flow, and C_{L2} the deviation from linear lift, caused by flow separation. These equations do not describe unsteady aerodynamics, as all the terms are functions of the instantaneous angle of attack. However, according to Khrabov et al, all of C_{L2}, x^W and ε are unsteady variables. To make Equation 7 valid in both steady and unsteady conditions, Khrabov et al. introduce three ODEs, given by Equation 9-11.

$$\tau_1(\alpha)\frac{\partial C_{L2}}{\partial t} + C_{L2} = C_{L20}(\alpha) \tag{9}$$

$$\tau_2(\alpha)\frac{\partial x^W}{\partial t} + x^W = x^W_0(\alpha) \tag{10}$$

$$\tau_3(\alpha)\frac{\partial \varepsilon}{\partial t} + \varepsilon = \varepsilon_0(\alpha) \tag{11}$$

$\tau_1(\alpha)$, $\tau_2(\alpha)$ and $\tau_3(\alpha)$ are three time lags that depend on the angle of attack, and $C_{L20}(\alpha)$, $x^W_0(\alpha)$ and $\varepsilon_0(\alpha)$ are three nonlinear functions that describe the steady behaviour of the three unsteady variables. Note that

Equation 7-11 requires the identification of 7 nonlinear functions, being $C_{m_q}(\alpha)$, $\tau_1(\alpha)$, $\tau_2(\alpha)$, $\tau_3(\alpha)$, $C_{L20}(\alpha)$, $x^W_0(\alpha)$, $\varepsilon_0(\alpha)$. Khrabov et al. were able to do so by means of wind tunnel tests using experiments with different configuration of the test aircraft (e.g. only body and wing or only body and tail). However, in flight testing it is not possible to obtain data sets for different configurations. Consequently, all the above described functions would need to be estimated from a single data set where effects due to wing, tail, body and their interaction are all intertwined. It is not realistic to expect to do so.

C. Modified Mathematical Pitching Moment Model

In order to mold Equation 7 in a form more suitable for identification from flight test data, it is first of all desirable to eliminate the ODEs in Equation 9-11. These ODEs effectively act as a time lag filter, such that the unsteady variables are not only a function of the instantaneous α, but also of that at previous time instances. This is how the pre-history of the angle of attack enters the model equations. Analogously, C_{L2} could also be directly modeled as a nonlinear function of the time trace of α, such as in Equation 12.

$$\begin{aligned} C^W_L &= C^W_{L1}(\alpha) + C^W_{L2}(\alpha(t)) \\ &= C^W_{L0} + C^W_{L1\alpha}\alpha + C^W_{L2}(\alpha(t)) \end{aligned} \tag{12}$$

Here the attached flow component C_{L1} has been rewritten in terms of the parameters from the nominal model. Similarly, this trick can be applied to x^W, C_L^T and ε. This gives Equation 13-15.

$$\begin{aligned} x^W &= x_1^W + x_2^W(\alpha(t)) \\ &= x_0^W + \frac{\partial x_1}{\partial \alpha}\alpha + x_2^W(\alpha(t)) \\ &\approx x_0^W + x_2^W(\alpha(t)) \end{aligned} \tag{13}$$

$$\begin{aligned} C_L^T &= C_{L1}^T(\alpha) + C_{L2}^T(\alpha) \\ &= C_{L0}^T + C_{L1_\alpha}^T \alpha_T + C_{L2}^T(\alpha) \end{aligned} \tag{14}$$

where

$$\alpha_T = \alpha - \varepsilon$$

$$\begin{aligned} \varepsilon &= \varepsilon_1(\alpha) + \varepsilon_2(\alpha) \\ &= \varepsilon_0 + \frac{\partial \varepsilon_1}{\partial \alpha}\alpha + \varepsilon_2(\alpha) \end{aligned} \tag{15}$$

The subscripts 0 and 1 in Equation 12-15 always correspond to the linear steady aerodynamic contribution, such as in nominal conditions. The subscript 2 refers to the deviation of these conditions, due to nonlinear and possibly unsteady aerodynamics. Note that $C_{L2}^T(\alpha)$ and $\varepsilon_2(\alpha)$ in Equation 14 and 15 are given as nonlinear steady functions, as evident by the omission of (t). The unsteadiness introduced by the downwash lag will be introduced later by including a $C_{m_{\dot{\alpha}}}$ parameter in addition to C_{m_q}.

Since $C_{L2}^T(\alpha)$ and $\varepsilon_2(\alpha)$ are steady functions, they must carry a value of zero up to some critical α. In the absence of a good model for the nonlinear tail lift and downwash contributions, they will be represented by a linear combination of spline basis functions in the subsequent:

$$C_{L2}^T(\alpha) = \sum_{i=1}^{n} c_i(\alpha - x_i)_+^i$$

where: (16)

$$(\alpha - x_i)_+^i = \begin{cases} (\alpha - x_i)^i & \text{if } \alpha \geq x_i \\ 0 & \text{if} \alpha < x_i \end{cases}$$

$$\varepsilon_2(\alpha) = \sum_{i=1}^{m} d_i(\alpha - y_i)_+^i$$

where: (17)

$$(\alpha - y_i)_+^i = \begin{cases} (\alpha - y_i)^i & \text{if } \alpha \geq y_i \\ 0 & \text{if } \alpha < y_i \end{cases}$$

c_i and d_i are to be determined coefficients, x_i and y_i are knots, and n and m are the highest order spline components included. Note that the lowest level contribution would be a first order term, i.e. $c_1(\alpha - x_1)_+$, representing a change of slope at $\alpha > x_1$.

Now we can write the contribution to the pitching moment in a pure α-motion using the expressions in Equation 12, 13 and 14:

$$\begin{aligned} C_m &= -C_L^W x^W - C_L^T \overline{V}_T^2 \overline{S}_T \overline{L}_T \\ &= -\left(C_{L1}^W(\alpha) + C_{L2}^W(\alpha(t))\right)\left(x_1^W + x_2^W(\alpha(t)\right) - \left(C_{L1}^T(\alpha) + C_{L2}^T(\alpha)\right)\overline{V}_T^2 \overline{S}_T \overline{L}_T \\ &= -C_{L1}^W(\alpha)x_1^W - C_{L1}^W(\alpha)x_2^W(\alpha(t)) - C_{L2}^W(\alpha(t))x_1^W - C_{L2}^W(\alpha(t))x_2^W(\alpha(t)) \\ &\quad - C_{L1}^T(\alpha)\overline{V}_T^2 \overline{S}_T \overline{L}_T - C_{L2}^T(\alpha)\overline{V}_T^2 \overline{S}_T \overline{L}_T \end{aligned} \tag{18}$$

Then the change in pitching moment with a change in the instantaneous α is given as:

$$\begin{aligned}
C_{m_\alpha} = &-\frac{\partial C^W_{L1}(\alpha)}{\partial \alpha} x^W_1 - \frac{\partial C^W_{L1}(\alpha)}{\partial \alpha} x^W_2(\alpha(t)) - C^W_{L1}(\alpha)\frac{\partial x^W_2(\alpha(t))}{\partial \alpha} \\
&- \frac{\partial C^W_{L2}(\alpha(t))}{\partial \alpha} x^W_1 - \frac{\partial C^W_{L2}(\alpha(t))}{\partial \alpha} x^W_2(\alpha(t)) - C^W_{L2}(\alpha(t))\frac{\partial x^W_2(\alpha(t))}{\partial \alpha} \\
&- \frac{\partial C^T_{L1}(\alpha)}{\partial \alpha}\overline{V}_T^2 \overline{S}_T \overline{L}_T - \frac{\partial C^T_{L2}(\alpha)}{\partial \alpha}\overline{V}_T^2 \overline{S}_T \overline{L}_T \\
= &-C^W_{L1_\alpha} x^W_0 - C^T_{L1_\alpha}\left(1 - \frac{\partial \varepsilon}{\partial \alpha}\right)\overline{V}_T^2 \overline{S}_T \overline{L}_T \\
&- \frac{\partial C^W_{L1}(\alpha)}{\partial \alpha} x^W_2(\alpha(t)) - C^W_{L1}(\alpha)\frac{\partial x^W_2(\alpha(t))}{\partial \alpha} - \frac{\partial C^W_{L2}(\alpha(t))}{\partial \alpha} x^W_1 \\
&- \frac{\partial C^W_{L2}(\alpha(t))}{\partial \alpha} x^W_2(\alpha(t)) - C^W_{L2}(\alpha(t))\frac{\partial x^W_2(\alpha(t))}{\partial \alpha} - \frac{\partial C^T_{L2}(\alpha)}{\partial \alpha}\overline{V}_T^2 \overline{S}_T \overline{L}_T \\
= &\, C_{m1_\alpha} + C_{m2_\alpha}(\alpha(t))
\end{aligned} \tag{19}$$

Now C_{m1_α} is identical to the α-derivative from the linear steady model in Equation 2, while C_{m2_α} is a nonlinear unsteady function that models the nonlinear unsteady deviation of the linear lift caused by flow separation and potential nonlinearity at the tail.

Similarly, alternative expressions for C_{m_q} and $C_{m_{\dot\alpha}}$ can be found. Taking the derivative of Equation 16 w.r.t. α and plugging into Equation 3 gives Equation 20.

$$\begin{aligned}
C_{m_q} &= \left(C^T_{L1_\alpha} + c_1 u(\alpha - x_1) + 2c_2(\alpha - x_2)_+ + 3c_3(\alpha - x_3)^2_+ + \ldots\right)\overline{V}_T^2 \overline{S}_T \overline{L}_T^2 \\
&= C_{m1_q} + C_{m2_q}(\alpha)
\end{aligned} \tag{20}$$

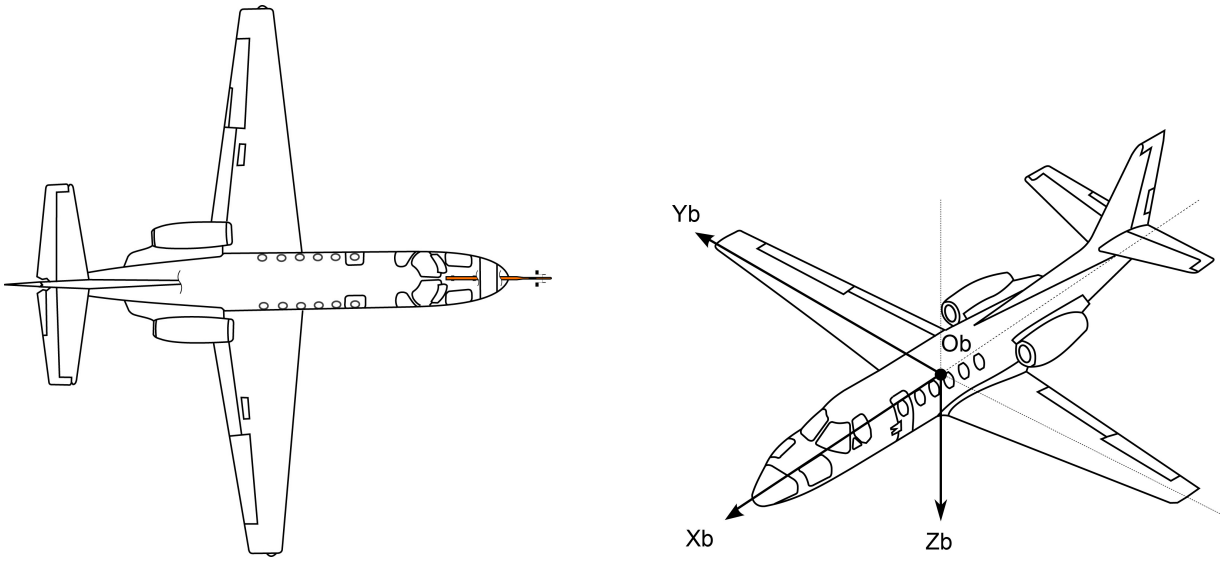
The Macaulay step function $u(\cdot)$ is the derivative with respect to α of the first order spline term in Equation 16. Again two derivatives result. C_{m1_q} is the nominal pitch damping and C_{m2_q} a nonlinear correction term. This term is nonlinear steady, similar to the parameter in Equation 7. Continuing, taking the derivative of Equation 16 and 17 w.r.t. α and plugging into Equation 4 gives us one last contribution.

$$\begin{aligned}
C_{m_{\dot\alpha}} = &\left(C^T_{L1_\alpha} + c_1 u(\alpha_t - x_1) + 2c_2(\alpha_t - x_2)_+ + 3c_3(\alpha - x_3)^2_+ + \ldots\right)\cdot \\
&\left(\frac{\partial \varepsilon}{\partial \alpha} + d_1 u(\alpha - y_1) + 2d_2(\alpha - y_2)_+ + 3d_3(\alpha - y_3)^2_+ + \ldots\right)\overline{V}_T^2 \overline{S}_T \overline{L}_T^2 \\
= &\, C_{m1_{\dot\alpha}} + C_{m2_{\dot\alpha}}(\alpha)
\end{aligned} \tag{21}$$

Also here the familiar nominal contribution appears, joint by a nonlinear correction term. At this point, all the familiar aerodynamic parameters have been rewritten as the sum of the linear steady parameters and a corresponding correction term facilitating nonlinear and unsteady aerodynamic modeling. Combining the above gives Equation 22.

$$\begin{aligned}
C_m = &\, C_{m_0} + C_{m1_\alpha}\alpha + C^*_{m1_q} q + C_{m1_{\delta_e}}\delta_e + \\
&+ C_{m2_\alpha}(\alpha(t))\alpha + C_{m2_q}(\alpha) q + C_{m2_{\dot\alpha}}(\alpha)\dot\alpha
\end{aligned} \tag{22}$$

Equation 22 is a new formulation of the pitching moment and a main contribution by this paper. For simplicity it is assumed that the elevator effectiveness $C_{m1_{\delta_e}}$ remains unaffected so that a correction $C_{m2_{\delta_e}}$ is unnecessary. If expressions can be found for C_{m2_α}, C_{m2_q} and $C_{m2_{\dot\alpha}}$ such that they are zero in the linear steady regime, then this model would have global modeling capacities.



(a) (b)

Fig. 3 Schematic top view (a) and diagonal view (b) of the Cessna Citation II PH-LAB operated by Delft university of Technology and NLR

Table 1 Cessna Citation II PH-LAB Dimensions and mass properties (BEW).

	Parameter	Value
Dimensions	S	30 m^2
	$\overline{b}$	15.9 m
	$\overline{c}$	2.09 m
Mass & Inertia	m	4,157 kg
	I_{xx}	12,392 kg $\cdot$ m^2
	I_{yy}	31,501 kg $\cdot$ m^2
	I_{zz}	41,908 kg $\cdot$ m^2
	I_{xz}	2,252.2 kg $\cdot$ m^2

Table 2 Flight test equipment onboard the Cessna Citation II PH-LAB, and the variables measured relevant for aerodynamic system identification.

Name	Explanation	Measures	Variables	Units
GPS	Global positioning system	Position in F_E	x_e, y_e, z_e	m
		Velocity in F_E	$\dot{x}_e$, $\dot{y}_e$, $\dot{z}_e$	m/s
DADC	Digital air data computer	Total airspeed	V_{TAS}	m/s
AHRS	Attitude and heading reference system	Aircraft attitude	ϕ, θ, ψ	rad
		Body rotation rates	p, q, r	rad/s
		Body specific forces	A_x, A_y, A_z	m/s^2
Synchro	Angle measurements	Control surface deflection	δ_a, δ_e, δ_r	rad
Boom	Air data boom	Air incidence angle	α, β	rad

V. Flight Test Data

The results in this paper are obtained from flight test data only. subsection V.A describes the test aircraft that was used to collect the data, subsection V.B details the specific maneuvers flown for the research, and finally in subsection V.C the data processing procedures and final data set is discussed.

A. Test Aircraft

Flight data was gathered with the Cessna Citation II PH-LAB. This former business jet is operated by the Delft University of Technology and the Netherlands Aerospace Centre (NLR). It has been fully repurposed to serve as a laboratory aircraft. Schematic drawings and an overview of its dimensions and mass properties are shown in Figure 3a-b and Table 1, respectively. It is equipped with an advanced Flight-Test Instrumentation System (FTIS), that logs the information measured by sensors in flight. These sensors include, amongst others, an air data boom on the nose of the aircraft, that can accurately measure the angle of attack. The available sensors and their measured variables relevant for aircraft system identification are listed in Table 2.

B. Experiment Design

Quasi-random input maneuvers were flown in order to excite the aircraft over a broad frequency range during stall. This flight test technique is based on a piloting technique described by Morelli et al. [29]. During the maneuver, stall

is initially entered by reducing airspeed by 1 kts/s untill the stall velocity is reached. At this point, the pilots apply disturbances to the control columns, doing this as randomly as possible. Additionally, they apply a bias in order to keep the aircraft in the desired flight condition. In all stalls the elevator was used by one of the two pilots. In most stalls, the other pilot would control the ailerons, while in a few cases this would be the rudder.

Part of the data set was taken from the one used by van Ingen et al., flown in 2018 [12]. These were partly wings-level symmetric stalls, and partly accelerated stalls. In the former, stall is entered in trimmed and level flight. In the latter, stall is entered at a bank angle of 45 degrees, at 1.4g. The main purpose of the accelerated stalls is do reduce the correlation between pitch rate q and angle of attack rate $\dot{\alpha}$. According to Jategaonkar et al., sufficient decorrelation for system identification is achieved at bank angles over 40 degrees [18]. On the other hand, wings level maneuvers can reach higher angles of attack, and are thus characterized by a deeper stall.

All maneuvers were flown in clean condition, i.e. no flaps and landing gear retracted. Also, the stalls were all flown in a small altitude range around FL170-180. This to not further complicate the data with altitude effects. It is not unlikely that altitude will effect the stall characteristics, as there is a relationship with Mach number [23], and the stall speed is a function of altitude.

C. Flight Data and Pre-Processing

7 accelerated stalls (1.4g) and 11 wings-level stall used by van Ingen et al. [12] were supplemented with 8 additional 1.4g stalls newly flown in early 2024. This has led to a data set containing 26 stalls, with 15 accelerated stalls and 11 wings-level stalls. All recorded data is pre-processed by means of a zero-phase low-pass Butterworth filter of order 4 and a cut-off frequency of 4 Hz. This ensures that vibrations originating from the stall buffet and noise amplification due to numerical differentiation are removed as much as possible. Modeling of the stall buffet can be done separately as van Horssen et al. showed [30], and is ignored in this paper. Numerical differentiation is required to obtain signals for the angle of attack rate. Furthermore, flight path reconstruction is applied to obtain an improved state estimate. An Unscented Kalman Filter (UKF) was used following up on van Horssen et al. and van Ingen et al. who successfully applied it in their stall modeling efforts [12, 30].

The portion of the flight envelope covered by the maneuvers is visualized in Figure 4 and 5. The data was split in identification and validation data, based on an 80-20 split. The sets 2, 6, 14, 19 and 24 were randomly selected as validation data. Finally, Figure 8 show the α time traces for an accelerated (stall set 1) and wings-level (stall set 21) stall. There is a clear difference in angle of attack attained. Figure 9 shows the $\dot{\alpha}$ and q time traces in the same sets. In the wings level stall they largely overlap, while in the accelerated stalls there is a clear distinction.

VI. Methodology

A. System Identification Framework

According to Equation 22 the pitching moment can be split into two parts: one as predicted by the classical linear stability derivative model representing the nominal flight envelope with attached flow, and one that describes the deviation from that flow regime due to stall. The nominal model C_{m1} for the Cessna Citation PH-LAB has been estimated and validated previously by van den Hoek et al. [31]. The parameter estimates can be found in Table 3.

Table 3 Nominal flight envelope parameter estimates by van den Hoek et al. [31].

Parameter	C_{m_0}	C_{m1_α}	C_{m1_q}	$C_{m1_{\delta_e}}$
Value	0.021	-0.488	-11.9	-1.25

Since this contribution to the pitching moment is already known, there is no need to estimate it in the stall model identification. Instead, the nominal contribution can be subtracted from the C_m measurements to then identify the stall terms. The nominal model residual is given by Equation 23.

$$
\begin{aligned}
&C_m - C_{m1} = C_{m0} + C_{m_\alpha}(\alpha(t))\alpha + C_{m_q}(\alpha)q + C_{m_{\dot{\alpha}}}(\alpha)\dot{\alpha} \\
&\text{where} \\
&C_{m1} = C_{m1_\alpha}\alpha + C^*_{m_q} q + C_{m_{\delta_e}}\delta_e
\end{aligned}
\tag{23}
$$

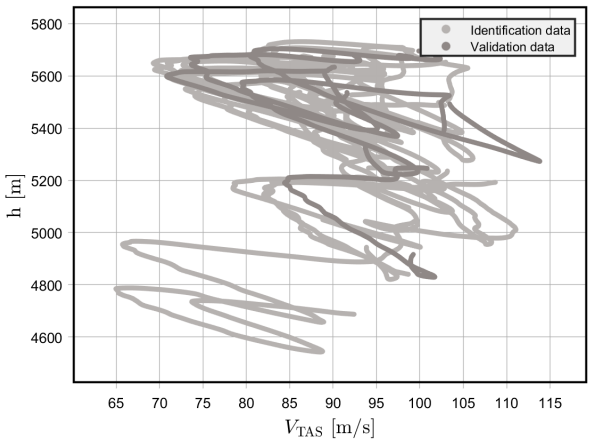


Fig. 4 V_{TAS}-h flight envelope covered by the 26 stall maneuvers.

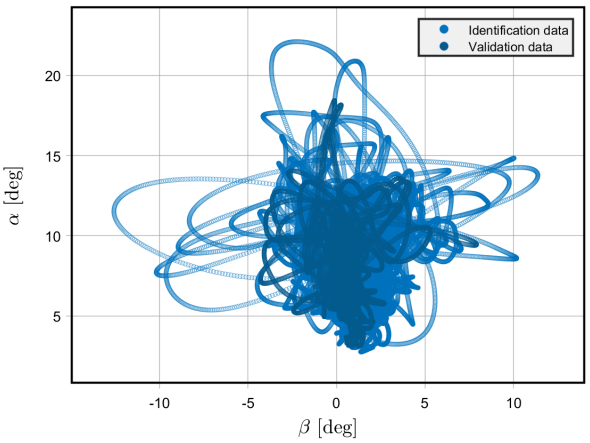


Fig. 5 α-β flight envelope covered by the 26 stall maneuvers.

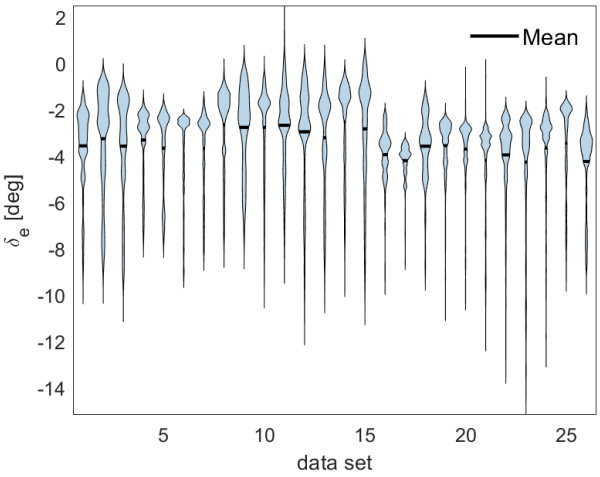


Fig. 6 Violin plots of elevator inputs during the 26 stall maneuvers in the data set.

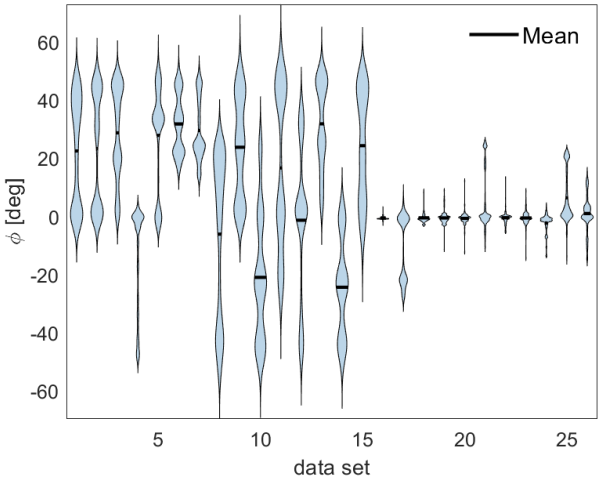


Fig. 7 Violin plots of bank angles during the 26 stall maneuvers in the data set.

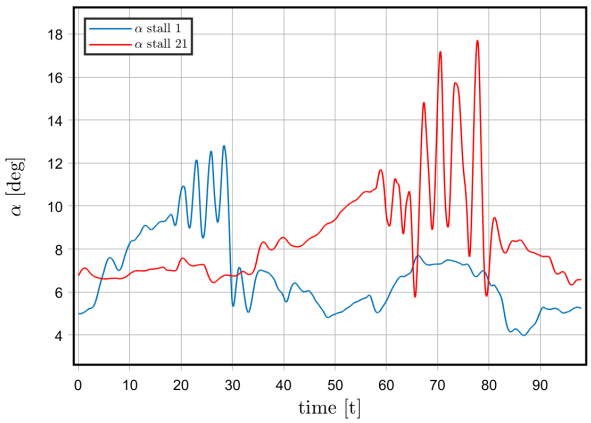


Fig. 8 Time traces of α for an accelerated stall (blue line) and a wings level stall (red line).

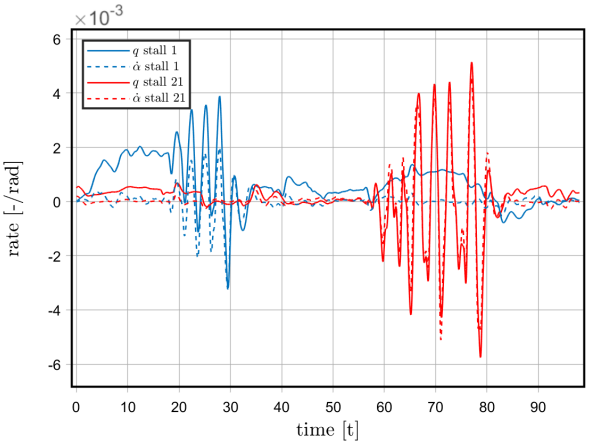


Fig. 9 Time traces of q and $\dot{\alpha}$ for an accelerated stall (blue lines) and a wings level stall (red lines).

Note that the bias term C_{m_0} from Table 3 will not be used, as the nominal model bias can simply be recomputed together with the stall bias.

The next step of the identification framework is to select a set of regressors that can best describe the stall dynamics. Multivariate Orthogonal Function (MOF) modeling will be used to select the stall regressors from a pool of candidates. This pool of candidates will be generated based on Equations 19, 20 and 21. With the model structure known, C_{m2} will be identified using linear regression. Finally, the total pitching moment can then be computed as the sum of the nominal model and the stall model. A flowchart of this system identification framework is provided in Figure 10.

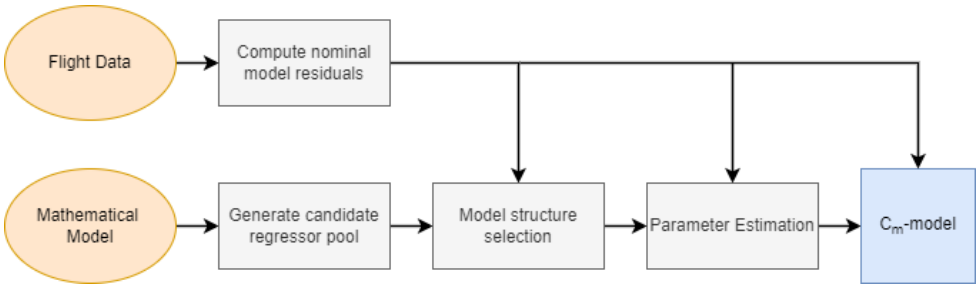


Fig. 10 Flow chart of system identification framework.

B. Candidate Regressor Pool

The mathematical model as formulated in Equation 22 provides some guidance on what terms to include in the model structure. Below the candidate regressors considered for the model are discussed. An overview of all the candidates is given in Table 4

1. α Regressors

$C_{m2_\alpha}(\alpha(t))$ represents a deviation of the pitching moment from from linear steady aerodynamics as predicted by C_{m1_α}. The deviation from linear is a direct result of a simultaneous loss in wing lift and moving COP, and potentially reduced tail lift. The parameter is a function of the time history of α because flow separation and movement of the COP are unsteady. $\alpha(t)$ can be approximated with a Taylor expansion including time derivatives of α up to a certain degree. Alternatively, Morelli shows that lag states of α can also be used as regressors [32]. In the subsequent α_{t-k} denotes such a lag state, where k is the lag. With a sampling frequency of 100 Hz, α_{t-k} represents the α measurement $\frac{k}{100}$ s prior to time t. To increase model complexity, higher order and cross terms of the lag states can be included in the model. It was decided to include a time lag up to 0.3 s, represented by α_{t-30}. A wide range of lags can be included because the MOF algorithm will be used to select a small subset of the candidates. The only drawback of many candidates is some additional computation time.

C_{m2_α} was derived as the derivative of C_{m2} with respect to α_t, therefore the total moment contribution for the α regressors is obtained by multiplying with α_t. Regressors up to the third order are included in the candidates.

2. q Regressors

$C_{m2_q}(\alpha)$ represents changes in the pitch damping due to a potentially nonlinear tail lift at high α. The contribution to C_{m2_q} by the wing is neglected. This can be justified by the very small moment arm of the wing lift compared to that of the tail. Equation 20 provides inspiration for good candidate regressors. It includes spline basic functions with zero order continuity. Because the point where the splines become active is unknown, if this point lies within the stall envelope at all, a large range of spline knots is included in the candidates. It was decided to choose a range between 0 and 20 deg. Examples of these spline terms are provided in Table 4. Again, a large range of candidates is included to cover any potential strong model contributions.

3. $\dot{\alpha}$ Regressors

$C_{m2_{\dot{\alpha}}}(\alpha)$ represents changes in the downwash lag damping due to a combined nonlinear tail lift and downwash gradient at high α. Similar to C_{m2_q}, spline terms are used to model it. In this case however, the regressors can be a

multiplication of two splines with different knots, as suggested by Equation 21.

4. Other Regressors

Finally, three more candidates were included in the candidate pool. First of all there is a bias. Secondly, the elevator deflection is included to compensate for any potential model errors in the model by van den Hoek et al.. Lastly, the thrust coefficient C_T is part of the pool. This model term was found to be beneficial for model quality by van Ingen et al. [12]. Since the current data set is largely the same, it is likely that C_T will again positively contribute. Physically it is not logical to include this term. Appendix C from the aforementioned paper by van Ingen et al. is referred to for an explanation on the inclusion of C_T.

Table 4 Candidate Regressors Pitching Moment Model

Order	Regressor	Order	Regressor
$\underline{C_m - \alpha}$		$\underline{C_m - \dot{\alpha}}$	
1	α_t	0	$u(\alpha - 0°)\dot{\alpha}, u(\alpha - 1°)\dot{\alpha}, \dots, u(\alpha - 20°)\dot{\alpha}$
2	$\alpha_t^2, \alpha_{t-1}\alpha_t, \alpha_{t-2}\alpha_t, \dots, \alpha_{t-30}\alpha_t$	1	$u(\alpha - 0°)(\alpha - 0°)_+\dot{\alpha}, u(\alpha - 0°)(\alpha - 1°)_+\dot{\alpha}, \dots,$ $u(\alpha - 11°)(\alpha - 14°)_+\dot{\alpha}, \dots, u(\alpha - 20°)(\alpha - 20°)_+\dot{\alpha}$
3	$\alpha_t^3, \alpha_{t-1}\alpha_t^2, \dots, \alpha_{t-18}\alpha_{t-24}\alpha_t, \dots, \alpha_{t-30}^2\alpha_t$		
$\underline{C_m - q}$		2	$(\alpha - 0°)_+^2\dot{\alpha}, (\alpha - 0°)_+(\alpha - 1°)_+\dot{\alpha}, \dots,$ $(\alpha - 9°)_+(\alpha - 17°)_+\dot{\alpha}, \dots, (\alpha - 20°)_+^2\dot{\alpha}$
0	$u(\alpha - 0°)q, u(\alpha - 1°)q, \dots, u(\alpha - 20°)q$		
1	$(\alpha - 0°)_+q, (\alpha - 1°)_+q, \dots, (\alpha - 20°)_+q$	*other*	
2	$(\alpha - 0°)_+^2q, (\alpha - 1°)_+^2q, \dots, (\alpha - 20°)_+^2q$		$1, \delta_e, C_T$

C. Model Structure and Parameter Estimation

Table 4 provides a finite set of candidate regressors for the stall model, however the question remains what subset yields the optimal model. Many of the candidates do not represent the real underlying physics that generated the data, and should therefore not be part of the model structure. Furthermore, a good model will include enough model terms to capture the complex stall dynamics but not overfit the data. However, the chosen model complexity must be justified by the information content in the data. If that is not the case, the model will overfit the identification data and not generalize well to unseen data. Another challenge is preventing highly correlated regressors. Including such variables in the model structure will also inevitably lead to overfitting and large parameter variance. These considerations are especially important in this problem, due to the size and the potential correlations between the entries of the candidate regressor pool.

1. Multivariate Orthogonal Function Modeling

Morelli developed a combined model structure selection and parameter estimation techniques for nonlinear aerodynamic modeling [20, 21]. This technique uses empirical multivariate orthogonal functions. Originally, it was applied in a quasi-steady context, i.e. using the instantaneous values of the candidate regressors [29]. Later, Morelli also deployed the same method to identify unsteady model terms, including lag states of α [32]. The MOF method picks model terms from a pool of candidates in an iterative fashion. On each step of the algorithm, all candidate regressors are assessed based on their potentially beneficial contribution to reducing a cost function. Then after every iteration the best candidate is added to the model structure until the cost function is minimized. The cost function penalizes both model error and model complexity. The candidate regressors are always orthogonalized with respect to the regressors in the model structure to avoid correlated entries.

In the subsequent, regressors in physical coordinates will be referred to by $\mathbf{x}_i$, whereas orthogonalized regressors will be referred to as $\mathbf{p}_i$. At initialization, a bias term is included into the model structure:

$$\mathbf{x}_1 = \mathbf{p}_1 = \mathbf{1} \tag{24}$$

At this point and after every iteration of the algorithm, A Gram-Schmidt orthogonalization procedure is applied to all remaining candidate regressors to give us the $\mathbf{p}_j$:

$$\mathbf{p}_j = \mathbf{x}_j - \sum_{k=1}^{j-1} \gamma_{k,j} \mathbf{p}_k, \quad j = 2, 3, \ldots, n$$

where: (25)

$$\gamma_{k,j} = \frac{\mathbf{p}_k^\intercal \mathbf{x}_j}{\mathbf{p}_k^\intercal \mathbf{p}_k}$$

All orthogonalized candidates are individually assessed based on the improvement that the inclusion of the candidate yields to the predicted square error (PSE) metric. The PSE constitutes two parts; one ensuring a close fit to the data trough an MSE term, and a regularization term that penalizes model complexity. The best candidate is added to the model structure. The PSE metric is given in Equation 26, and the contribution by the jth candidate is given in Equation 27.

$$\text{PSE} = \frac{(\mathbf{y} - \hat{\mathbf{y}})^\intercal (\mathbf{y} - \hat{\mathbf{y}})}{N} + \sigma_y^2 \frac{n}{N} \tag{26}$$

$$\Delta\text{PSE}_j = -\frac{(\mathbf{p}_j^\intercal \mathbf{y})^2}{\mathbf{p}_j^\intercal \mathbf{p}_j} + \sigma_{\max}^2 \frac{1}{N} \tag{27}$$

$\hat{\mathbf{y}}$ is the model output using the model structure at the current iteration, and $\mathbf{y}$ the measurement vector, in this case always C_m. N is the number of data points, n the number of terms already included in the model and σ_{max}^2 is the upper bound variance of $\mathbf{y}$. the upper bound was taken as $\sigma_{\max}^2 = 25 \cdot \sigma_y^2$. The candidate providing the largest reduction in PSE is selected and added to the model structure. Subsequently, the algorithm is repeated until $\Delta\text{PSE}_j \geq 0$ for all j. Namely, if that condition is satisfied, there are no more candidates that justify an increased model complexity by their contribution to the model fit. The set of orthogonal vectors is now collected in matrix $\mathbf{P} = \begin{bmatrix}\mathbf{p}_1 \mathbf{p}_2 \ldots \mathbf{p}_n\end{bmatrix}$. Using ordinary least squares, a set of parameters $\hat{\boldsymbol{\phi}}$ that minimize the MSE (not the PSE) can be found.

$\hat{\boldsymbol{\phi}}$ is associated with the orthogonalized regressors in $\boldsymbol{P}$. Both can be easily transformed to physical space using Equation 28 and Equation 29, where the transformation matrix $\boldsymbol{\Gamma}$ is given by Equation 30.

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\Gamma}^{-1} \hat{\boldsymbol{\phi}} \tag{28}$$

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{\Gamma} \tag{29}$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,n} \\ 0 & 1 & \gamma_{2,3} & \cdots & \gamma_{2,n} \\ 0 & 0 & 1 & \cdots & \gamma_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{30}$$

Then the model output is computed as:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\theta}} = \boldsymbol{P}\hat{\boldsymbol{\phi}} \tag{31}$$

Due to the iterative procedure, some of the model terms added early on in the process might end up negatively contributing to the MSE. Therefore, after the model structure is finalized, it is checked for each regressor how the MSE changes if it is excluded. If the reduction in Root Mean Square (RMS) of the model output falls below a threshold of 0.5%, the term is removed, with the RMS calculated by Equation 32.

$$\text{RMS} = \frac{1}{N} \sqrt{\hat{\mathbf{y}}^\intercal \hat{\mathbf{y}}} \tag{32}$$

2. Practical Considerations

Applying MOF modeling to the entire pool of regressors in Table 4 turns out to be problematic in practice. At early iterations of the algorithm, some of the highest order model terms are selected. Then many iterations follow where a large array of model terms is added to the model structure. As a result, the model can include over 20 model terms, and parameter standard deviations of over 10 times the parameter estimates themselves. This negates the entire purpose of MOF modeling: preventing an overly complex model structure and large uncertainties in the parameter estimates. It seems that the nonlinearity of the underlying physics lies at the base of the issue. Whereas the higher order terms can be beneficial to the fit locally, they might introduce some fitting errors elsewhere. This allows for a large number of further regressors to be induced into the model structure, outweighing the model complexity penalty.

To combat the overfitting issue, the MOF algorithm is applied in two stages. Firstly, only the first two rows of the α, q and $\dot{\alpha}$ candidates from Table 4 are included in the candidate pool. Then the model terms picked in the first stage are frozen into the model structure. Now on the second stage the highest order regressors are tested. If any of them can still provide a positive contribution to the PSE, they are included as well.

Furthermore, all identification data is placed on one heap and served to the MOF algorithm at once, opposed to the approach of considering the stall sets one by one, such as other stall modeling efforts (e.g. van Ingen et al.) have done.

D. Validation Approach

The stall dynamics are very complex and each stall may very in various respects, consequently it might be difficult to see what trends the stall model has picked up from the data by merely looking at validation data. Each stall set consists of a series of cycles, as discussed in subsection V.C. The cycles all have one thing in common, i.e. the angle of attack time trace starts at some minimum, reaches a maximum, to end with another minimum. This way the aircraft oscillates in and out of the stall regime. The angle of attack can be approximated as a cosine with some mean α_0 in rad, amplitude α_1 in rad and radial frequency ω in rad/s:

$$\alpha(t) = \alpha_0 + \alpha_1 \cos \omega t \tag{33}$$

The model will be evaluated against an array of such cosine functions, with different amplitudes and frequencies.

To assess the model performance on real flight data individual stall sets will be looked at. A comparison is made to the model by van Ingen et al. [12]. The model structure that was found by van Ingen for the pitching moment is given by Equation 34.

$$C_m = C_{m_0} + C_{m_\alpha}\alpha + C_{m_{\delta_e X}} \max\left(\frac{1}{2}, X\right)\delta_e + C_{m_{C_T}} C_T \tag{34}$$

Here X represents the point of flow separation on the wing. It is an internal state and requires a nonlinear model optimization as described by van Ingen et al. and others [13, 30].

VII. Results

A. Dynamic Stall Phenomena in Flight Test Data

Some of the dynamic stall effects that have been observed in wind tunnel tests with airfoils are also present in the flight test data collected for this paper. The lift and moment hysteresis for a stall cycle in the first stall set as given by Figure 11. This stall is classified as a light stall because the counterclockwise loop at the end of the upstroke of the moment curve is missing. The plateau and subsequent vortex lift can be clearly recognized, indicating the formation and shedding of the DSV. Thereafter, the lift stall is very abrupt. Furthermore, the moment stall has been indicated as the first point where C_m decreases. The lift and moment stall occur at samples 134 and 147 of this data segment, respectively. With a sampling frequency of 100 Hz, the lag between the two events is 0.13 s.

In most deep stall cycles the lift plateau is not visible, although it is in some. Also the lift and moment stalls are much less abrupt. Figure 12 illustrates this. Note the small loop in the bottom right of the moment hysteresis, indicating deep stall. This stall only slightly penetrates into the deep stall regime, yet the stall behaviour is completely different compared to that of the light stall. The moment stall again precedes the lift stall. The moment stall indicates the point where the slope of the moment curve changes to a more negative slope. This point can be found easily as a negative peak in the second (numerical) time derivative of C_m. In this case the two events occur 0.17 s apart.

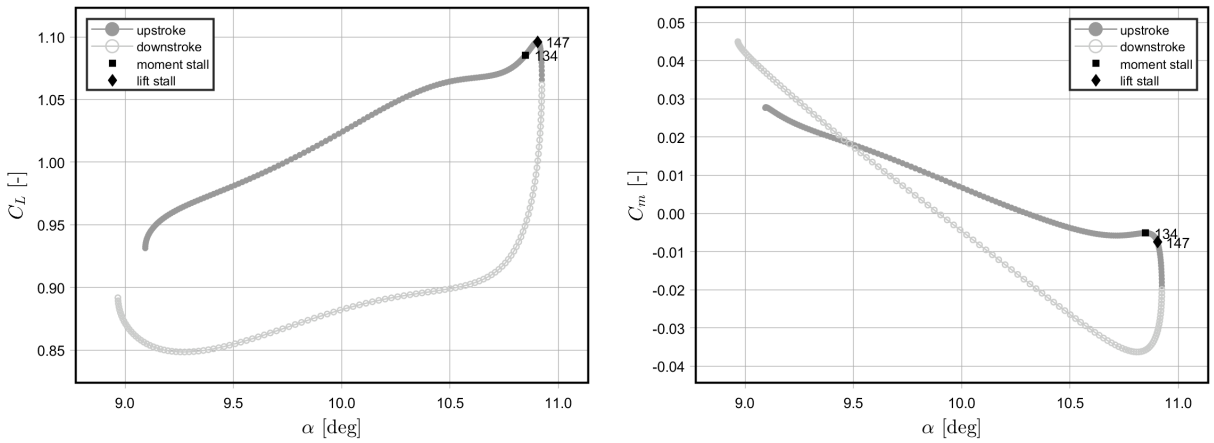


Fig. 11 Lift (left) and moment (right) hysteresis for a light stall with the PH-LAB in stall set 1.

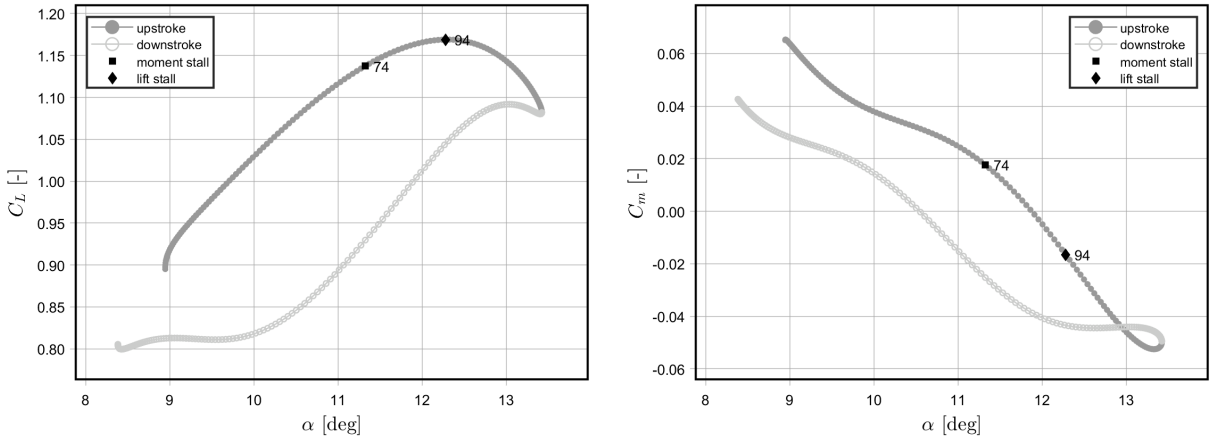


Fig. 12 Lift (left) and moment (right) hysteresis for a stall with the PH-LAB just penetrating the deep stall regime in stall set 2.

Figure 11 and Figure 12 are two stall cycles taken from accelerated stall maneuvers. Due to the large bank angle these maneuvers can not reach large values for angle of attack. The wings level stall can penetrate stall much further. Figure 13 shows one of the most severe stalls in the data set. Whereas the size of counterclockwise deep stall loops typically appears to increase with increasing angle of attack, it completely disappears for these deepest stalls. Again the lift stall, and the moment stall as defined for deep stall previously, are indicated. The lift and deep stall are now 0.63 s apart. This lag does not relate logically to moment stall caused by the DSV. Instead the moment stall in the deep stall might be a nonlinearity in the pitching moment, occurring at around 12 deg angle of attack. Yet, there is the appearance of a small plateau in the lift hysteresis. It seems that vortex formation continues to occur at higher angles of attack, but having a smaller impact. This also explains the lesser degree of abruptness in the lift and moment stalls.

B. Pitching Moment Model

The first iteration of the MOF algorithm yields the model structure in Equation 35. This model was selected from the candidate pool in Table 4, minus the highest order regressors for each α, q and $\dot{\alpha}$. On the second iteration, that did include those highest order regressors, no additional model terms were selected.

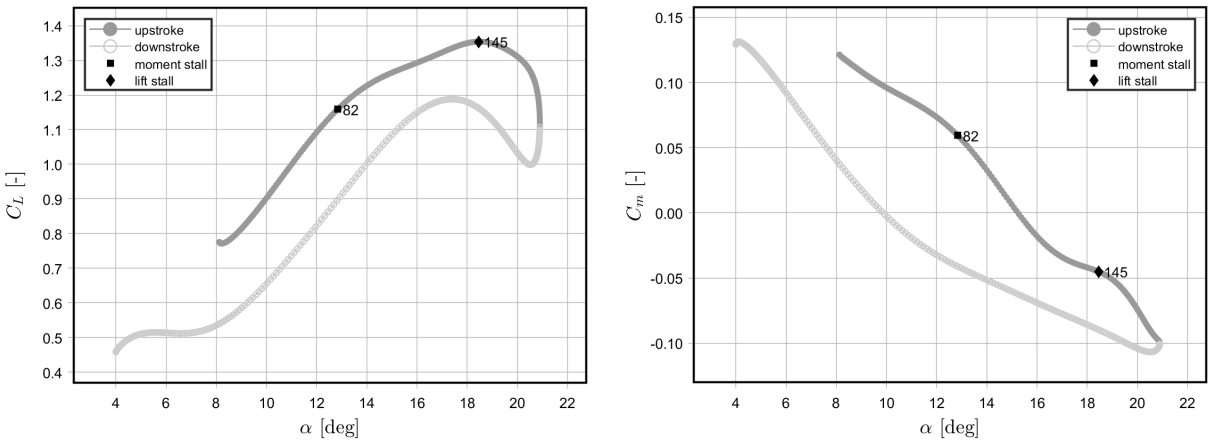


Fig. 13 Lift (left) and moment (right) hysteresis for a deep stall with the PH-LAB in stall set 22.

$$
\begin{aligned}
C_m &= C_{m0} + C_{m1\alpha}\alpha + C^*_{m1q}q + C_{m1\delta_e}\delta_e \\
&\quad + C_{m2_\alpha}\alpha + C_{m2_q}u(\alpha_t - 5^\circ)q + C_{m2_{\dot\alpha}}u(\alpha_t - 8^\circ)\dot\alpha + C_{m2_{\delta_e}}\delta_e + C_{m2_{C_T}}C_T \\
&\text{where} \\
C_{m2_\alpha} &= C_{m2_{\alpha_t}} + C_{m2_{\alpha_t^2}}\alpha_t + C_{m2_{\alpha_{t-7}\alpha_t}}\alpha_{t-7} + C_{m2_{\alpha_{t-30}\alpha_t}}\alpha_{t-30}
\end{aligned}
\tag{35}
$$

It can be seen that C_{m2_α} consists of four model terms, and three different states of α. Furthermore, a correction on $C^*_{m1q}q$ is activated starting at an angle of attack of five degrees and a dependency on $\dot\alpha$ is introduced for angles of attack larger than eight degrees. Finally, an additional elevator term is selected, and a thrust coefficient term. Parameter estimates and the corresponding parameter standard deviations are provided in Table 5.

Table 5 Stall model parameter estimates and corresponding parameter standard deviations.

Parameter	C_{m0}	$C_{m2_{\alpha_t}}$	$C_{m2_{\alpha_t^2}}$	$C_{m2_{\alpha_{t-7}\alpha_t}}$	$C_{m2_{\alpha_{t-30}\alpha_t}}$	C_{m2_q}	$C_{m2_{\dot\alpha}}$	$C_{m2_{\delta_e}}$	$C_{m2_{C_T}}$
Estimate	0.0145	0.271	93.4	-125	32.0	11.5	4.83	0.965	0.182
Standard Error	0.0216	0.297	35.9	42.7	8.66	3.57	9.92	0.153	0.0508

The correlation matrix of the stall model regressors is given in Table 6. The correlations between the various α states are striking. the cross correlations between each of them are practically unity, as is to be expected. The selection of all these regressors underlines the richness of the information content in the data. Despite these high correlations, 4 model terms could be selected without excessive parameter standard deviations. Furthermore, the correlation between the q and $\dot\alpha$ splines shows the benefit of the spline terms. Over all the identification data, the correlation between q and $\dot\alpha$ is 0.825, whereas between $u(\alpha - 5^\circ)q$ and $u(\alpha - 8^\circ)\dot\alpha$ it is only 0.729.

C. Aerodynamic Parameters

1. C_{m_α}

As C_{m2_α} is a function of multiple lag states, it is in essence a combined α and derivative parameter. In fact, this parameter includes information on the first and second derivative, because it includes three different α states. In other words, it could also be written as $C_{m2_\alpha}(\alpha, \dot\alpha, \ddot\alpha)$. This is a logical outcome since the stall cycles each have a cosine like time trace. Given the value and first two derivatives at any point of a cosine fully defines the function. That means that with the first two derivatives the time history of α in a real stall is also more or less defined.

Table 6 Stall model regressor correlation matrix.

	Variable	1	2	3	4	5	6	7	8
1	α_t	1							
2	α_t^2	0.974	1						
3	$\alpha_{t-7}\alpha_t$	0.973	0.999	1					
4	$\alpha_{t-30}\alpha_t$	0.968	0.991	0.994	1				
5	$u(\alpha - 5°)q$	0.169	0.139	0.114	0.037	1			
6	$u(\alpha - 8°)\dot{\alpha}$	-0.004	-0.005	-0.035	-0.126	0.729	1		
7	δ_e	-0.418	-0.399	-0.399	-0.401	-0.129	-0.000	1	
8	C_T	-0.826	-0.784	-0.782	-0.780	-0.178	-0.074	0.358	1

Interestingly, in the steady state C_{m2_α} reduces to a pure α parameter. Once the angle of attack is no longer subject to change, the transient washes out in at most 0.3 s. After 0.07 s the second derivative information disappears, since now α_{t-7} equals α_t, and thus only two different α values are known. After 0.3 s, α_{t-30} will be equal to α_t and C_{m2_α} is merely a function of the instantaneous angle of attack. Similarly, in slow motions where unsteady phenomena are not to be expected, the same logic holds up. The three α states will then also be approximately equal.

$$C_{m2_\alpha}\alpha = -0.0333\alpha^2 + 0.271\alpha \tag{36}$$

Equation 36 represents the moment contribution for a constant α. The parameter estimates for $C_{m2_{\alpha_t^2}}$, $C_{m2_{\alpha_{t-7}\alpha_t}}$ and $C_{m2_{\alpha_{t-30}\alpha_t}}$ approximately add up to 0, resulting in a constant steady state C_{m2_α}. Figure 14 shows this relationship.

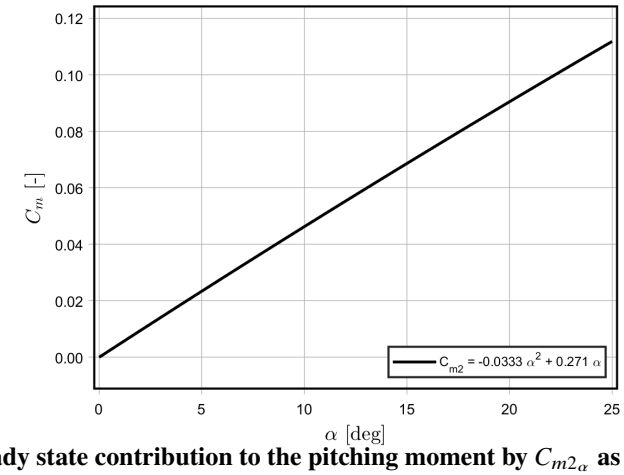


Fig. 14 Steady state contribution to the pitching moment by C_{m2_α} as a function of α.

C_{m2_α} was intended to represent the stall dynamics' deviation from the moment generated by the nominal wing and tail lift. The positive value in the steady state can be explained by a reduced wing and tail lift, leading to a smaller pitching moment around the CG.

In order to gain an insight in the unsteady modeling capacity of C_{m2_α}, the response to cosine signals with various frequencies are analyzed. Figure 15 shows five low amplitude oscillations (α_0=10° and α_1=2.5°) with various frequencies. These low amplitude oscillations are a typical representation of the α range in a 1.4g accelerated stall. The middle and lower plot show C_{m2_α} and its contribution to C_m for one period of each oscillation, respectively. The slowest signal approximates a constant C_{m2_α}, making it a steady motion since there is no dependency on time. Furthermore, it can be seen that there is a relationship between the frequency and the peak value of the negative moment. This is the same relationship that is mentioned in the works of the likes of McCroskey and Carr [14, 22].

Figure 16 shows the exact same relationships but for a large amplitude oscillation with a higher mean (α_0=15° and α_1=5°), that is more representative of a wings-level stall. The relationship between the negative moment peak and the oscillation frequency is also here clear. The major difference is the magnitude of the peaks. The peaks in the large amplitude oscillations reach values four times higher than the low amplitude oscillations.

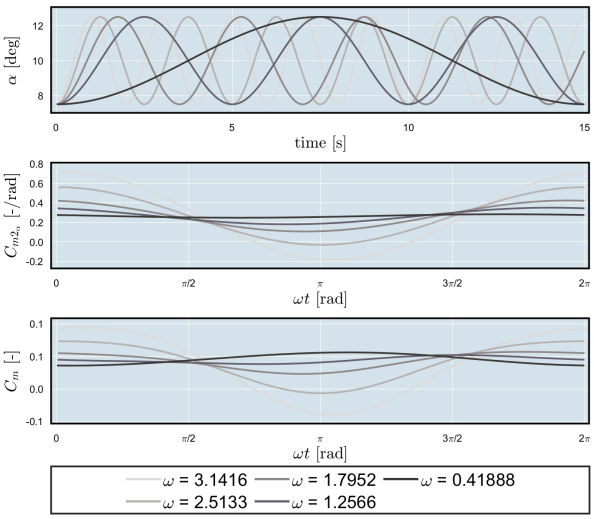


Fig. 15 Signal $\alpha = 10 - 2.5\cos\omega t$ for 5 frequencies (top) and corresponding C_{m2_α} (middle) and C_m (bottom) throughout the stall cycle.

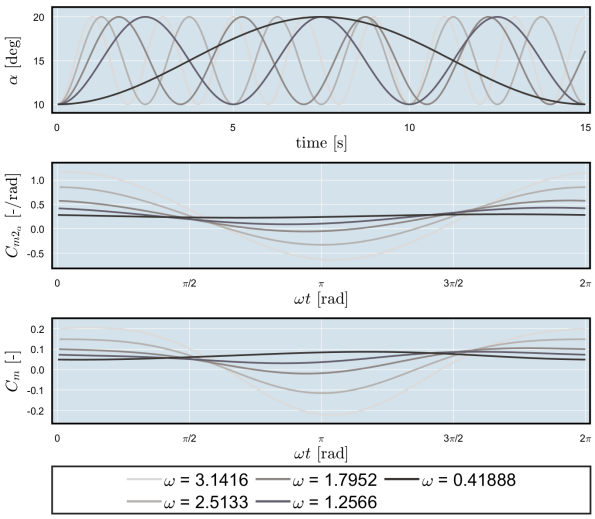


Fig. 16 Signal $\alpha = 15 - 5\cos\omega t$ for 5 frequencies (top) and corresponding C_{m2_α} (middle) and C_m (bottom) throughout the stall cycle.

Figure 17 show the contributions to the pitching moment plotted against α as modeled by C_{m2_α} for the high amplitude motion from Figure 16, but only for the three highest frequencies. An interesting pattern emerges. As we had seen before, the contribution to C_{m2} from the slower motions is not significant. However, Figure 17-a reveals that this slower motion contributes to a counter clockwise hysteresis loop, associated with stable damping. On the other, the slightly faster motion in Figure 17-b exhibits no hysteresis. Finally, the fastest motion in Figure 17-c is characterized by a loop reversal, i.e. the hysteresis is now clockwise. The offset in moment can be explained by different wing lift force and moment arm during the up- and downstroke. The tail lift is assumed independent of the prehistory of the angle of attack, and therefore no different during up- and downstroke. With only the moment curves, it is difficult to say what phenomenon exactly causes the loop reversal. A model for the lift hysteresis could be insightful since it would exclude the effects of the moving center of pressure.

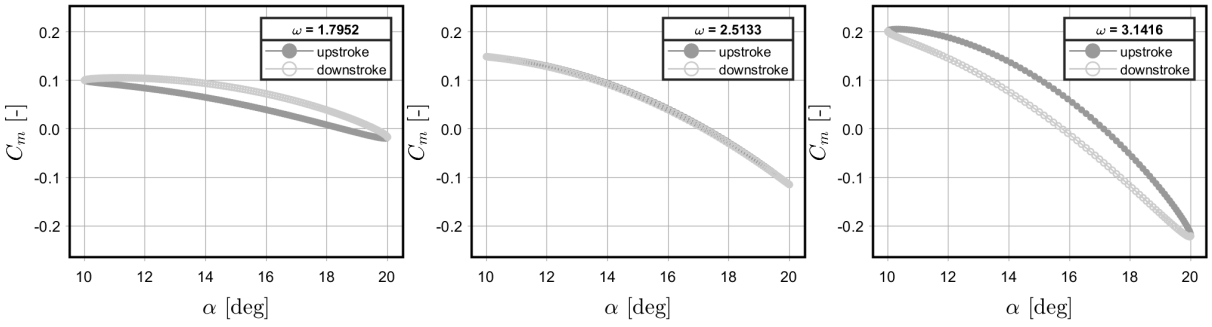


Fig. 17 Moment hysteresis as modeled by C_{m2_α} for signal $\alpha = 15 - 5\cos\omega t$ with frequencies from left to right: $\omega = 1.7952$, $\omega = 2.5133$ and $\omega = 3.1416$.

2. C_{m_q} and $C_{m_{\dot\alpha}}$

As indicated in Equation 35 both an additional pitch rate as angle of attack rate parameter were selected in the model structure selection process. The stall induced pitch damping is modeled by C_{m2_q} as a spline of degree 0 in α, so that an additional constant contribution is introduced at angle of attack larger than 5°.C_{m2_q} practically cancels the nominal pitch damping C_{m1_q}: The nominal value stands at -11.9 and the stall parameter increases it by 11.5 to -0.4. Witch the

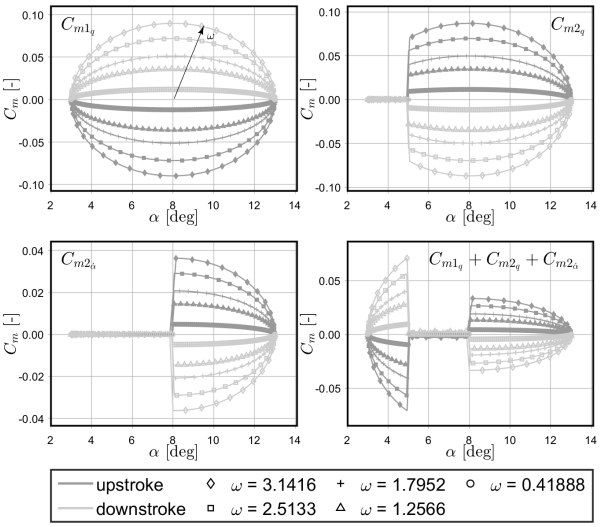


Fig. 18 Moment hysteresis as modeled by the three aerodynamic rate derivatives for signal $\alpha = 8 - 5cos(\omega t)$ with 5 frequencies.

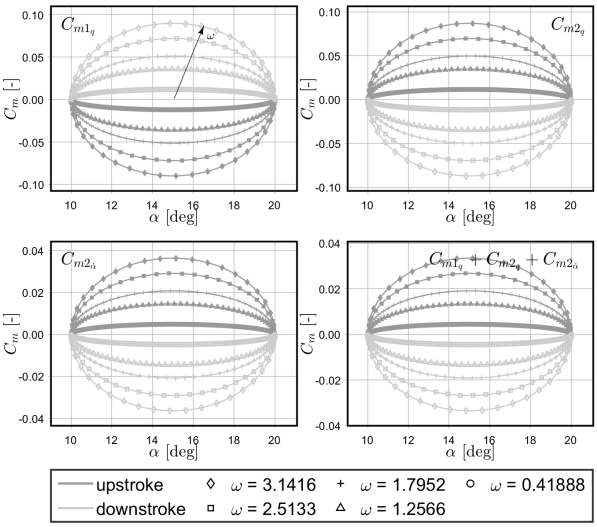


Fig. 19 Moment hysteresis as modeled by the three aerodynamic rate derivatives for signal $\alpha = 15 - 5cos(\omega t)$ with 5 frequencies.

pitch damping defined is given by Equation 3, this result implies that the tail lift becomes constant at angles of attack above 5°. It is not very probable that this is indeed the case as this point falls well within the attached flow regime. It should be noted that this result is only a low order approximation and it is likely that more and higher order pitch damping terms are needed to capture the exact relationship. However, given the data this is not realistic.

A similar zero order spline term is selected for the downwash lag coefficient $C_{m2_{\dot\alpha}}$. With a value of 4.83 the total damping from the rate derivatives is taking a positive, and thus aerodynamically unstable, value at angles of attack above 8°. Given Equation 4 and the earlier observed pitch damping, this result suggest a decreased downwash gradient.

Despite the introduction of spline terms, q and $\dot\alpha$ are significantly correlated (see Table 6). Additionally, the combined α derivatives also hold information about the angle of attack rate. It is therefore plausible that some of the contributions are exchanged between the various model terms.

Similar to subsubsection VII.C.1, some artificial cosine signals are used to investigate the rate contributions. The derivative of the standard cosine signal from Equation 33 is normalized by $\frac{\bar{c}}{V_{\text{stall}}}$ with $V_{\text{stall}} \simeq 75$ m/s to obtain a unitless $\dot\alpha$. Figure 18 shows hysteresis plots of osillations that cover both the C_{m2_q} and $C_{m2_{\dot\alpha}}$ knots. The contributions from the nominal and 2 stall contributions are shown. In the bottom right the combined effect is shown. The advantage of this model is that once the angle of attack falls below 5° the exact same damping characteristics are restored as in the nominal model. It can be seen that higher frequency motions have a larger swept out area, as is to be expected.

Figure 19 shows the same hysteresis plots for a signal with a higher mean at $\alpha_0 = 10°$. This is a good representation of a deeper stall. It shows that these stall completely operate in the regime with the unstable aerodynamic damping.

The dependency in Figure 18 and 19 on the motion frequency is straightforward. These derivatives only describe a quasi-steady contribution to the pitching moment. Therefore it makes sense that there are no surprising frequency dependent effects. These terms model the volume in the moment hysteresis, while C_{m2_α} can capture more intricate time dependent relationships.

D. Hysteresis Modeling

The motions considered in subsection VII.C are artificial and idealized. In the subsequent, the modeling capacity of the α and rate derivatives is analysed for some real flight test data, to see how the model handles the hysteresis observed in the pitching moment.

Figure 20 shows the contributions by the α related model parameters for the modeling of the stall cycle in Figure 12 discussed previously. In the top row, the various lag states of α are shown, as well as their individual contributions to the pitching moment. $C_{m_{\alpha_t}}$ has a small contribution in the stall. As explained in subsubsection VII.C.1, this parameter mainly provides a correction on the steady state behaviour, where the other three contributions add up to approximately

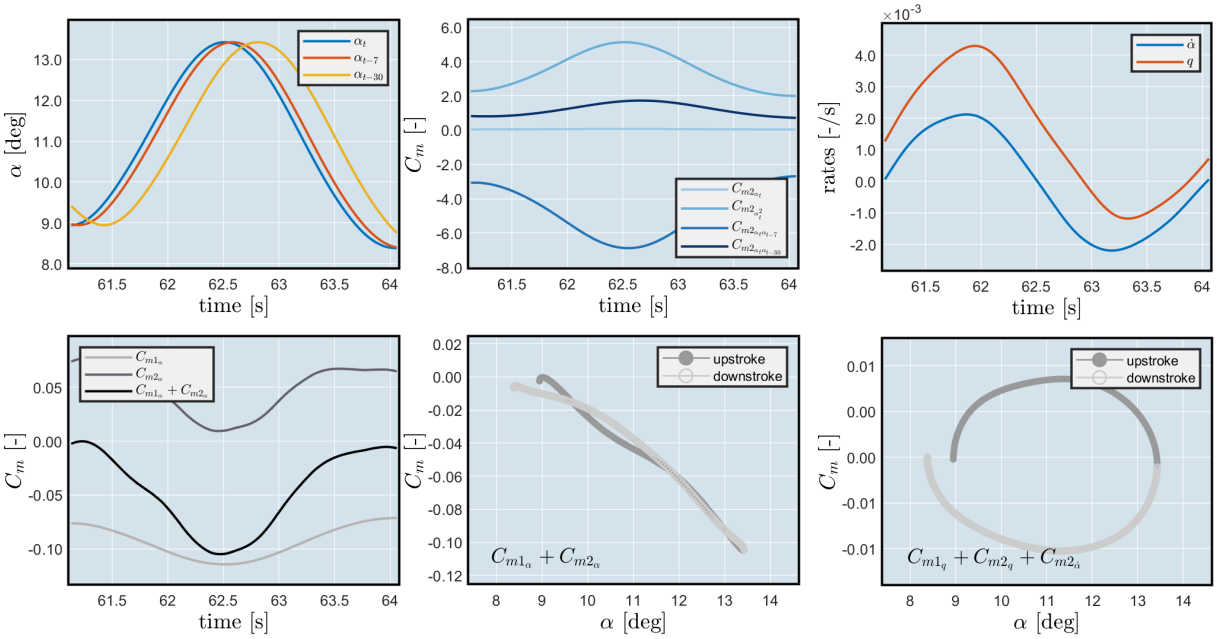


Fig. 20
Contribution to the pitching moment by by α for the deep stall in Figure 12.

Fig. 21 Contribution to the pitching moment by by q and $\dot\alpha$ for the deep stall in Figure 12.

zero. $C_{m_{\alpha_t^2}}$, $C_{m_{\alpha_{t-7}\alpha_t}}$ and $C_{m_{\alpha_{t-30}\alpha_t}}$ have large contributions that individually lack real physical interpretation. However, jointly they can capture subtle changes in the pitching moment. Compare the moment hysteresis curve in Figure 12 to the one in Figure 20 and note how subtle changes in the slope of the pitching moment are modeled. Barely any hysteresis is modeled by C_{m2_α}. That contribution is provided by q and $\dot\alpha$, as shown in Figure 21.

Figure 23 is an even better display of the collaborative effort provided by $C_{m_{\alpha_t^2}}$, $C_{m_{\alpha_{t-7}\alpha_t}}$ and $C_{m_{\alpha_{t-30}\alpha_t}}$. It shows the stall model contributions for the modeling of the data in Figure 13. Whereas the α contributions in Figure 20 joined to form a single negative peak, in Figure 22 three small peaks form. The α regressors also model various hysteresis loops: a small and large counter clockwise, and a small clockwise one. Again the volume in the moment hysteresis is captured by the rate derivatives as shown in Figure 23.

E. Model Validation

The model presented in the form of Equation 35 provides an alternative to the Kirchoff based method. To assess its predictive capabilities the MSE of the C_m predictions resulting from this model and the one by van Ingen et al. [12] are compared. Table 7 displays the model performances in terms of MSE for both identification and validation data.

Clearly, the model presented in this paper is the undisputed champion when it comes down to minimizing the MSE. Over all the identification data it provides an improvement of 64.3% while over all validation data an improvement of 55.9% can be reported. It is important to note that a large part of the performance improvement is thanks to the modeling performance outside the stalls. Every stall set includes a significant portion of data pre- and post-stall. The top plot in Figure 24 shows an example. These are the time traces of C_m of validation stall 14. Especially in the post stall, the model is much closer to the measurements. In the isolated stall cycles, it outperforms the model by van Ingen as well, although the difference is much less striking in the wings-level stalls.

The bottom plot in Figure 24 reveals a model deficiency in both models, although much more severe in the model by van Ingen. This shows the same data as the top plot, but zoomed in on the stall cycles. The model predictions seem extremely noisy. This is the direct result of the quasi-random inputs applied to the elevator. It is likely that the quasi random inputs are too high frequency, not allowing the pitching moment to adapt to the input. As a result only the bias applied to the control column exerts a direct influence on the pitching moment. This new model accounts for that with a

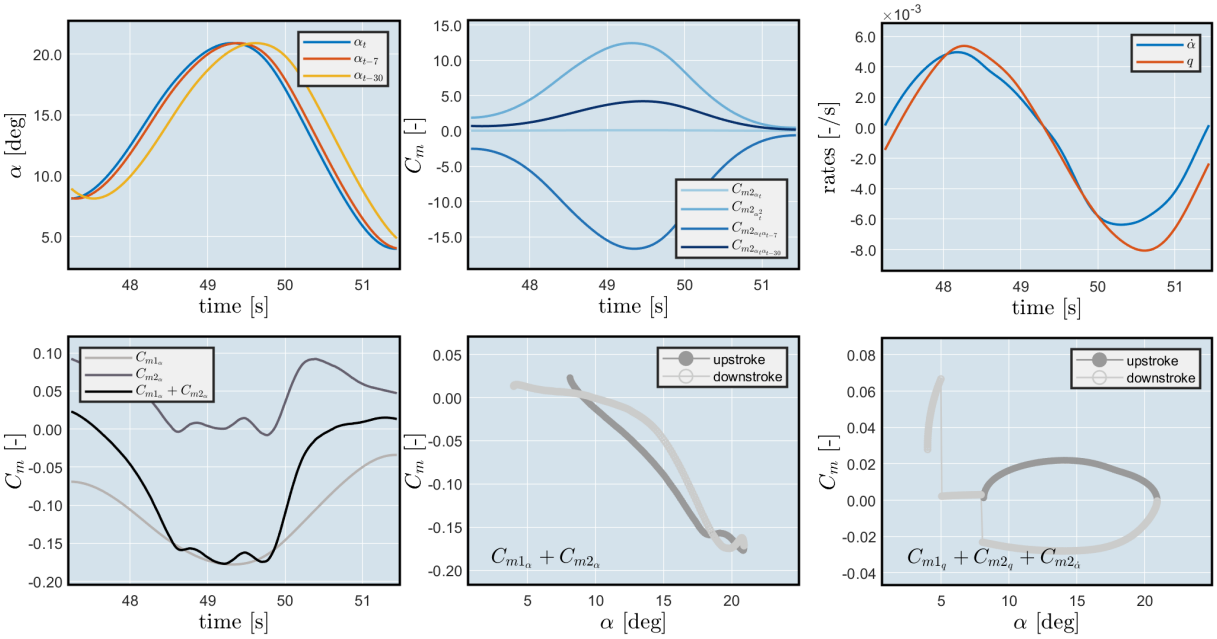


Fig. 22
Contribution to the pitching moment by by α for the deep stall in Figure 13.

Fig. 23 Contribution to the pitching moment by q and $\dot{\alpha}$ for the deep stall in Figure 13.

Table 7 Identification and Validation MSE for the model presented in this paper and the model by van Ingen et al. [12].

Data	MSE model current paper	MSE model van Ingen et al. [12]	Difference [%]
Identification data			
Asymmetric stalls (12 stalls)	$4.50{\cdot}10^{-5}$	$1.82{\cdot}10^{-4}$	-75.2
Wings-level stalls (9 stalls)	$6.07{\cdot}10^{-5}$	$1.06{\cdot}10^{-4}$	-42.7
All identification data	$5.15{\cdot}10^{-5}$	$1.44{\cdot}10^{-4}$	-64.3
Validation data			
Stall 2	$3.93{\cdot}10^{-5}$	$4.36{\cdot}10^{-5}$	-9.92
Stall 6	$3.67{\cdot}10^{-5}$	$7.17{\cdot}10^{-5}$	-48.8
Stall 14	$4.36{\cdot}10^{-5}$	$2.69{\cdot}10^{-4}$	-83.8
Stall 19	$2.32{\cdot}10^{-5}$	$1.43{\cdot}10^{-4}$	-83.8
Stall 24	$1.08{\cdot}10^{-4}$	$8.45{\cdot}10^{-5}$	+28.4
All validation data	$4.89{\cdot}10^{-5}$	$1.11{\cdot}10^{-4}$	-55.9

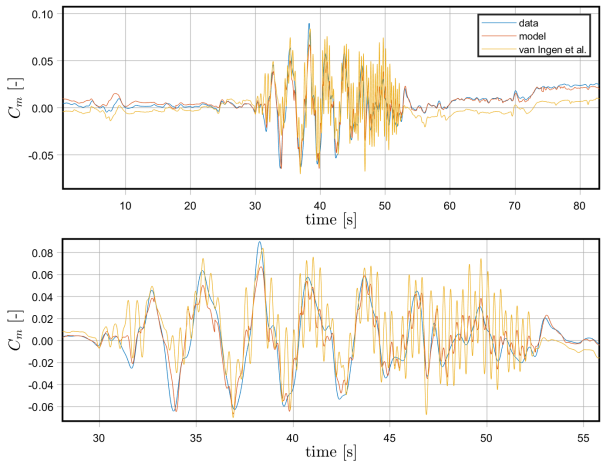


Fig. 24 Model fit on validation set 14 of full time trace (top) and close up stall maneuvers (bottom).

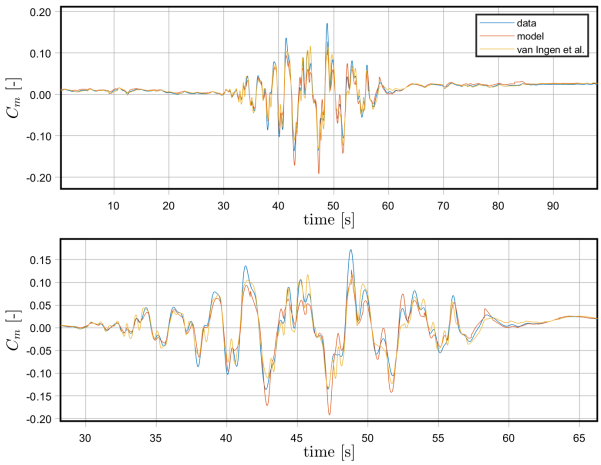


Fig. 25 Model fit on validation set 24 of full time trace (top) and close up stall maneuvers (bottom).

drastic reduction in the elevator effectiveness through $C_{m2_{\delta_e}}$. Whereas the nominal model reports $C_{m1_{\delta_e}} = -1.25$, the stall model reduces it with 0.965 to only -0.285. On this stall, the model improves over 83% on van Ingen's model mainly due to its inability to model the elevator contribution.

The large modeling errors due to the elevator disturbances only seem to appear when the inputs exceed a certain magnitude. For smaller amplitude inputs the problem does not become evident. Only stall sets 14 and 19 exceeded this apparent threshold, leading to a MSE improvement of 83.4 % in both cases.

The model by van Ingen outperforms the new model in one of the 5 validation sets, i.e. set 24. Figure 25 shows the C_m time traces and model predictions. Although the new model picks up on variations in the pitching moment, it has trouble correctly predicting the C_m peaks, both positive and negative.

Finally, the time traces and fits by the model presented in this paper and the model by van Ingen for the remaining three validation sets can be found in Figure 26-28.

VIII. Conclusion

The aim of this paper was to present an aerodynamic model that can capture the nonlinear unsteady characteristics of the pitching moment during stall, especially aerodynamic damping, of the Cessna Citation II. Additionally, the model was intended to possess global modeling capacity and high physical interpretability. This model is meant to serve as an alternative to Kirchoff-based models that are lacking with regard to those objectives. The model proposed is based on aerodynamic lag states to capture time varying effects. Furthermore, it is based on the linear stability derivative formulation that has been prevalent since the early days of aviation. It includes contributions identical to the nominal pitching moment model, augmented with correction terms that model deviations from linear attached flow in the nonlinear unsteady regime. Consequently, the model is characterized by high physical interpretability and global modeling capability. This mathematical model was used to generate a large pool of candidate regressor terms, out of which the best model terms were selected using Multivariate Orthogonal Function modeling. The combined effect of flow separation and the moving center of pressure is modeled by means of lag states of the angle of attack, where lags of 0.3 s, 0.07 s as well as the instantaneous angle of attack were included. Three lag states means that the model picks up on information up to the second derivative of the angle of attack. It is demonstrated how this model is sensitive to the frequency of the angle of attack time trace, and how it represents nonlinear and time-dependent aerodynamics. Furthermore, contributions were found for the pitch damping in stall for angles of attack over 5° and a downwash lag damping term for angles of attack over 8°. Additionally, a thrust coefficient parameter and a correction on the elevator effectiveness were found to be beneficial for model fit. This model was found to provide a good model fit on both identification and validation data. In terms of validation MSE, this model provides an improvement of 55.9% over the Kirchoff based pitching moment stall model by van Ingen et al. [12].

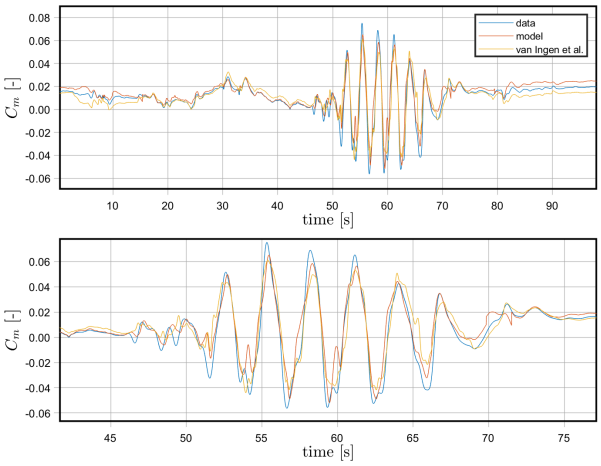


Fig. 26 Model fit on validation set 2 of full time trace (top) and close up stall maneuvers (bottom).

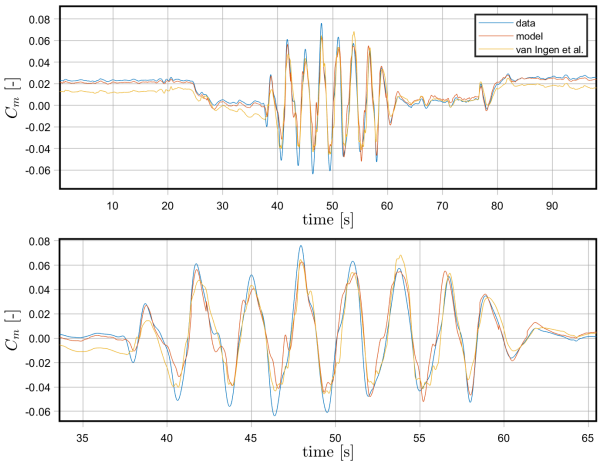


Fig. 27 Model fit on validation set 6 of full time trace (top) and close up stall maneuvers (bottom).

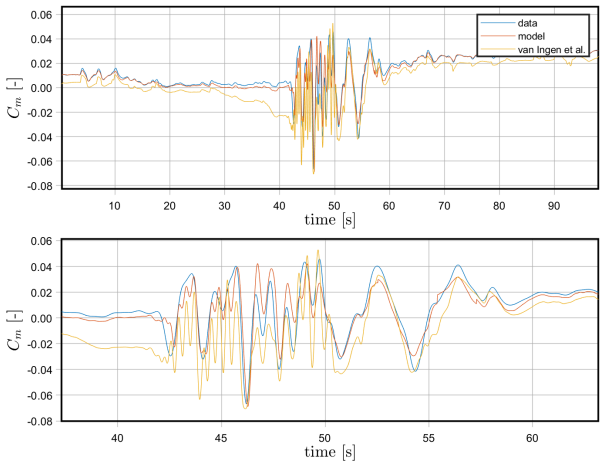


Fig. 28 Model fit on validation set 19 of full time trace (top) and close up stall maneuvers (bottom).

References

[1] "Safety Report 2019," Safety Report Edition 56, International Air Transport Association, Montreal, Quebec, Canada, 2020.

[2] "Federal Aviation Administration. Advisory Circular 120-109A - Stall Prevention and Recovery Training," Tech. rep., Federal Aviation Administration, 2015.

[3] "EASA. Certification specifications for aeroplane flight simulation training devices," Tech. Rep. CS-FSTD(A), European Aviation Safety Agency, 2018.

[4] Advani, S., and Field, J., "Upset Prevention and Recovery Training in Flight Simulators," Portland, US, 2011. https://doi.org/10.2514/6.2011-6698.

[5] J. R. Chambers, and S. B. Grafton, "Aerodynamic Characteristics of Airplanes at High Angles of Attack," *Journal of Aircraft*, 1977, pp. 1109–1115.

[6] Tobak, M., and Schiff, L. B., "On the formulation of the aerodynamic characteristics in aircraft dynamics," Techmical Report TR R-456, NASA, Moffet Field, California, USA, 1976.

[7] Klein, V., and Noderer, K. D., "Modeling of Aircraft Unsteady Aerodynamic Characteristics. Part I - Postulated Models," Technical Memorandum TM-109120, Langley Research Center, NASA, Hampton, Virginia, USA, 1994.

[8] Goman, M., and Khrabrov, A., "State-space representation of aerodynamic characteristics of an aircraft at high angles of attack," *Journal of Aircraft*, Vol. 31, No. 5, 1994, pp. 1109–1115. https://doi.org/10.2514/3.46618, URL https://doi.org/10.2514/3.46618, publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/3.46618.

[9] Abramov, N., Goman, M., Khrabrov, A., and Soemarwoto, B., "Aerodynamic Modeling for Poststall Flight Simulation of a Transport Airplane," *Journal of Aircraft*, Vol. 56, 2019, pp. 1–14. https://doi.org/10.2514/1.C034790.

[10] Klein, V., "Modeling of Longitudinal Unsteady Aerodynamics of a Wing-Tail Combinat," Contractor Report CR-1999-209547, Langley Research Center, NASA, Hampton, Virginia, USA, 1999.

[11] Murphy, P., and Klein, V., "Estimation of Longitudinal Unsteady Aerodynamics of a Wing-Tail Combination from Wind Tunnel Data," *AIAA Atmospheric Flight Mechanics Conference and Exhibit*, American Institute of Aeronautics and Astronautics, Keystone, Colorado, 2006. https://doi.org/10.2514/6.2006-6154, URL https://arc.aiaa.org/doi/10.2514/6.2006-6154.

[12] van Ingen, J., de Visser, C. C., and Pool, D. M., "Stall Model Identification of a Cessna Citation II from Flight Test Data Using Orthogonal Model Structure Selection," *AIAA Scitech 2021 Forum*, American Institute of Aeronautics and Astronautics, 2021. https://doi.org/https://doi.org/10.2514/6.2021-1725.

[13] Fischenberg, D., "Identification of an unsteady aerodynamic stall model from flight test data," *20th Atmospheric Flight Mechanics Conference*, American Institute of Aeronautics and Astronautics, 1995. https://doi.org/10.2514/6.1995-3438, URL https://arc.aiaa.org/doi/10.2514/6.1995-3438.

[14] Mccroskey, W. J., "The Phenomenon of Dynamic Stall." Technical Memorandum, Ames research center, NASA, Moffet Field, California, USA, 1981.

[15] Khrabrov, A., Vinogradov, Y., and Abramov, N., "Mathematical Modelling of Aircraft Unsteady Aerodynamics at High Incidence with Account of Wing-Tail Interaction," AIAA Atmospheric Flight Mechanics Conference, Provedence, Rhode Island, USA, 2004.

[16] Corke, T., and Thomas, F., "Dynamic Stall in Pitching Airfoils: Aerodynamic Damping and Compressibility Effects," *Annual Review of Fluid Mechanics*, Vol. 47, 2015. https://doi.org/10.1146/annurev-fluid-010814-013632.

[17] Maine, R. E., and Iliff, K. W., "Maximum Likelihood Estimation of Translational Acceleration Derivatives from Flight Data," *Journal of Aircraft*, Vol. 16, No. 10, 1979, pp. 674–679. https://doi.org/10.2514/3.58588, URL https://arc.aiaa.org/doi/10.2514/3.58588, publisher: American Institute of Aeronautics and Astronautics.

[18] Jategaonkar, R. V., and Gopalratnam, G., "Two complementary approaches to estimate downwash lag effects from flight data," *Journal of Aircraft*, Vol. 28, No. 8, 1991, pp. 540–542. https://doi.org/10.2514/3.46060, publisher: American Institute of Aeronautics and Astronautics.

[19] Grauer, J. A., Morelli, E. A., and Murri, D. G., "Flight Test Techniques for Quantifying Pitch Rate and Angle of Attack Rate Dependencies," *Journal of Aircraft*, Vol. 54, No. 6, 2017.

[20] Morelli, E. A., "Global nonlinear aerodynamic modeling using multivariate orthogonal functions," *Journal of Aircraft*, Vol. 32, No. 2, 1995, pp. 270–277. https://doi.org/10.2514/3.46712.

[21] Klein, V., and Morelli, E. A., *Aircraft system identification - theory and practice*, AIAA education series, Vol. 2, Sunflyte Enterprises, Williamsburg, VA, USA, 2016. OCLC: 664103034.

[22] Carr, L. W., "Progress in analysis and prediction of dynamic stall," *Journal of Aircraft*, Vol. 25, No. 1, 1988, pp. 6–17. https://doi.org/10.2514/3.45534.

[23] Choudhry, A., Leknys, R., Arjomandi, M., and Kelso, R., "An insight into the dynamic stall lift characteristics," *Experimental Thermal and Fluid Science*, Vol. 58, 2014, pp. 188–208. https://doi.org/10.1016/j.expthermflusci.2014.07.006.

[24] Ericsson, L. E., and Reding, J. P., "Fluid mechanics of dynamic stall part I. Unsteady flow concepts," *Journal of Fluids and Structures*, Vol. 2, No. 1, 1988, pp. 1–33. https://doi.org/10.1016/S0889-9746(88)90116-8.

[25] McCroskey, W. J., McAlister, K. W., Carr, L. W., Pucci, S. L., Lambert, O., and Indergrand, R. F., "Dynamic Stall on Advanced Airfoil Sections," *Journal of the American Helicopter Society*, Vol. 26, No. 3, 1981, pp. 40–50. https://doi.org/10.4050/JAHS.26.40.

[26] Liiva, J., Davenport, F., Gray, L., and Walton, I., "Two-dimensional tests of airfoils oscillating near stall. Volume 1. Summary and evaluation of results," 1968.

[27] Higgins, R., Barakos, G., and Filippone, A., "A review of propeller stall flutter," *Aeronautical Journal*, Vol. 126, 2022. https://doi.org/10.1017/aer.2022.12.

[28] Etkin, B., *Dynamics of Atmospheric Flight*, Dover Publications, 2000. URL https://app-knovel-com.tudelft.idm.oclc.org/kn/resources/kpDAF00013/toc.

[29] Morelli, E., Cunningham, K., and Hill, M., "Global Aerodynamic Modeling for Stall/Upset Recovery Training Using Efficient Piloted Flight Test Techniques," 2013. https://doi.org/10.2514/6.2013-4976.

[30] Van Horssen, L., De Visser, C. C., and Pool, D. M., "Aerodynamic Stall and Buffet Modeling for the Cessna Citation II Based on Flight Test Data," *2018 AIAA Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, Kissimmee, Florida, 2018. https://doi.org/10.2514/6.2018-1167, URL https://arc.aiaa.org/doi/10.2514/6.2018-1167.

[31] van den Hoek, M., de Visser, C., and Pool, D., "Identification of a Cessna Citation II Model Based on Flight Test Data," *4th CEAS Specialist Conference on Guidance, Navigation and Control*, Warsaw, Poland, 2017.

[32] Morelli, E. A., "Nonlinear Unsteady Aerodynamic Modeling using Empirical Orthogonal Functions," Technical Memorandum TM–20210009593, Langley Research Center, NASA, Hampton, Virginia, USA, 2021.

Part III

Preliminary Analysis

*This part has been assessed for the course AE4020 Literature Study.

1

Introduction

1.1. Introduction

Colgan Air Flight 3407 crashed on the 12th of February in 2009 with 4 crew members and 45 passengers on board. During final approach, the pilot in command was being warned by the stick shaker that he was near stall entry. The pilot reacted by abruptly pulling back on the control column and increasing thrust to 75% power. Consequently, the aircraft pitched up even further. To make matters worse, the pilot overrode the stick pusher that was meant to aid recovering stall, and the first officer retracted the flaps making recovery even more challenging. The pilots lost control of the aircraft and moments after crashed into a home near the airport. No one inside the aircraft survived. [1].

The pilots did not follow the right protocol for stall recovery. This example is not an incident; Loss of Control in Flight (LOC-I) bears the most common cause of fatal accidents, and the main contributor is aerodynamic stall [2]. After the Colgan Air crash, the Airline Safety and Federal Aviation Administration Extension Act was adopted in the United States. Part of this act focuses on realizing improved stall and upset recovery training [3]. Later, in 2019, flight simulator-based stall prevention and recovery training has become obligatory in the US [4] and similar legislation was passed in the European Union [5]. Simulator based training requires stall models [6]. This gives rise to an increased demand in accurate stall models for aircraft.

Aerodynamic models can be obtained from computational fluid dynamics (CFD), wind tunnel tests and flight-test data based methods. CFD and wind tunnel test approaches currently can not capture critical damping and nonlinear aerodynamic characteristics [10]. Flight-test data based stall model identification does not suffer from these limitations, and therefore it is an essential source for stall models. These flight-test data based stall models have drawbacks of their own. It can be a difficult task to collect high quality data with the right identification content. Also making global models that cover both the nominal and high angle of attack envelope is a source for headaches.

In the pursuit of developing a good stall model from flight-test data, a stall task force was founded at the section of Control & Simulation at Delft University of Technology. The task force has developed a number of stall models in recent years for the Cessna Citation II PH-LAB. This laboratory aircraft has been employed numerous time for the purpose of collecting stall model identification data. Van Horssen et al. [7] presented a buffet model and a first longitudinal stall model. Van Ingen et al. [8] further improved this longitudinal stall model and de Fuijk et al. [9] added a lateral stall model. The models have a high accuracy in the stall regime, but are lacking in other flight conditions. These stall models are local models that have been developed for a specific purpose. This leaves the need for a global stall model, so that pilots can be exposed to a fully continuous experience from nominal flight to pre-stall, stall and post-stall.

One promising modeling method is the multivariate simplex B-spline [11]. They can be used to construct global models with an arbitrarily high approximation power and can efficiently handle large amounts of data. The main limitations of these splines is that they tend to have problems fitting collinear and sparse data sets. Problematically, stall flight-test data sets tend to be sparse. Steiner et al. presented a solution by implementing Tikhonov regularization in the simplex B-spline system identification framework [12]. This finally opens the gates for global stall modeling using multivariate simplex B-splines. The remainder of this report provides an overview of all the relevant literature stall modeling from flight-test data, multivariate simplex B-splines and regularization.

1.2. Research Goal and Questions

The main goal of this thesis is to develop a global aerodynamic model that can be used to describe both nominal flight and all phases of stall. The main hypothesis for this goal is that this can best be done using multivariate simplex B-splines in combination with Tikhonov regularization.

To establish the scope of this thesis the main research question is posed as follows:

How can a full flight envelope model including both steady and unsteady aerodynamics of a Cessna Citation II aircraft best be identified from flight test data using a global modeling method?

In order to give a comprehensive and well substantiated answer to this question, and to provide focus points for the research, the main research question is broken down in the following sub-questions:

1. What are the advantages and disadvantages of using Multivariate Simplex B-splines for identifying and validating a full flight envelope model from sparse flight-test data of the Cessna Citation II aircraft?

2. What flight test maneuvers are suitable for identifying both steady and unsteady, and both longitudinal and lateral aerodynamics?

3. Which geometric and polynomial model structure can best be used for steady and unsteady aerodynamic modeling with multivariate simplex B-splines?

4. How does a multivariate simplex B-spline model compare to the local stall models by van Ingen et al. [8] and de Fuijk et al. [9]?

5. How can local regularization enable simplex B-spline modelling of sparse and ill-conditioned flight datasets?

1.3. Report Outline

First of all, the stall phenomenon and existing stall models are discussed in Chapter 2. In Chapter 3 the acquisition and processing of stall data is explained. Next, the Simplex B-Spline theory is covered and compared to other global modeling methods in Chapter 4. Then in Chapter 5 the concept of ill-posed regression is talked about and a review of regularization is found in Chapter 6. Lastly, in Chapter 7 conclusions to the answers of the research questions are provided as far as possible with the knowledge from the literature review. This chapter also provides a research plan that will be leading in fully answering the remaining questions.

2

Stall Theory and Models

Accurate stall models are vital for a high fidelity stall simulation. This chapter is concerned with providing an overview of the aerodynamic stall phenomenon, and an overview of the state of the art in stall modeling. Section 2.1 talks about all the aerodynamic aspects of stall and Section 2.2 about what is expected of a good stall model. Section 2.3 covers a theoretical stall model by Kirchoff, and Section 2.4 and Section 2.5 some application of Kirchhoff's model to flight test data based models. Section 2.6 discusses the difference between local and global stall models, and finally a conclusion on all findings in this chapter is found in Section 2.7.

2.1. Aerodynamics of Stall

To understand what a stall is and how it arises, some aerodynamic principles need to be defined. This is done in Section 2.1.1. To understand how it affects the aircraft throughout the flight envelope, a comprehension of how stall develops longitudinally and laterally is required. This is the aim of Section 2.1.2. Lastly, some more advanced stall phenomena are highlighted in Section 2.1.3

2.1.1. Stall Defined

It is well known that lift is a linear function of angle of attack. At the same time, the induced drag increases quadratically with angle of attack. That is, up to the stall angle of attack. After this point, the air flow starts to separate from the wing and highly non-linear phenomena can be observed. The most striking effects are a drastic decrease in lift and a sudden increase in drag [13]. This development in the lift-curve for a symmetric airfoil is shown in Figure 2.1. The non-linear flight regime is in fact not only non-linear, it also highly unpredictable and time-varying. The time-varying response is determined by the motion pre-history and is referred to as unsteady aerodynamics [14, 15]. At the point where the lift starts to decrease, the aircraft is said to have been stalled. When the aircraft has stalled, the airflow has separated from the wing. This flow separation is due to viscous flow effects. A viscous flow experiences dissipative effects due to friction, thermal conduction, or mass diffusion [13].

In front of the wing the air flow is at free-stream velocity. The aircraft wing body interferes with the flow once the air reaches the vicinity of the wing. Near the surface a boundary layer forms. The airflow right at the surface interacts with the wing body, causing friction. This friction retards the flow velocity and this effect propagates away from the wing surface. At the wing surface the flow velocity will be zero to then further increase until free stream velocity is reached at the edge of the boundary layer. This is called a velocity profile.

One characteristic of viscous flow is the formation of an adverse pressure gradient on the wing. This means that the static pressure increase in the direction of the flow. In other words, at some point in the boundary layer the pressure downstream is greater than upstream, and the flow has to withstand this increasing pressure to progress. The adverse pressure gradient further retards the flow and the velocity profile becomes a function of the longitudinal location on the wing. As the angle of attack increases, the adverse pressure gradient at the top of the wing grows stronger. At some point, the adverse pressure gradient has grown so strong that it forces the flow direction to reverse. At this point where flow reversal occurs, the boundary layer separates from the wing and the aircraft enters stall. A separation region forms where a large turbulent wake is present. This process is visualized in Figure 2.2.

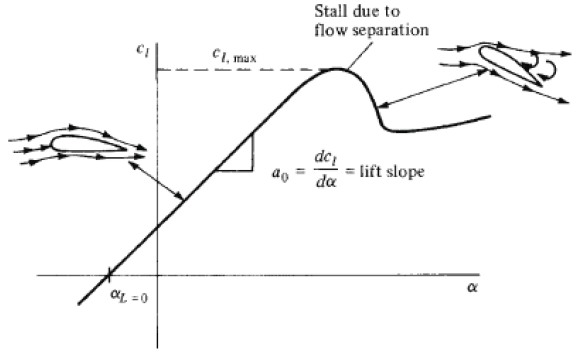


Figure 2.1: Static lift curve for a cambered airfoil, taken from [13].

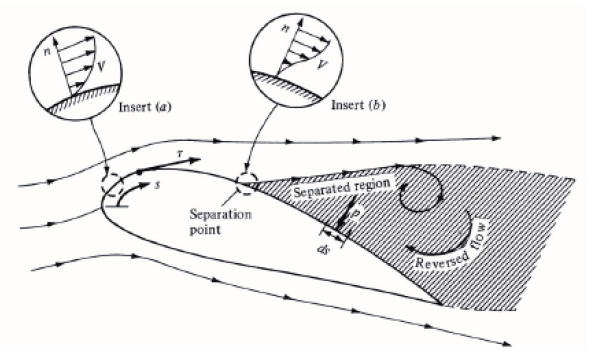


Figure 2.2: Airfoil with partially separated boundary layer, taken from [13].

In addition to drastic changes in aerodynamic forces, the stall is accompanied by another unmistakable phenomenon. This is the stall buffet. The buffet is active at high angles of attack and can be felt as vertical vibrations in the aircraft. It is a result of pressure fluctuations resulting from the flow separation [10]. The buffet acts as a natural warning cue to the pilot indicating stall entry.

When the airspeed is decreased, angle of attack needs to be increased in order to generate sufficient lift to defy the inevitable force of gravity. That means there will be air speed V_{stall}, where the angle of attack is equal to the stall angle of attack [13]. This airspeed increases with altitude due to the decrease in air density. High lift devices, such as flaps, slats and slots, can be used to keep the airflow attached for longer and therefore increase $C_{L_{\max}}$ [13]. This logically also reduces V_{stall}. Stall characteristics are therefore a function of altitude and aircraft configuration.

2.1.2. Stall Propagation

Flow separation occurs at the point where the airflow reverses direction as a result of it not being able to withstand the adverse pressure gradient. This typically does not mean that the flow is separated over the entire wing, although it may be true. For thick airfoils, the point of desperation initially occurs at the trailing edge, while the rest of the flow is still attached. When the angle of attack is increased past the stall angle of attack, the separation point moves upstream until the separation point reaches the leading edge. This is referred to as a trailing edge stall.

On the other hand, symmetric airfoil with a moderate thickness are often associated with a leading edge stall. Here the separation starts at the leading edge, resulting in a sudden and dramatic loss in lift.

So far longitudinal stall propagation has been discussed. Also in the lateral direction, stall propagation can be different. In this case the characteristics are determined by the geometry of the wing planform. The direction of propagation for 6 different wing planforms is shown in Figure 2.3.

Note that higher taper and higher sweep result in the stall initiating more towards the wingtips. On first glance this might seem beneficial, as more lift is generated at the roots of the wing. However, the ailerons are located near the tips of wing. As a result, the ailerons will be less effective during stall and lateral maneuverability will be compromised.

The lateral stalls in Figure 2.3 are valid for aircraft in wings level flight. If the aircraft flies at an angle of side slip, the backward wing will have a higher local angle of attack than the forward one. Similarly, when the aircraft is flying at a bank angle, the down wing has a higher local angle of attack than the up wing [7]. In both situations, one of the two wings will stall before the other. This can induce roll and yaw moments during stall.

2.1.3. Unsteady Stall effects

The lift curve shown in Figure 2.1 is the static stall curve and does not cover unsteady flow effects. The stall angle of attack of this curve is referred to as the static stall angle of attack. Unsteady stall comprises

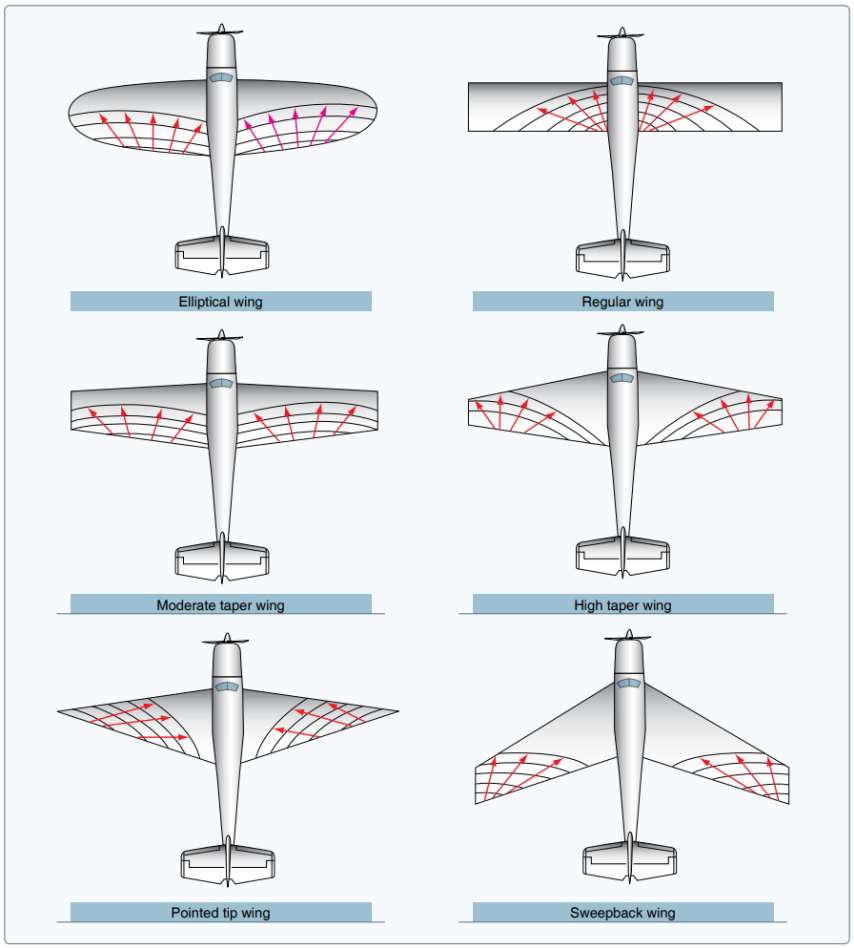


Figure 2.3: Lateral stall propagation for a number of different wing planforms, taken from [16].

of two distinct flow phenomena . One is of quasi-steady nature, and one of transient. The first effect is associated with dynamic delay of flow separation and the second with effects due to the moving of the separation point from trailing to leading edge [17]. The quasi-steady effect is caused by time lags and boundary layer improvement effects. On the one hand, the time lag simply leads to the stall occurring at a higher angle of attack, but not a higher maximum lift. There are three types of time lags: circulation lag, convective viscous flow time lag, and moving separation point effect. On the other hand, the boundary layer improvement effect bring the stall angle of attack up, but also the associated maximum lift. The stall angle of attack and maximum lift increment is proportional to the angle of attack rate. This means that when the stall is entered at a higher angle of attack rate, also a higher maximum lift is achieved.

Once stall has finally been entered, the transient effects become apparent. The first of two transient effects is a moving-wall effect. That means, some additional relative velocity is present at the wing surface due to the separation of the boundary layer. The boundary layer sees a downstream moving surface at the static point of separation, pushing the separation point back. As a result, the separation point will lag its static separation point [17]. A second transient effect only happens when the angle of attack is increased much past the point of stall. Then a vortex that has built up at the leading edge “spills” and moves towards the trailing edge, giving rise to additional normal force and moment peaks [17].

Lastly, one particularly curious stall phenomenon can be observed when the angle of attack is brought back down to below the stall angle of attack. It turns out that the observed aerodynamic forces observed in the stall regime are different during the up- and downstroke. This means that the aircraft has two distinct aerodynamic forces associated to a single angle of attack. Which of the two it is at a specific instance, depends on whether the flow was previously attached or separated. Here the motion pre-history of the aircraft enters the stall description. This phenomenon is referred to as stall hysteresis [15]. Also, the angle

of attack of reattachment can be much lower than the critical angle of attack. The forces during up- and downstroke together form a hysteresis loop. As described in this subsection, the angle of attack where separation is initiated, is dependent on the angle of attack rate, due to the boundary layer improvement effect. The boundary layer improvement effect is a reversible process [17]. meaning a higher angle of attack rate during the downstroke will also lead to a lower reattachment angle of attack and thus a larger hysteresis loop.

2.2. Stall Modeling Regulations and Requirements

The aerodynamics of stall is highly non-linear and unpredictable. This makes it hard to control for pilots. Hazardous situations can arise if the stall is not timely recognized and recovered, even fatal ones. In fact, the number one cause of fatal accidents in commercial aviation is in flight Loss of Control (LOC-I) [2]. The main contributor to these accidents is the inability to recover a stall. This has led to a call for improved Upset Prevention and Recovery Training (UPRT) for pilots in civil aviation.

Recognizing the urgency of an appropriate response, flight simulator-based stall prevention and recovery training became obligatory from 2019 onward in the US by the Federal Aviation Administration (FAA) [4]. Similarly the European Union introduced legislation on UPRT in flight simulators [5].

Training pilots in Flight Simulator Training Devices (FSTD) instead of in aircraft makes training more safe, cost-efficient and scalable. UPRT in FSTD requires accurate stall models. If the model is faulty, the training might be ineffective. The International Committee for Aviation Training in Extended Envelopes formulated a list of characteristics to be included in the stall model to ensure a satisfactory model fidelity [18]. The list is as follows:

- Degradation of static and dynamic lateral-directional stability.
- Degradation of control surface effectiveness.
- Uncommanded lateral-directional response.
- Apparent randomness or non-repeatability.
- Changes in pitch stability.
- Mach effects.
- Buffeting.
- Unsteady effects.

Naturally, including all these characteristics does not automatically yield a good stall model. Each point also needs to be accurately modeled and integrated, which is far from a trivial task. However, this list gives a good baseline of focus points to consider while developing a stall model.

2.3. Kirchhoff's Flow Separation Theory

The theoretical modelling of unsteady aerodynamics during high lift maneuvers can be done using Kirchoff's theory of flow separation. This theory describes non-linear effects on the lift curve by incorporating an additional variable X that varies between zero and one. This unit-less variable describes the point of separation on the wing. If the flow is fully attached, X is one. If it is fully detached, X is zero. The relationship according to Kirchoff's theory between lift on the one hand, and angle of attack and flow separation on the other, is given by Equation 2.1.

$$C_L = C_{L_\alpha} \left(\frac{1 + \sqrt{X}}{2} \right)^2 \alpha \tag{2.1}$$

The steady flow separation X_0 can be found experimentally as a function of solely α. Alternatively, an analytical approximation for the steady separation that is especially useful for identification from flight test data is given by Equation 2.2 [15].

$$X_0 = \frac{1}{2} \left\{ 1 - \tanh \left(a_1 \left(\alpha - \alpha^* \right) \right) \right\} \tag{2.2}$$

a_1 and α^* are some parameters that affect the behavior of the separation model, and will be explained later in this section. Unsteady aerodynamics such as hysteresis and transient effects can not be modeled if X_0 is used in Equation 2.1. An extension of Equation 2.2 is needed to describe unsteady separation. In order to include hysteresis and transient effects, an ODE is to be solved that governs X. This ODE introduces two additional stall parameters τ_1 and τ_2, and is given by Equation 2.3.

$$\tau_1 \frac{\mathrm{d}X}{\mathrm{d}t} + X = X_0 \left(\alpha - \tau_2 \dot{\alpha}\right) \tag{2.3}$$

The unsteady effects are introduced when $\frac{\mathrm{d}X}{\mathrm{d}t}$ and $\dot{\alpha}$ are non-zero. The four parameters a_1, α^*, τ_1 and τ_2 are vital for describing the steady and unsteady process of flow separation. The collective name for these parameters is the 'X-parameters'.

Each of the X-parameters has a physical meaning and affects the stall in a different way. The way they influence the lift curve and the separation as a function of α is visualized by Figure 2.4. The influence of each of the parameters can be explained as follows [15]:

- a_1 [-]: the abruptness parameter. The higher this parameter, the less gradual the stall will be. An abrupt stall will have attached flow up to high angle attack values, after which the flow conditions suddenly very quickly change to a state of full separation. As a result of the flow staying attached for longer, higher C_L values can be achieved.
- α^* [rad]: The separation delay parameter. This parameter sets the stall angle of attack. It is the angle of attack at which the flow is separated at half the chord. This again means that a higher parameter means attached flow for longer, and thus higher C_L values.
- τ_1 [s]: transient effect parameter. τ_1 is a time constant. This parameter influences the slope of X and this way models a lag in the separation process.
- τ_2 [s]: hysteresis effect parameter. τ_2 is also a time constant. It determines at what α the flow starts to detach, but also attaches back to the wing.

Using Equation 2.1 an equation for the roll moment coefficient as a function of flow separation can be obtained too. This provides a relationship between flow separation and asymmetric aerodynamics. During asymmetric stall the lift between wings is different, leading to a roll moment. As a result of the wings not stalling at the same time, this can be quit aggressive. De Fuijk et al. [9] modeled the additional roll moment due to the stall using Equation 2.4, where only the effect of the main wing was taken into account and the tail neglected.

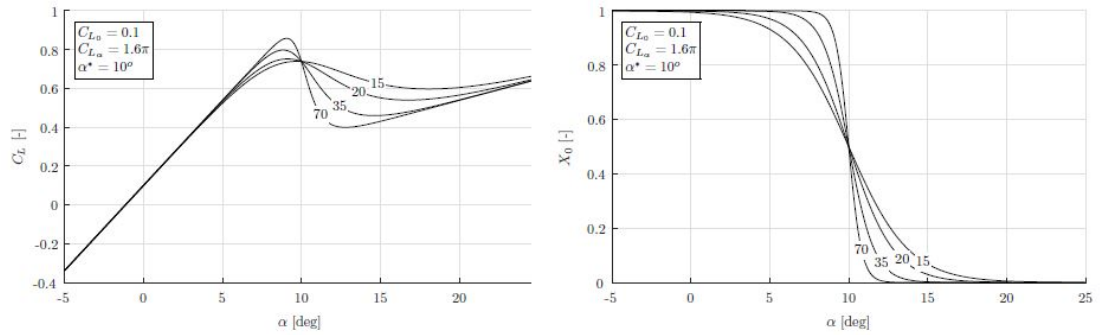
$$\Delta \hat{C}_l = \left[(C_{L_\alpha})_L \left(\frac{1+\sqrt{X_L}}{2} \right)^2 \alpha_L - (C_{L_\alpha})_R \left(\frac{1+\sqrt{X_R}}{2} \right)^2 \alpha_R \right] \frac{y_w}{b} \tag{2.4}$$

Equation 2.4 basically multiplies the lift differential between the wings with the moment arm y_w. A more simple alternative to Equation 2.4 does not take into account the local angles of attack of both wings, but merely the degree of separation at both wings. This is Equation 2.5.

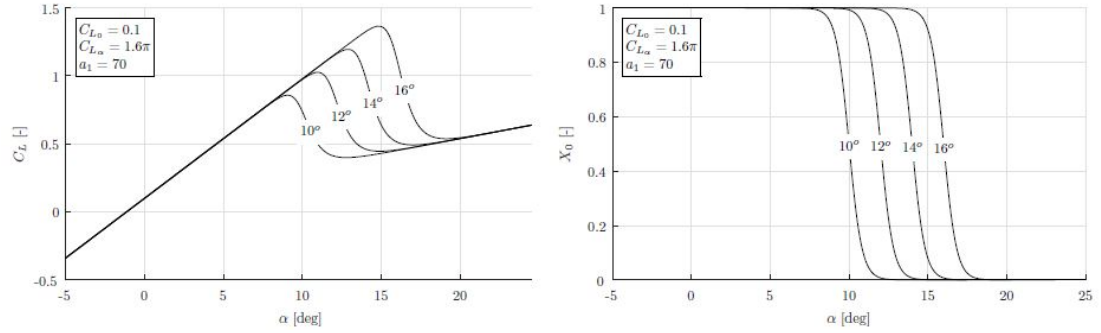
$$\begin{aligned} \Delta \hat{C}_l &= C_{l_{\Delta X}} \Delta X \frac{y_w}{b} \\ \text{with } \Delta X &= X_L - X_R \end{aligned} \tag{2.5}$$

2.4. X-Parameter Estimation

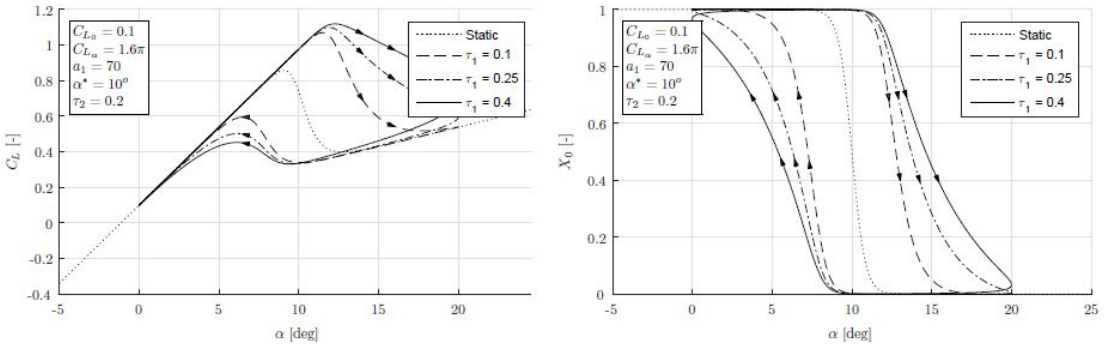
The variable X can not be measured directly. Instead X can be computed as a function of time by solving the ODE in Equation 2.3. values for the X-parameters are required knowledge to solve this ODE. The X-parameters can each be determined from experimental data [15]. This is typically a nonlinear optimization. In this section symmetric and asymmetric X-parameter estimation methods are discussed (Section 2.4.1 and 2.4.2), as well as techniques to determine what data to use for the estimation (Section 2.4.3).



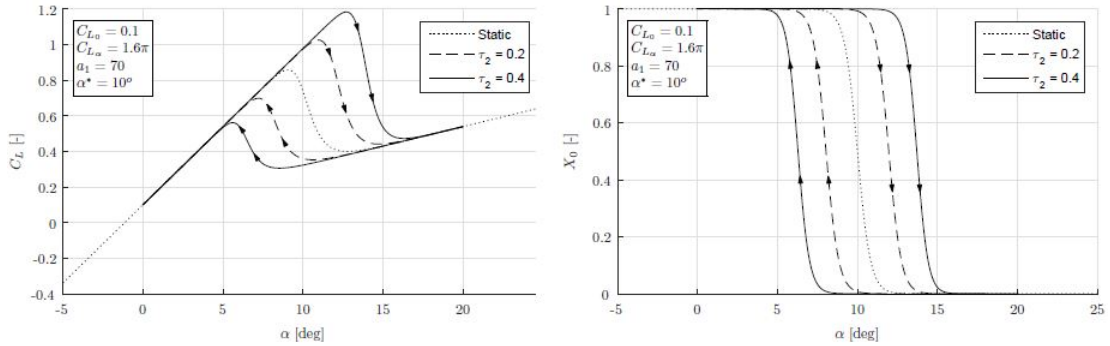
(a) Effect a_1



(b) Effect α^*



(c) Effect τ_1



(d) Effect τ_2

Figure 2.4: Effect of X-parameters on lift-curve (left column) and the point of separation X (right column), taken from [7].

2.4.1. Lift-Based Kirchoff Method

Van Ingen et al. estimated the X-parameters by solving a non-linear optimization problem [8]. They leveraged the dependency of Kirchoff's C_L on X, and thus the X-parameters, to obtain a suitable cost function. The cost function is given by Equation 2.6.

$$J(\theta, x) = \frac{1}{N}\left(C_L - \hat{C}_L(\theta, x)\right)^T \left(C_L - \hat{C}_L(\theta, x)\right) \tag{2.6}$$

θ is the vector of to be estimated parameters and x the vector of aircraft states. $J(\theta, x)$ describes the squared difference between the lift coefficient C_L as observed during a test flight and a lift coefficient estimate $\hat{C_L}$ obtained from some model depending on X. The initial model structure used for $\hat{C_L}$ is given by Equation 2.7. This model structure can change during the optimization process, as will be explained shortly The parameters C_{L_0}, C_{L_α} and C_{L_q} are estimated alongside the X-parameters. Lower and upper bounds are imposed on all parameters.

$$\hat{C_L}(\theta, x) = C_{L_0} + C_{L_\alpha}\left(\frac{1+\sqrt{X}}{2}\right)^2 \alpha + C_{L_q}\frac{q\bar{c}}{V} \tag{2.7}$$

This is a nonlinear optimization problem. The matlab routine `fmincon.m` was used to solve the optimization problem using an active set algorithm. This optimization problem is not a trivial one and it was found that the outcome can vary significantly depending on the initial conditions. Therefore the algorithm was run a number of times from different initial conditions to only keep the best runs. This reduced the variance in the estimates.

Equation 2.7 was used as the initial model structure. Once the X-parameters were determined, it was reviewed whether the model structure had to be adjusted. This was done by means of Multivariate Orthogonal Function (MOF) modeling [3]. This method selects model terms after orthogonalizing all candidate model terms from some model pool. The orthogonalization helps to avoid correlations between the regressors. The benefit of this method is that regressors are selected automatically based on the identification information content in the data. This process of alternating parameter estimation and model structure selection was repeated untill the model structure did no longer change. This only took a few iterations. A model term was added to the model structure if at least half the stalls in the data set selected a parameter.

2.4.2. Roll-Based Kirchoff Method

Similarly to the Lift-based method by van Ingen et al., again the MOF algorithm was used for iteratively selecting the correct model structure.

$$J(\theta, x) = \frac{1}{N}\left(C_l - \hat{C}_l(\theta, x)\right)^T \left(C_l - \hat{C}_l(\theta, x)\right) \tag{2.8}$$

Furthermore, they used 2 candidate model structures for the roll moment coefficient C_l. The first (Model A), is given by Equation 2.9 and the second (model B) by Equation 2.10.

$$\begin{aligned} \hat{C}_l(\theta, x) &= \hat{C}_l(\theta, x)_{lin} + C_{l_{\Delta K_\alpha}} \Delta K_\alpha \frac{y_w}{b} \\ \text{and } \Delta K_\alpha &= \left(\frac{1+\sqrt{X_L}}{2}\right)^2 \alpha_L - \left(\frac{1+\sqrt{X_R}}{2}\right)^2 \alpha_R \end{aligned} \tag{2.9}$$

$$\hat{C}_l(\theta, x) = \hat{C}_l(\theta, x)_{lin} + C_{l_{\Delta X}} \Delta X \frac{y_w}{b} \quad \text{with} \quad \Delta X = (X_L - X_R) \tag{2.10}$$

Model A and B are both extensions of a linear roll-moment model structure $\hat{C}_l(\theta, x)_{lin}$. The initial model structure chosen was taken from Klein and Morelli [19] and is given by Equation 2.11.

$$\hat{C}_l(\theta, x)_{lin} = C_{l_0} + C_{l_\beta}\beta + C_{l_p}p + C_{l_r}r + C_{l_{\delta_a}}\delta_a + C_{l_{\delta_r}}\delta_r \tag{2.11}$$

Model B is significantly simpler then A, as it does not require the computation of local angles of attack of both wings. Again a-priori lower and upper bounds were imposed on the parameter vector. The problem was optimized similarly to van Ingen's approach, but with the `fmincon.m` interior point algorithm De Fuijk et al. also found that model B is superior in terms of C_l validation Mean Squared Error (MSE).

2.4.3. Optimal Data Window Kirchoff Method

Both in the works by van Ingen et al. and de Fuijk et al. flight test data was gathered and used to estimate the X-parameters. For each stall in their databases they repeated the process described above. The amount of data assigned to each maneuver was chosen arbitrarily and included pre-stall, stall and post-stall dynamics. This is not an optimal way of estimating the X-parameters, and there is a lot to gain by smartly choosing the data windows used to estimate each parameter [20].

Brill et al. developed a slice based method for the Lift-based Kirchoff method to identify the optimal window for estimating each X-parameter. They did this by comparing simulation data and real flight data. Additionally they looked at the Fisher information in the data, a metric that quantifies information content in signals [21]. For each X-parameter they recommend a certain data window to be used for the estimation. They did this for 3 inputs types: no input, 3-2-1-1 inputs and quasi-random inputs (wiggle). The test results of this method varied greatly. In general a positive result was observed, with more than half of the validation cases improving 10% to 35% in MSE of the C_L parameter. However, in one of the test cases an increase of 116% in MSE was observed. The method is thus not guaranteed to yield beneficial results, yet it provides some very useful guidelines on estimating X-parameters.

They found that it is optimal to estimate α^* from stalls with no inputs, while using all available data. Furthermore, they found that τ_1 and τ_2 estimation is best served with a short window. For τ_2 this is data very close to the stall and mostly pre-stall, where the hysteresis effect is most prominent. The wiggle input stalls were found to contain more Fisher information than 3-2-1-1 stalls. Furthermore, the τ_1 estimation was found to be very sensitive. The Fisher information is high during the stalls and disappears post, after which the τ_1 estimate quickly increases to a wrong value.

Wang et al. also provide some guidelines on estimating the X-parameters in their 2023 paper. They too applied a lift based Kirchoff method applied to stall data. They note that a_1 and α^* only describe the quasi-steady part of the stall where $\frac{\mathrm{d}X}{\mathrm{d}t} \approx 0$ and $\dot{\alpha} \approx 0$. Therefore they suggest a_1 and α^* should first be estimated from only quasi-steady stall data. Their estimates should only contain random errors, as data measurement errors and atmospheric disturbance are small in the quasi-steady region. On the other hand, τ_1 and τ_2 mainly govern the behavior during stall and recovery. In their method, they estimate these time constants after a_1 and α^* have been fixed, to eliminate the effect of them changing as well. This can help reducing parameter variance. For this second phase they use data where transient and hysteresis characteristics are clearly present. In mathematical terms, data with $\frac{\mathrm{d}X}{\mathrm{d}t} \neq 0$ and $\dot{\alpha} \neq 0$.

2.5. Kirchoff Based Stall Model Identification

In recent years significant advancements have been made with regard to stall models from flight test data based on Kirchhoff's theory of flow separation. A common factor between all these methods is the use of the separation point variable X in the aerodynamic force and moment equations. In this section an overview will follow including the most important contributions.

Van Horssen et al. presented a low- and high-frequency stall model in 2018, based on flight tests with the Cessna Citation II PH-LAB from TU Delft. This data set contained mostly quasi-static stalls with no identification inputs and no side-slip measurements. The high frequency model predicts the violent vibration felt during the buffet. It was developed by fitting stall data to a second-order-filter in the frequency domain. By passing a white noise signal through this filter, vibrations are simulated. The low frequency model uses linear polynomials for the force and moment equations. The regressor X was used to capture the non-linear dynamics in stall. For each stall in the data set a polynomial model was estimated. The final model was obtained by averaging the resulting force and moment coefficients. This low-frequency model was fit to the same data as the flight data from the buffet model, but with a low pass filter applied. A zero-phase Butterworth filter was used. This way the effects of the buffet are excluded from the force and moment model. By superimposing the two, the complete stall can be characterized.

Van Horssen et al. only developed a longitudinal stall model. They were able to capture hysteresis effects and the buffet. On the other hand, they were not able to capture the degradation in longitudinal or lateral

stability during stall. They note that the methodology used in this paper does provide the right pathway towards a full stall model and observing stability degradation. It can be achieved if the method is applied to dynamic stall data, with appropriate inputs during the stall. For lateral models side-slip measurements were found to be essential too.

Van Ingen et al. continued working on the low frequency stall model. They developed a full longitudinal and lateral aerodynamic stall model and were able to estimate longitudinal stability degradation during stall [8]. The data was again gathered with the Cessna Citation II PH-LAB. This time, wings-level and accelerated stalls were performed with quasi-random inputs during the stall. For the model structure selection of all six force and moment equations, the MOF algorithm was used. In some cases proposed candidates were ignored based on a lack of physical explanation, and in some cases model terms were added based on engineering insight and/or a beneficial effect on validation MSE. The MOF algorithm did not select pitch rate q for the longitudinal models, even though this is a common model term [19]. They blamed this on the lack of dynamic excitation outside the stall and rather small variation inside the stall, and proposed dynamic maneuvers right before the stall in future flight tests. In the lateral model no X related model term was selected. This indicates that the stall dynamics are not fully captured and hysteresis or transient effects are missed. They were able to identify elevator control effectiveness degradation in stall. The data set only contained data at an altitude near 5500 meter with clean configuration, therefore the model is not validated outside this range.

Continuing the work, de Fuijk et al. specifically focused on the asymmetric stall model [9]. They noted that the main limitation of the Kirchoff model developed thus far is that it considers X as one dimensional coordinate, making it useless for describing asymmetric flow separation. The degree of stall can be different between the wings, inducing rolling and yawing motions. They identified the separation parameter at the left and right wing, X_L and X_R respectively, by using the roll-based Kirchoff method described in Section 2.4.2. In the flight tests executed for this research with the Cessna Citation PH-LAB only accelerated stalls were performed. This way sufficient side slip and aileron variability could be expected in the data. In some of the stalls fully automated 3-2-1-1 inputs were put on the aileron. Again, the stalls were flown at an altitude around 5500 meter with clean configuration. The model obtained trumped the one by van Ingen significantly for the lateral part: an improvement in MSE of 48% on average. On the contrary, the longitudinal model deteriorated by 88%. This is why de Fuijk et al. propose to use their roll-based Kirchoff method for the lateral model and the lift based method from van Ingen. In practise this would mean 2 sets of parameters X-parameters. Furthermore, the research was able to identify a degradation in aileron effectiveness of 41%.

All the three works mentioned above only modeled effects of the main wing. Van Horssen's model can be used to simulate buffet vibrations, van Ingen's model is best for longitudinal modeling, while de Fuijk's is better for asymmetric. None of them can accurately describe the aircraft behavior outside the stall region. In Table 2.1 the variables picked for their respective models can be found. $(\alpha - 6^\circ)_+^2$ is a spline term that is non-zero and active after an angle of attack of 6 degrees. C_T is the thrust-coefficient, and δ_a, δ_e and δ_r are the aileron, elevator and rudder deflections.

Table 2.1: Variables selected for three Kirchoff based stall models [7, 8, 9].

Data set	Coefficient	van Horssen	van Ingen	de Fuijk
Longitudinal	C_L	1, $\left(\frac{1+\sqrt{X}}{2}\right)\alpha$	1, $\left(\frac{1+\sqrt{X}}{2}\right)\alpha$, $(\alpha-6^\circ)_+^2$	1, $\left(\frac{1+\sqrt{X}}{2}\right)\alpha$, $(\alpha-6^\circ)_+^2$
	C_D	1, α, δ_e, 1-X	1, α, δ_e, 1-X, C_T	1, α, δ_e, 1-X, C_T
	C_m	1, α, δ_e,(1-X)	1, α, max($\frac{1}{2}$,X)δ_e,C_T	1, α, max($\frac{1}{2}$,X)δ_e,C_T
Lateral	C_Y	-	1, β, p, r, δ_a	1, β, r, δ_a, δ_r, ΔX
	C_l	-	1, β, p, r, δ_a	1, β, r, δ_a,ΔX
	C_n	-	1, β, r, δ_r	1, β, r, δ_a, δ_r, ΔX, α

Lastly, Wang et al. presented a full flight envelope stall model. This effort aims to solve the problems with accurately modeling the steady aerodynamics using the stall model. The data used was gathered

with a civil aircraft. Also, aerodynamic coefficients for the tail were taken into account. They propose to use a pre-stall and a stall model, with different model structures. A switch is made between the models once some critical angle of attack is exceeded. This is taken as the point where $X = 0.9$. At this point the 2 models have the same aerodynamic characteristics. After this point, some correction terms, involving higher degrees of X, start to evolve as a function of α. This is done by modeling them as a univariate spline with zero order continuity. This way the 2 models are compatible at the point of switching.

Wang et al. claim that their model is able to accurately describe steady and stall longitudinal aerodynamics. However, their model structure in the steady aerodynamic regime is near identical to the models by van Ingen et al. and de Fuijk et al.. They found that their models were not suitable for those flight conditions. Their approach of adding higher order correction terms after some critical angle of attack is promising, however.

2.6. Local Versus Global Stall Model

A global model can instantaneously predict system output and spans the entire feasible domain of this system [11]. A local model on the other hand is only valid on a subset of that domain. Local models can be very useful in predicting system output in their subdomains, but there is no guarantee of them being accurate outside this region.

The Kirchoff stall models discussed in Section 2.5 [7, 8, 9] are local models. The flight test experiments used for these models only included dynamic excitation in the stall regime. Consequently, the model can not be used to describe nominal flight [7]. Yet, it is very important to accurately cover the full flight envelope. This way, a pilot undergoing training in a FSTD will be able to feel the transition from steady to unsteady aerodynamics. Furthermore, the flight tests used by the above were all conducted at approximately the same altitude and clean configuration. This means that those models are not only local in the sense that they can not be used for nominal flight, they are also local in the sense that they can only be used for stalls with specific flight conditions.

To make the model global such that the full flight envelope is encompassed, [7] proposes to use two sigmoid blending functions. These scale two models, a steady and unsteady one, as a function of angle of attack. Unfortunately, this approach has not lead to satisfactory results. Morelli et al. identify a key requirement for obtaining a global aerodynamic model [3]. During the flight tests, it is vital to gather dynamic data over a large range of the independent variables, such as angle of attack, pitch rate and control surface deflections. To achieve this, van Ingen et al. proposed to include a dynamic maneuver before the stall maneuver [8]. Then, a model such as multivariate simplex B-splines can be used to construct a valid model over the full-flight envelope [7].

2.7. Conclusions

If an aircraft enters stall, the air flow separates from the wing. Stall is a highly nonlinear, time-varying and unpredictable aerodynamic phenomenon. During the stall the aerodynamic forces change drastically. Most noticeably, the lift is decreased and the drag increased. In the stall the aerodynamic forces can take different values based on whether the flow was previously attached or detached. Nominal aircraft models do not suffice in the stall regime and nonlinear models are required. A number of such models have been developed based on Kirchoff's theory of flow separation. These models use the flow separation point variable X as a regressor in polynomial models. X can not be directly measured but needs to be computed from a non-linear model that introduces four additional parameters, known as the X-parameters. The current state of the art is made up of a buffet model by van Horssen et al. a longitudinal model by van Ingen et al. and a lateral model by de Fuijk et al.. These models are so called local models and are only valid in the stall regime. Wang et al. presented a global longitudinal Kirchoff stall model in 2023 by combining a low and high angle of attack model.

3

Data Acquisition and Processing

A high accuracy flight-test data model can only be obtained from high-quality measurements from a real aircraft. This chapter describes how a stall data set can be obtained using the Cessna Citation II PH-LAB. Section 3.1 talks about this aircraft and its specifications, Section 3.2 about flight-test experiments suitable for collecting stall data and Section 3.3 about the current stall database gathered with the PH-LAB. Furthermore, Section 3.4 and Section 3.5 cover Flight Path Reconstruction (FPR) methods to clean noisy data. Lastly, a conclusion on the findings in this chapter can be found in Section 3.6.

3.1. Laboratory Aircraft and Instrumentation

Flight data is gathered with the Cessna Citation II PH-LAB. This former business jet is operated by the Delft University of Technology and the Netherlands Aerospace Centre (NLR). It has been fully repurposed to serve as a laboratory aircraft. Schematic drawings and an overview of its dimensions and mass properties are shown in Figure 3.1 and Table 3.1, respectively. It is equipped with an advanced Flight-Test Instrumentation System (FTIS), that logs the information measured by sensors in flight. These sensors include, amongst others, an air data boom on the nose of the aircraft, that can accurately measure the angle of attack and side slip angles. The available sensors and their measured variables relevant for aircraft system identification are listed in Table 3.2. Furthermore, the PH-LAB features a fly-by-wire system developed by the university, allowing fully automated inputs.

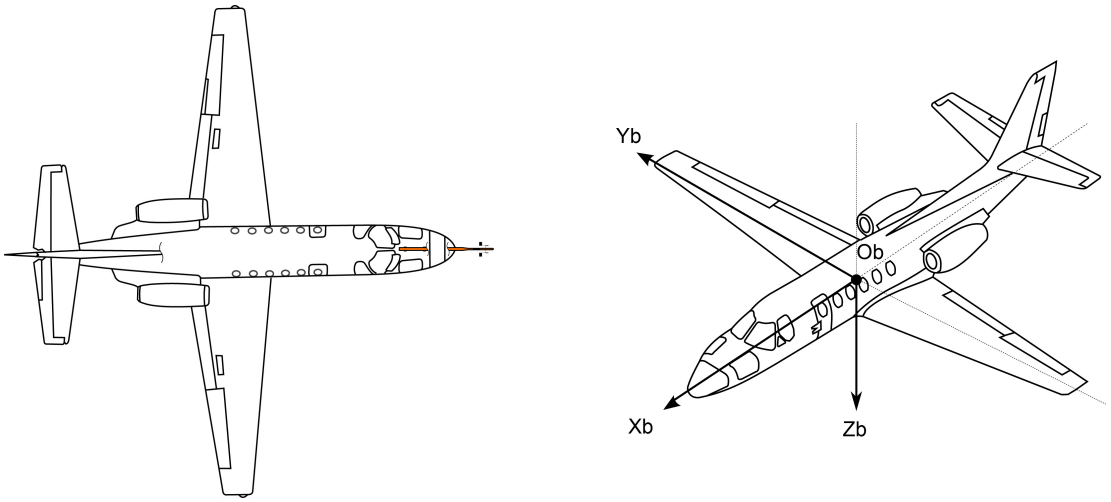


Figure 3.1: schematic top view (left) and diagonal view (right) of the Cessna Citation II PH-LAB operated by Delft university of Technology and NLR.

3.2. Experiment Design

In recent years, several flight tests have been carried out with the PH-LAB for stall system identification. The results and also flaws in the resulting models have lead to a number of recommendations for future

Table 3.1: Cessna Citation II PH-LAB Dimensions and mass properties (BEW).

	Parameter	Value
Dimensions	S	30 m^2
	$\bar{b}$	15.9 m
	$\bar{c}$	2.09 m
Mass & Inertia	m	4,157 kg
	I_{xx}	12,392 kg $\cdot$ m^2
	I_{yy}	31,501 kg $\cdot$ m^2
	I_{zz}	41,908 kg $\cdot$ m^2
	I_{xz}	2,252.2 kg $\cdot$ m^2

Table 3.2: Flight test equipment onboard the Cessna Citation II PH-LAB, and the variables measured relevant for aerodynamic system identification.

Name	Explanation	Measures	Variables	Units
GPS	Global positioning system	Position in F_E	x_e, y_e, z_e	m
		Velocity in F_E	$\dot{x}_e$, $\dot{y}_e$, $\dot{z}_e$	m/s
DADC	Digital air data computer	Total airspeed	V_{TAS}	m/s
AHRS	Attitude and heading reference system	Aircraft attitude	ϕ, θ, ψ	rad
		Body rotation rates	p, q, r	rad/s
		Body specific forces	A_x, A_y, A_z	m/s^2
Synchro	Angle measurements	Control surface deflection	δ_a, δ_e, δ_r	rad
Boom	Air data boom	Air incidence angle	α, β	rad

flight tests. First of all, van Horssen et al. found that side-slip measurements are an absolute necessity to be able to estimate a lateral model [7]. They also found the need for dynamic stall maneuvers around the critical angle of attack for good fitting. Furthermore, van Ingen et al. found that the pitch rate q was not suitable for model identification due to a lack of excitation [8]. He proposed a dynamic maneuver such as a doublet in a flight condition well before the stall. Also van Horssen et al. recommend this practise, to reduce correlation between the longitudinal biases (C_{L_0}, C_{D_0}, C_{m_0}) and α derivatives (C_{L_α}, C_{D_α}, C_{m_α}).

Moreover, some recommendations for stall flight tests are provided by the FAA [22]. They recommend to fly two types of maneuvers: wings-level and accelerated stalls (1.1g or 1.3g), and to fly both maneuvers in different flight conditions. Excitation in all three axes is required, especially during stall approach and stall itself. The same goes for high quality measurements, sufficient repetitions, and a broad coverage of the flight envelope. It should be aimed for to reach an as high as possible angle of attack.

The stall data used by van Ingen and de Fuijk et al. include two types of dynamic excitation. These inputs were placed on the aileron and elevator channels. The rudder channel was unused due to safety issues. The inputs types are explained below.

- **Quasi-Random (QR) disturbance inputs** Inspired by the work of [3], this input type requires the pilots to give fast random inputs to the control column. The pilots should make a conscious effort to give inputs as randomly as possible. In the ideal case, the inputs signal is white-noise, meaning all frequencies are excited during the maneuver. In reality, this is not possible, and the inputs are thus quasi-random.
- **3-2-1-1 inputs** This is a multi-step input. The numbers in the name correspond to the number of seconds the control column is held positively or negatively, before moving it to the opposite direction. With the fly-by-wire system installed on the PH-LAB, this maneuver can be flown automatically. This provided a high degree of consistency between different stalls during a flight test. Also, this maneuver divides power over a wide frequency range.

The 3-2-1-1 inputs are present in the data set used by de Fuijk et al.. In the stall flight-tests the 3-2-1-1 input was never used on both the aileron and elevator control column, instead it was combined with a QR input. In Figure 3.3 examples of the 3-2-1-1 and QR inputs are shown during one stall.

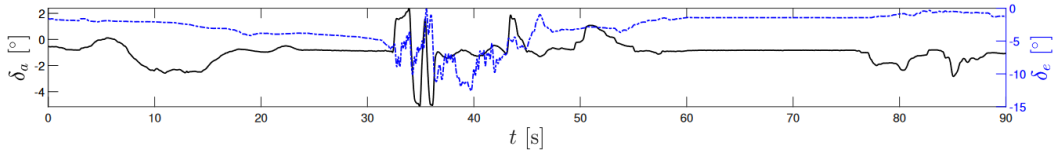


Figure 3.3: Illustrative stall identification inputs, with a 3-2-1-1 input on the aileron and QR input on the elevator, taken from data de Fuijk et al. [9].

3.3. Stall Database

The stall models developed by van Ingen et al. [8] and de Fuijk et al. [9] each used a dataset from flight tests with the PH-LAB, designed for their specific model objectives. The first two blocks in Table 3.3 provides an overview of these two data sets. The data sets are referred to as *van Ingen* and *Asym*, respectively. The *van Ingen* data set contains mostly data around FL180, and a handful of stalls at lower altitudes. Half these stalls were accelerated and half wings-level. Some of the accelerated stalls were executed at 1.3g. The *Asym* data set consists fully of accelerated stalls. The 1.3g stalls originate from the *van Ingen* data set. 10 of these stalls have the 3-2-1-1 input on the aileron. These stalls were, with one exception, all flown near FL180.

Table 3.3: Stall maneuvers in stall data sets *van Ingen*, *Asym* and *Global*.

Data set	h [m]	h [FL]	n_z	Input(δ_a | δ_e)	Reps
van Ingen	2900	95	1.0	QR|QR	2
	4300	140	1.0	QR|QR	3
	4500	145	1.0	QR|QR	1
	4700	155	1.0	QR|QR	1
	5000	165	1.0	QR|QR	1
	∼5500	∼ 180	1.0	QR|QR	9
	∼5500	∼ 180	1.1	QR|QR	10
	∼5500	∼ 180	1.3	QR|QR	7
Asym	4900	160	1.1	QR|QR	1
	∼5500	∼ 180	1.1	QR|QR	12
	∼5500	∼ 180	1.1	3211|QR	10
	∼5500	∼ 180	1.3	QR|QR	7
Global	5500	180	1.1	QR|QR	1
	4600	150	1.1	QR|QR	2
	4300	140	1.1	QR|QR	2
	4000	130	1.1	QR|QR	2
	3700	120	1.1	QR|QR	2
	3400	110	1.1	QR|QR	2
	3000	100	1.1	QR|QR	2
	2700	90	1.1	QR|QR	2

A new data set was compiled in December 2023, the *Global* data set. An overview of the stalls flown is found in the last portion of Table 3.3. The stalls in this data set were purposely flown at varying altitudes, mostly between FL150 and FL90, so that the effect of altitude on stall can be studied. One stall was recorded at FL180, to have some overlap with the *Asym* data set. All stalls are asymmetric 1.1g stalls. Each stall at FL150 and below was flown in a left- and right-hand turn. The FL180 stall only in a right-hand turn. The inputs given were QR disturbance inputs.

In addition to stalls, also some low angle of attack maneuvers were recorded in the *Global* data set. These maneuvers had a dynamic input simultaneous on the aileron and elevator, at $\sim$130 knots. Figure 3.4 and 3.5 show examples for the time traces of the aileron and elevator inputs. An overview of these maneuvers is provided in Table 3.4.

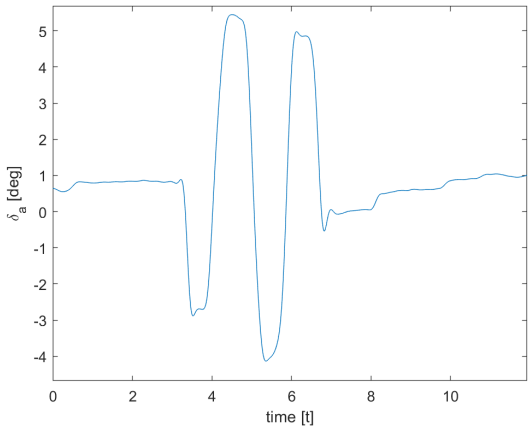


Figure 3.4: Example aileron steady identification input in *Global* data set at an airspeed of about 130 knots.

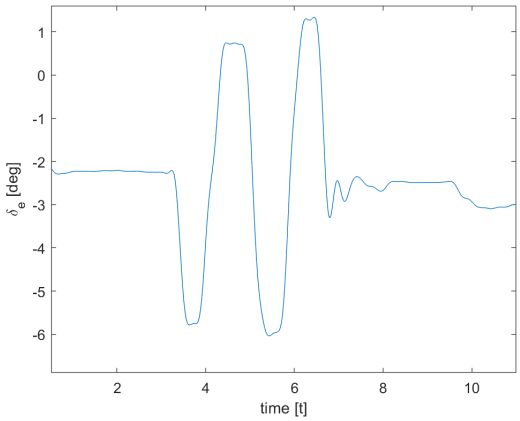


Figure 3.5: Example elevator steady identification input in *Global* data set at an airspeed of about 130 knots.

Table 3.4: Steady maneuvers recorded in the *Global* data set.

Maneuver	h [m]	h [FL]
1	5500	180
2	4600	150
3	4300	140
4	4000	130
5	3700	120
6	3400	110
7	3000	100
8	2700	90

The flight envelope (altitude h against Mach M and α against β) from the three aforementioned data set are shown in Figure 3.6-3.11. The sparsity in the data is clearly visible. In the (h, M) envelope there is no data between the stalls flown at different altitudes. In the (α, β) envelope there is no sparse regions at the edges, and very dense regions at the core.

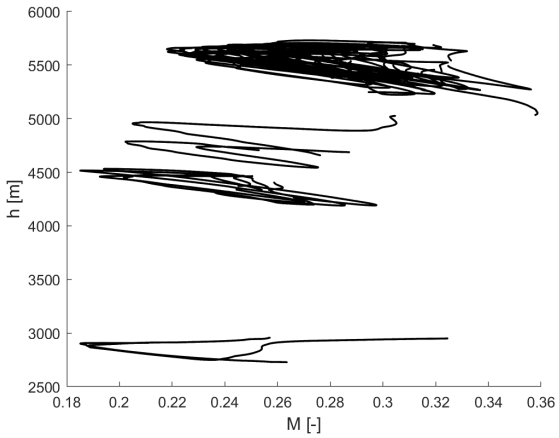


Figure 3.6: *Van Ingen* data set flight envelope (h, M).

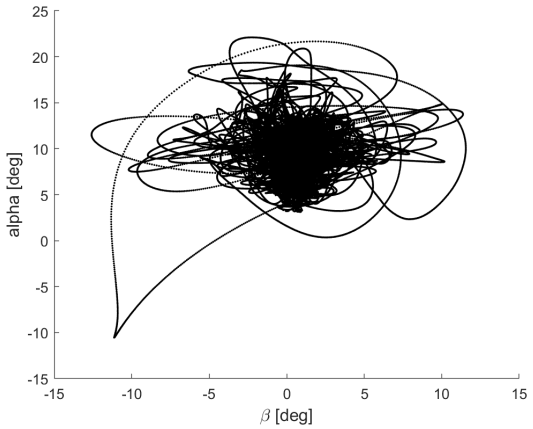


Figure 3.7: *Van Ingen* data set flight envelope (α, β).

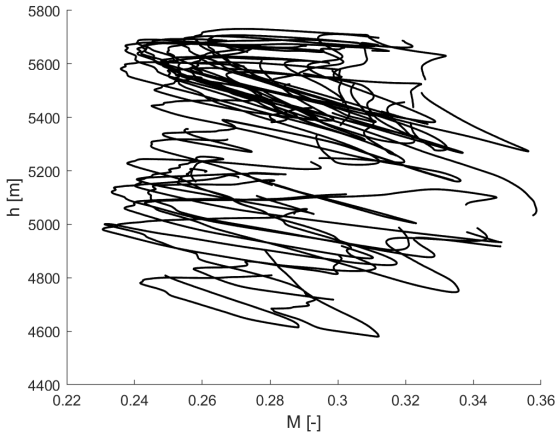


Figure 3.8: *Asym* data set flight envelope (h, M).

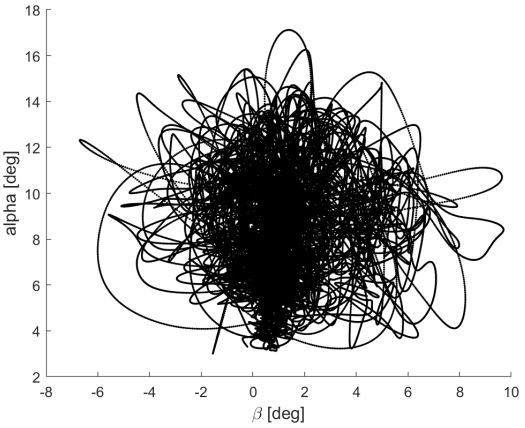


Figure 3.9: *Asym* data set flight envelope (α, β).

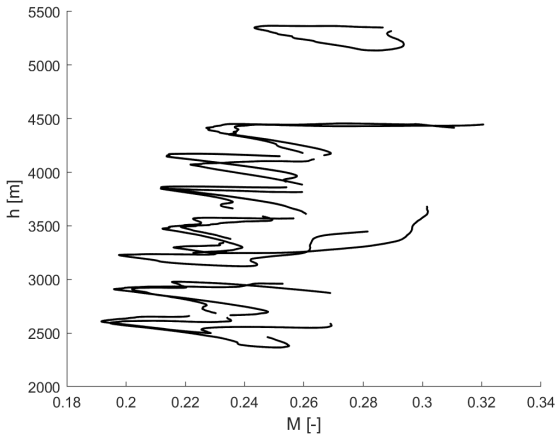


Figure 3.10: *Global* data set flight envelope (h, M).

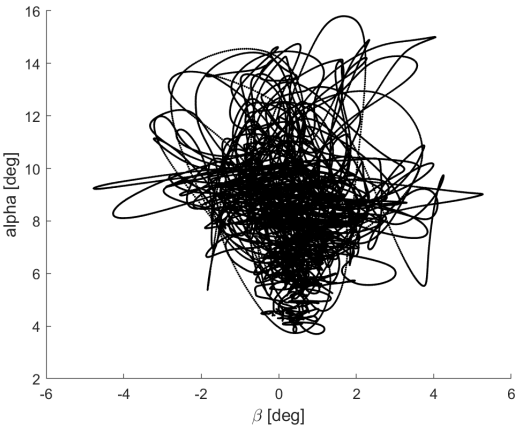


Figure 3.11: *Global* data set flight envelope (α, β).

3.4. Kalman Filter

The Kalman Filter (KF) is an important tool in the flight path reconstruction process. Kalman filters are the most common state estimation methods. Many variations of the Kalman filter exist, but the main principle is the same. It is a recursive algorithm that estimates the state of dynamic systems, such as aircraft, from noisy measurements. To do this it uses two fundamental steps: prediction and update. During the prediction step, the system's state is forecasted based on the previous state and the system's dynamics. The system's dynamics can be provided with a state space representation. During the update step, the predicted state is adjusted by taking into account new measurements, while also taking into account the reliability of these measurements. The original Kalman filter is an optimal filter [23], in the sense that it minimizes the expected mean squared error. Unfortunately, this filter is not suitable for more complex, nonlinear systems. The most common extensions for nonlinear systems are the Extended Kalman Filter (EKF), the Iterated Extended Kalman Filter (IEKF) and the Unscented Kalman Filter (UKF). The EKF uses linearization of the nonlinear system. It has the unfavourable property that convergence to the optimal solution depends on the choice of the initial conditions. The IEKF follows the same principle as the EKF but uses additional iterations to improve convergence. Lastly, the UKF is different to the other mentioned nonlinear Kalman Filters. It does not include linearization in its algorithm. Instead, it is centered around the scaled unscented transformation [24].

The UKF and IEKF are two good options for stall data state reconstruction. In previous stall model research by van Horssen et al. the UKF was selected as the most suitable state reconstruction tool [7], for the reasons that the UKF is less sensitive to initial conditions than the IEKF. Also van Ingen et al. and de Fuijk et al. succesfully used the UKF in their stall models. The UKF is not a suitable tool to remove vibrations caused by the buffet. As mentioned before, for this purpose a zero-phase Butterworth filter is an effective option.

3.5. Aerodynamic Coefficients

The aerodynamic model that is used in the various stall models developed at TU Delft use the three force and moment equations expressed in the aerodynamic frame of reference. These forces and moments can not be measured directly. In order to obtain them, first the force equations in the aircraft body frame of reference need to be calculated with Equation 3.1 using the measured body specific forces A_x, A_y and A_z and .

$$\begin{aligned} C_X &= \frac{(A_x - X_T)}{\overline{q}S} \\ C_Y &= \frac{(A_y)}{\overline{q}S} \\ C_Z &= \frac{(A_z - Z_T)}{\overline{q}S} \end{aligned} \tag{3.1}$$

The wing surface S was provided in Table 3.1. $\overline{q}$ is the dynamic pressure. X_T and Z_T are the components of the thrust force along the X and Y axis. Z_T is neglected.

Next, C_D and C_L can be calculated from the forces in the body frame and measurements of α and β with Equation 3.2. C_Z is used to complete the force coefficients in the aerodynamic frame

$$\begin{aligned} C_D &= -C_X \cos\alpha \cos\beta + C_Y \sin\beta - C_Z \sin\alpha \cos\beta \\ C_L &= -C_Z \cos\alpha + C_X \sin\alpha \end{aligned} \tag{3.2}$$

Lastly, the moment coefficients can be calculated by using the measurements of the rotational rates p, q and r. These calculations also require the rotational accelerations. The latter are not available from measurements. Instead, they are computed by differentiating the rotational rate signals. The resulting accelerations are low-pass filtered to remove any amplified noise.

$$
\begin{aligned}
C_l &= \frac{I_{xx}}{\overline{q}Sb}\left[\dot{p} - \frac{I_{xz}}{I_{xx}}\left(pq + \dot{r}\right) + \frac{I_{zz} - I_{yy}}{I_{xx}}qr\right] \\
C_m &= \frac{I_{yy}}{\overline{q}Sb}\left[\dot{q} + \frac{I_{xx} - I_{zz}}{I_{yy}}pr + \frac{I_{xz}}{I_{yy}}\left(p^2 - r^2\right)\right] + \frac{1}{\overline{q}S\overline{c}}X_T\left(z_e - z_{cg}\right) \\
C_n &= \frac{I_{zz}}{\overline{q}Sb}\left[\dot{r} - \frac{I_{xz}}{I_{zz}}\left(\dot{p} - qr\right) + \frac{I_{yy} - I_{xx}}{I_{zz}}pq\right]
\end{aligned}
\tag{3.3}
$$

The inertia of the aircraft, the average chord length and wing span b are found in Table 3.1.

The complete FPR process involves many more steps, but these have been extensively researched and validated. Therefore these steps are not included in this chapter. Instead the reader is referred to the master theses by either van Horssen or Brill [25, 20].

3.6. Conclusions

Stall data can be gathered by flying stalls in the Cessna Citation II PH-LAB. During such experiments an accurate side-slip measurement is crucial for asymmetric identification. Identification inputs such as QR disturbance inputs or 3-2-1-1 inputs need to be applied in the stall regime. For global aerodynamic modeling also a dynamic maneuver outside the stall is necessary. A zero-phase Butterworth filter needs to be applied to remove vibrations due to the buffer, and the UKF is the best option for state reconstruction.

4

Simplex B-Splines

Constructing accurate global models of complex systems characterized with severe nonlinearity and a large domain can only be done with a suitable modeling method. Multivariate simplex B-splines are function approximators that lend themselves very well for a global nonlinear system identication framework. Namely, this method has some favourable properties that other global model methods lack. In this chapter, these simplex B-splines and their use in system identification are discussed. The comparison between simplex B-splines and other global modeling methods is first made in Section 4.1, followed by an introduction to univariate B-splines in Section 4.2, the multivariate simplex B-spline theory in Section 4.3 and a variation of simplex B-splines in Section 4.4: the physical splines. Finally a conclusion is provided in Section 4.5.

4.1. Simplex B-splines as System Identification Framework

There are not many available function approximators that are suitable for global and nonlinear modeling. These available methods are neural networks, kernel methods, polynomial methods and spline models [11].

Especially neural networks are a widely applied data modeling tool. Neural networks have a very high approximation power and can in fact approximate any function [26]. However, neural networks are also black-box models lacking transparency. Furthermore, neural networks do not have local basis functions and therefore all basis functions need to be taken into account while evaluating and estimating the model. This makes efficient parameter estimation troublesome.

Kernel methods generally require a coefficient vector as large as the number of data points. This makes global model identification practically infeasible, as a large data set is typically required.

Polynomial methods are great for local modeling but not so much for global modeling. Combining local polynomial models into a global model is possible but the combination results into gaps between the different sub-models. Even though it is possible to blend them into a continuous model, doing this is impractical and requires expert knowledge.

This leaves the spline models. There is many variations on spline models, including unsuitable options. However, the simplex B-splines are especially suitable for system identification. They can be used to construct global models with an arbitrarily high approximation power. Furthermore, these splines have local basis functions and are parametric models. Consequently, parameter estimation and model evaluation is efficient and feasible for large data sets. Lastly, the parameter estimation method is rooted in linear regression.

4.2. Univariate B-splines

Univariate splines have been around for much longer than the multivariate splines. Univariate B-spline theory has been extensively studied and described, such as by de Boor and Dierckx [27, 28]. A univariate B-spline is a smooth 1-dimensional curve that consists of connected piecewise polynomial basis functions. A first order, linear B-spline basis function is shown on the left of the top axis in Figure 4.1. It is build up from two linear pieces, one between x_1 and x_2, and one between x_2 and x_3. The x are called knots, and outside the domain these knots span, the basis function is zero valued. At the knots, the linear pieces are connected. By introducing more knots, more basis functions can be added. This is done to the right

of this same depiction. To create higher order B-splines, the degree of the polynomial pieces can be increased. An example is shown in the bottom axis of Figure 4.1. On the left, a single quadratic basis function is shown, defined on four knots and consisting of 3 pieces. Now, the pieces are not only connected at the knots, they have the same first derivative as well. On the right, multiple overlapping quadratic basis functions are shown.

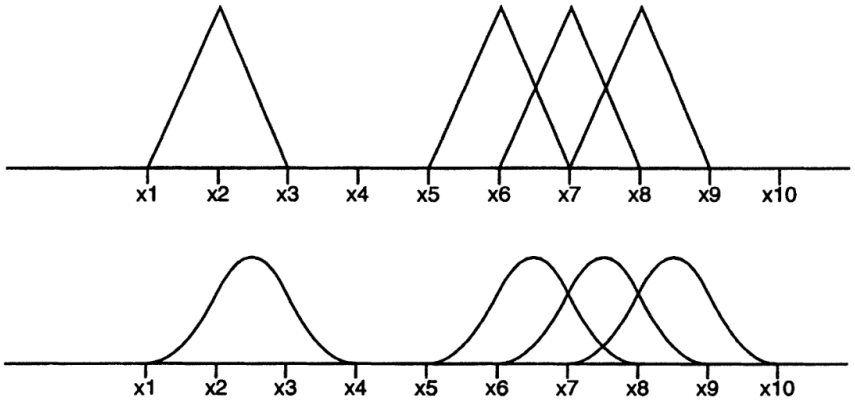


Figure 4.1: Examples of isolated univariate B-splines and overlapping ones, on the top of degree 1 and on the bottom of degree 2, taken from [29].

In general the following general properties can be derived for a B-spline of degree q:

- A degree q B-spline basis function consists of q + 1 polynomial pieces, each of degree q.
- The polynomial pieces join at q inner knots.
- The derivatives up to order q-1 are continuous at the joining points of the polynomial pieces.
- The basis functions are zero outside the domain spanned by its knots.
- If multiple B-splines basis functions are connected into one B-spline, each polynomial piece overlaps with 2q neighbours, except at the boundaries.
- at a given x, q + 1 basis functions are nonzero.

So the B-spline is made up of a series of piecewise polynomial basic functions that are structured and connected in a special way. The B-spline can easily be used in univariate parametric regression [29]. Making a linear combination of basis-functions yields a smooth curve.

Mathematically, B-splines are also attractive. De Boor introduced an algorithm that allows the computation of any degree basis-function from lower degree ones. By iterating this algorithm, eventually any B-spline can be written in terms of zero-degree basis functions. Basis function of degree zero are simply constant, making it easy to compute any degree B-spline. If a degree q basis function, active between knots x_i and x_{i+1} and evaluated at x, is denoted by $B_{i,q}(x)$, then it can be rewritten in terms of lower degree basis functions by Equation 4.1.

$$B_{i,q}(x) = \frac{x - x_i}{x_{i+q+1} - x_i} B_{i,q-1}(x) + \frac{x_{i+q} + x}{x_{i+q} - x_{i+1}} B_{i+1,q-1}(x) \tag{4.1}$$

The works by De Boor and Dierckx can be reffered to for a complete overview of univariate B-splines as functions and in a regression framework [27, 28].

4.3. Simplex B-spline Theory

The univariate B-splines are a great tool for creating simple global models. However, when modeling more complex systems where more regressands are needed to explain all the system dynamics, a multivariate theory is required. In this subsection the multivariate simplex B-spline theory is discussed, and why it is the most suitable multivariate spline theory for system identification. In Section 4.3.1 an introduction to

and comparison between multivariate spline methods is provided. In Section 4.3.2, 4.3.3 and 4.3.4 the construction of a global spline model is explained. Finally, in Section 4.3.5 some mathematical properties of simplex B-splines are listed.

4.3.1. Background

It has proven difficult to generalize the theory of univariate splines to a multivariate one. Even now, despite extensive efforts being put into this topic, the research field is still active. There are currently four spline theories that carry at least some of the desirable properties of univariate splines [11]. These are tensor product splines, thin plate splines, the polyhedral spline and the simplex spline. Elements and concepts taken from each of these spline theories will surface throughout the subsequent of this study. However, there is one spline type that will dominate and provide the main system identification framework. This is the simplex spline, as it is the only one that possesses all the necessary properties required for a successful nonlinear system identification method. These properties are listed here:

- Local Polynomial basis
- Scattered data fitting
- Efficient parameter estimation

The local polynomial basis allows local curve fitting based on the data near the point of interest. The scattered data fitting property is a prerequisite as flight test data are generally scattered. Efficient parameter estimation is needed to allow modeling of large data sets. This is especially important for global modeling of systems with a large domain.

The simplex B-spline consists of local piecewise Bernstein polynomials that are defined on a geometric structure known as a triangulation. the triangulation is built up from triangular patches called simplices. Inside the simplices a local Barycentric coordinate system is used. When compiling a spline basis the Barycentric coordinate system leads to a a sparse system of equations. This is beneficial because it provides the required property of efficient parameter estimation.

4.3.2. B-Form Polynomials

It should be clear now that simplex B-splines are piecewise polynomials. Each polynomial is defined on a simplex, together a so called simplex polynomial. A simplex is a geometric structure that provides a minimal, non-degenerate span of n-dimensional space [30]. The set of $n+1$ unique points that matches the above description is called V and the convex hull of V, being the n-simplex, is called t. V and t are defined by Equation 4.2 and Equation 4.3, respectively.

$$V := \{v_0, v_1, \ldots, v_n\} \tag{4.2}$$

$$t := \langle V \rangle \tag{4.3}$$

In Figure 4.2 examples of simplices for zero to three dimensional space are shown.

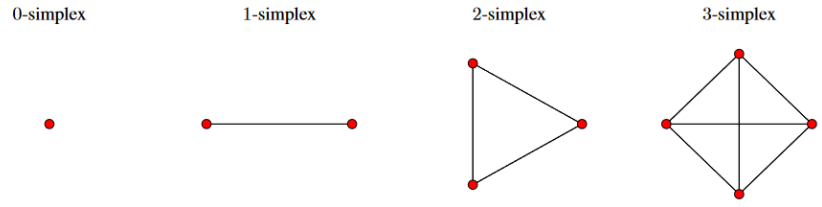


Figure 4.2: Graphical depiction of 0- to 3-dimensional simplex representations, taken from [11].

Not the Cartesian coordinate system, but another is used to define points in- and outside the simplex. This is the Barycentric coordinate system. The Barycentric coordinates of point $x = (x_1, x_2, \cdots, x_n)$ can

be seen as the weights put on the vertices of t required to put the center of gravity in the point x. The Barycentric coordinates can be denoted by $b(x) = (b_0, b_1, \cdots, b_n)$. The transformation from Cartesian to Barycentric coordinates is made using Equation 4.4.

$$x = \sum_{i=0}^{n} b_i v_{p_i}, \quad \sum_{i=0}^{n} b_i = 1 \tag{4.4}$$

The reason for using this unconventional coordinate system is that polynomials of any degree d can be elegantly expressed in Barycentric coordinates with the multinomial theorem. A polynomial of degree d in Barycentric coordinates can be expressed using Equation 4.5.

$$(b_0 + b_1 + \cdots + b_n)^d = \sum_{|\kappa|=d} \frac{d!}{\kappa_0!\kappa_1! \ldots \kappa_n!} b_0^{\kappa_0} b_1^{\kappa_1} \ldots b_n^{\kappa_n} = \sum_{|\kappa|=d} B_\kappa^d(\boldsymbol{b}) \tag{4.5}$$

κ in Equation 4.5 is a multi-index B_κ^d a Bernstein basis polynomial. κ is defined by Equation 4.6 and has the properties shown in Equation 4.7.

$$\kappa := (\kappa_0, \kappa_1, \ldots, \kappa_n) \in \mathbb{N}^{n+1} \tag{4.6}$$

$$\begin{aligned} \kappa! &= \kappa_0!\kappa_1! \ldots \kappa_n! \\ |\kappa| &= \kappa_0 + \kappa_1 + \cdots + \kappa_n \end{aligned} \tag{4.7}$$

The summation in Equation 4.5 actually turns out to always sum to unity, as shown by de Boor in 1987 [27]. In other words, the Bernstein polynomials provide a stable basis. As a result of the stable basis, any polynomial of degree d can be written as a linear combination of Bernstein polynomials. This polynomial representation is called the B-form and it is shown in Equation 4.8.

$$p(\boldsymbol{b}) = \sum_{|\kappa|=d} c_\kappa B_\kappa^d(\boldsymbol{b}) \tag{4.8}$$

The c_κ are called the B-coefficients. The B-form can also be molded into vector form. This vector form takes the form of Equation 4.9.

$$p(\boldsymbol{b}) = \boldsymbol{B}_t^d(\boldsymbol{b})\, \boldsymbol{c}_t \tag{4.9}$$

In Equation 4.9 $\boldsymbol{B}_t^d(\boldsymbol{b})$ is the vector of Bernstein basis polynomials for simplex t. This vector acts as the foundation of the spline basis that will be used in a regression framework. $\boldsymbol{c}_t$ is the corresponding vector of B-coefficients that scale the basis polynomials. They can be expanded to Equation 4.10 and 4.11, respectively.

$$\boldsymbol{B}_t^d(\boldsymbol{b}) = \begin{bmatrix} B_{(d,0,\ldots,0)}^d(\boldsymbol{b}) & B_{(d-1,1,\ldots,0)}^d(\boldsymbol{b}) & \cdots & B_{(0,0,\ldots,d)}^d(\boldsymbol{b}) \end{bmatrix} \tag{4.10}$$

$$\boldsymbol{c}_t = \begin{bmatrix} c_{(d,0,\ldots,0)} & c_{(d-1,1,\ldots,0)} & \cdots & c_{(0,0,\ldots,d)} \end{bmatrix}^T \tag{4.11}$$

Note that the subscripts in Equation 4.10 and Equation 4.11 are the multi-index κ written out in full, ordered in lexicographical order.

4.3.3. Building a Global Spline Model

The concept of the simplex was explained in the the previous (Section 4.3.2), as well as the mathematical definition (Equation 4.2 and 4.3). It is even more interesting to look at a set of simplices. A triangulation is

formed by connecting simplices together in a non-overlapping manner. This way the entire domain can be partitioned into smaller portions. Only the data that falls inside a simplex is used to fit that simplex polynomial and thus that piece of the global spline, hence the local polynomial basis property. A triangulation contains T simplices and is defined by Equation 4.12. In this equation $\tilde{t}$ is an edge simplex with a dimension no higher than $n-1$.

$$\mathcal{T} := \bigcup_{i=1}^{T} t_i, \quad t_i \cap t_j \in \{\varnothing, \tilde{t}\}, \quad \forall t_i, t_j \in \mathcal{T} \tag{4.12}$$

Each simplex has its B-coefficients spatially organised in a special structure, called the B-net. This property unique to the simplex B-splines allows the curve to be affected even locally within the simplex by varying the value of a specific B-coefficient. In Figure 4.3 an example of a B-net for a fourth degree spline is shown. This spline has 15 B-coefficients but the number of parameters $\hat{d}$ can be found for any spline degree d with Equation 4.13. $\hat{d}$ is also the size of the vectors in Equation 4.10 and Equation 4.11.

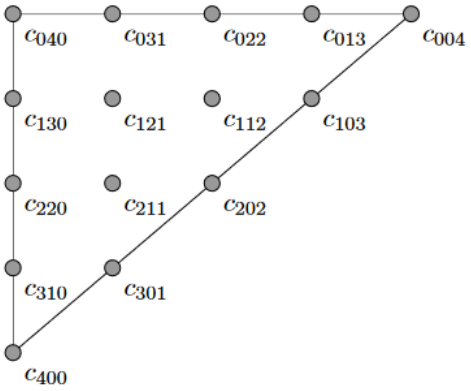


Figure 4.3: Degree 4 spline B-net, taken from [11].

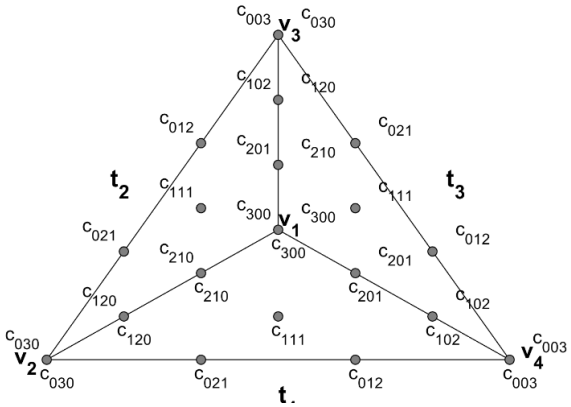


Figure 4.4: Degree 4 spline triangulation with 3 simplices, taken from [11].

$$\hat{d} = \binom{d+n}{n} = \frac{(d+n)!}{n!d!} \tag{4.13}$$

In Figure 4.4 a triangulation of 3 simplices each with their B-net is shown. Indexing the B-coefficients in a triangulation is quit a religious task and should be done with the necessary care. The indexing is subject to some rules, as explained in [11]. The indices will determine the order the Bernstein polynomials and B-coefficients appear in the vector B-form (Equation 4.10 and 4.11).

Obtaining an appropriate triangulation is not easy. There are four simplex metrics that need to have a sufficient value to prevent numerical issues [11]. These metrics boil down to some necessary conditions that are to be met for a good triangulation. It is important that the simplices are not too slender, have sufficient volume and that there is at least $\hat{d}$ data points inside the simplex. Bad simplices that are not meeting the conditions are called sliver simplices.

Delaunay triangulations are the most common triangulation method. However, in dimensions higher than two, the Delaunay tiangulation leads to sliver simplices [11]. Kuhn triangulations are an alternative that is especially suitable for aircraft system identification [31, 32]. The Kuhn triangulization methhod is a recursive algorithm that is initialized with an n-dimensional hypercube. Then the hypercube is subdivided in n-simplices, where their vertices are determined recursively. Then, simplices can be removed that are outside the feasible domain, or do not contain data, to obtain a non-rectangular triangulation. Figure 4.5 displays a Kuhn triangulation in $\mathbb{R}^3$.

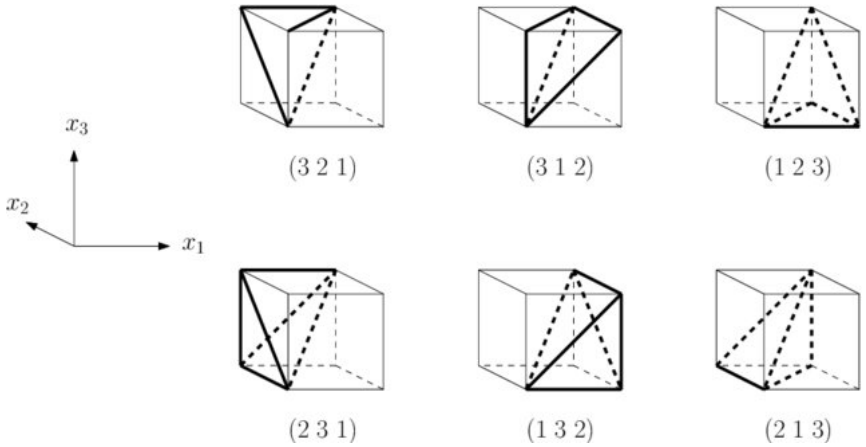


Figure 4.5: Kuhn triangulation in $\mathbb{R}^3$, taken from [32].

Once the triangulation is set, a matrix made up of B-form regressors (Equation 4.10) can be constructed for each simplex, with one row for each data point falling inside the simplex. This gives Equation 4.14 for simplex j, taken form [33].

$$\boldsymbol{B}^d_{t_j}(b(\boldsymbol{x}_{t_j})) = \begin{bmatrix} B^d_{d,0,\ldots,0}(b(\boldsymbol{x}_{1_{t_j}})) & B^d_{d,0,\ldots,0}(b(\boldsymbol{x}_{1_{t_j}})) & \cdots & B^d_{d,0,\ldots,0}(b(\boldsymbol{x}_{1_{t_j}})) \\ \vdots & \vdots & & \vdots \\ B^d_{d,0,\ldots,0}(b(\boldsymbol{x}_{M_{t_j}})) & B^d_{d,0,\ldots,0}(b(\boldsymbol{x}_{M_{t_j}})) & \cdots & B^d_{d,0,\ldots,0}(b(\boldsymbol{x}_{M_{t_j}})) \end{bmatrix} \tag{4.14}$$

By collecting these matrices for each simplex in a single global matrix, $\boldsymbol{X}$ is constructed. $\boldsymbol{X}$ has a high degree of sparseness [34, 10], as the $\boldsymbol{B}^d_{t_j}(b(\boldsymbol{x}))$ are zero when $\boldsymbol{x}$ does not fall inside t_j. In fact, $\boldsymbol{X}$ is a block diagonal matrix, where the off-diagonal blocks contain only zeros. Also a global B-coefficient vector $\boldsymbol{c}$ can be constructed with B-coefficient vectors corresponding to the blocks. $\boldsymbol{X}$ and $\boldsymbol{c}$ are defined by Equation 4.15.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{B}^d_{t_1}(b(\boldsymbol{x}_{t_1})) & 0 & \cdots & 0 \\ 0 & \boldsymbol{B}^d_{t_2}(b(\boldsymbol{x}_{t_2})) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{B}^d_{t_T}(b(\boldsymbol{x}_{t_T})) \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_{t_1} \\ \boldsymbol{c}_{t_2} \\ \vdots \\ \boldsymbol{c}_{t_T} \end{bmatrix} \tag{4.15}$$

adjacent simplices share the B-coefficient that lie on their shared facet. Here the concept of continuity comes into play. By enforcing these shared B-coefficients to becoming same valued, zeroth order continuity is achieved. By involving the next layers of coefficients higher orders can be achieved as well. The conditions to be imposed for a certain continuity order m can be captured by Equation 4.16 [30].

$$c^{t_1}_{(\kappa_0,\ldots,\kappa_{n-1},m)} = \sum_{|\gamma|=m} c^{t_2}_{(\kappa_0,\ldots,\kappa_{n-1},0)} B^m_\gamma(v_*), \quad 0 \leq m \leq r \tag{4.16}$$

In Equation 4.16, γ is a multi-index similar but independent of κ. v_* is the out-of-edge vertex of simplex t_1. For all neighbouring simplices these continuity equations need to be set up for orders 0 to r to achieve C^r continuity, for a total of R continuity conditions. All continuity conditions should be assembled in smoothness matrix H such that Equation 4.17 holds.

$$\boldsymbol{H}\boldsymbol{c} = 0, \quad \boldsymbol{H} \in \mathbb{R}^{R \times T\hat{d}} \tag{4.17}$$

4.3.4. B-Coefficient Estimation

The B-coefficients can be estimated efficiently using linear regression. The regression problem is formulated as Equation 4.18 [30].

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{c} + \boldsymbol{r} \in \mathbb{R}^N \tag{4.18}$$

The $\boldsymbol{Y}$ in Equation 4.18 is the observations vector, $\boldsymbol{X}$ and $\boldsymbol{c}$ are the global spline matrix and vector from Equation 4.15. Due to the sparseness of $\boldsymbol{X}$, efficient solvers exist for solving the regression problem, allowing the estimation of many coefficients. Furthermore, $\boldsymbol{r}$ is the residual vector that captures unmodeled variation in the observations. An optimal estimate $\hat{\boldsymbol{c}}$ in terms of squared residuals is found by solving the ordinary least squares (OLS) problem in Equation 4.19. This solution can be easily extended to generalized or weighted least squares if need be.

$$\hat{\boldsymbol{c}} = \underset{\boldsymbol{c}}{\operatorname{argmin}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{c}\|_2^2 \quad \text{subject to } \boldsymbol{G}\boldsymbol{c} = \boldsymbol{g} \tag{4.19}$$

$\boldsymbol{G}$ and $\boldsymbol{g}$ are given by Equation 4.20. These constraints impose equality constraints that enforce smoothness with $\boldsymbol{H}$ (Equation 4.17), also a sparse matrix, and potentially additional non-zero constraints through constraint matrix and vector $\boldsymbol{W}$ and $\boldsymbol{w}$. The latter constraint type also introduces the possibility of differential constraints [34].

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{W} \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} 0 \\ \boldsymbol{w} \end{bmatrix} \tag{4.20}$$

The equality constraints make it no trivial task to calculate $\hat{\boldsymbol{c}}$. De visser, Chu and Mulder introduce a Lagrange method that does the trick. The solution for $\hat{\boldsymbol{c}}$ and the Lagrange multipliers $\boldsymbol{\lambda}$ in matrix form is given by Equation 4.21, where the dispersion matrix $\boldsymbol{Q}$ is given by 4.22.

$$\begin{bmatrix} \hat{\boldsymbol{c}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{G}^\top \\ \boldsymbol{G} & 0 \end{bmatrix}^+ \cdot \boldsymbol{X}^\top \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \boldsymbol{C}_3 & \boldsymbol{C}_4 \end{bmatrix}^+ \cdot \boldsymbol{X}^\top \boldsymbol{Y} \tag{4.21}$$

$$\boldsymbol{Q} = \boldsymbol{X}^\top \boldsymbol{X} \in \mathbb{R}^{T\hat{d} \times T\hat{d}} \tag{4.22}$$

The pseudo-inverse in Equation 4.21 can be a computationally inefficient operation and to reduce run time it is beneficial to use an iterative solver. the solver in Equation 4.23 is such a method, called multiplier-based Matrix Iterative Solver (MIS) [35, 36].

$$\begin{aligned} \boldsymbol{c}^{(1)} &= \left(2\boldsymbol{Q} + \frac{1}{h}\boldsymbol{G}^\top \boldsymbol{G}\right)^{-1} \left(2\boldsymbol{X}^\top \boldsymbol{Y} + \frac{1}{h}\boldsymbol{G}^\top \boldsymbol{g} - \boldsymbol{G}^\top \boldsymbol{\lambda}^{(0)}\right) \\ \boldsymbol{c}^{(k+1)} &= \left(2\boldsymbol{Q} + \frac{1}{h}\boldsymbol{G}^\top \boldsymbol{G}\right)^{-1} \left(2\boldsymbol{Q}\boldsymbol{c}^{(k)} + \frac{1}{h}\boldsymbol{G}^\top \boldsymbol{g}\right), \quad k = 0, 1, 2 \ldots, M \end{aligned} \tag{4.23}$$

Not only can the least squares solution for each parameter be found, also the variance of each parameter and the covariance between parameters is easy to obtain once the matrices in Equation 4.21 are formed. Without additional efforts the covariance matrix of the B-coefficients is found by Equation 4.24.

$$\text{Cov}(\hat{\boldsymbol{c}}) = \boldsymbol{C}_1 \tag{4.24}$$

Equation 4.24 can be used to detect local model deficiencies. Another tool for such analysis is the bounding property of the simplex B-spine. The minimum and maximum value of the spline surface at any point on each simplex, is bounded by the minimum and maximum B-coefficient value in that simplex. This property is defined by Equation 4.25.

$$\min \boldsymbol{c}_{t_j} \leq p(\boldsymbol{b}) \leq \max \boldsymbol{c}_{t_j} \quad \forall\, \boldsymbol{b} \in t_j \tag{4.25}$$

As it turns out, sparse data sets with large regions without data often have inflated B-coefficients values and variance. This is especially problematic in constructing global aerodynamic models, where the edges of the flight envelope are typically very sparse.

4.3.5. Derivatives and Integrals

Simplex B-splines are an incredibly effective and versatile data modeling technique, that has a whole bunch of nifty applications. Many of these applications are described in [11]. 2 of these are discussed in the remainder of this section.

First of all, analytical directional derivatives are available for simplex B-splines, expressed in terms of the original B-coefficients [11, 34].

$$D_{\boldsymbol{u}}^{m} p(\boldsymbol{b}) = \frac{d!}{(d-m)!} \boldsymbol{B}_t^{d-m}(\boldsymbol{b}) \boldsymbol{P}^{d,d-m}(\boldsymbol{a}) \cdot \boldsymbol{c}_t \tag{4.26}$$

$\boldsymbol{P}^{d,d-m}$ is the degree m Casteljau matrix, that can be employed to achieve a transformation from a degree m B-coefficient vector to a degree $d-m$ B-coefficient vector. $\boldsymbol{P}^{d,d-m}$ lives in $\mathbb{R}^{\hat{d}^* \times \hat{d}}$ with $\hat{d}^*$ given by Equation 4.27. $\boldsymbol{B}_t^{d-m}(\boldsymbol{b})$ and $\boldsymbol{c}_t$ are the local spline basis and B-coefficients on simplex t. Furthermore, $\boldsymbol{u}$ is the direction of the derivative. Since $\boldsymbol{u}$ is a vector, it has no Barycentric coordinate. To make the coordinate transformation possible, $\boldsymbol{u}$ is decomposed in $\boldsymbol{v}$ and $\boldsymbol{w}$, see Equation 4.28. Then $\boldsymbol{a}$ is the directional coordinate of $\boldsymbol{u}$, as given by Equation 4.29.

$$\hat{d}^* = \frac{(d-m+n)!}{(d-m)!n!} \tag{4.27}$$

$$\boldsymbol{u} = \boldsymbol{v} - \boldsymbol{w} \tag{4.28}$$

$$b(\boldsymbol{u}) = \boldsymbol{a} = b(\boldsymbol{v}) - b(\boldsymbol{w}) \in \mathbb{R}^{n+1} \tag{4.29}$$

In addition to directional derivatives, exact integral solution can also be found from B-coefficients [11]. The surface under a simplex polynomial is given by Equation 4.30.

$$\int_t p(\boldsymbol{b}) d\boldsymbol{b} = \frac{A_t}{\hat{d}} \sum_{|\kappa|=d} c_k^t \tag{4.30}$$

A_t is the volume of the concerned n-simplex and c_k^t the B-coefficient in simplex t at local coordinate κ in the B-net. Extending this concept, the integral of the global spline is simply found by applying the same trick over the entire triangulation, see Equation 4.31.

$$\sum_{j=1}^{J} \int_{t_j} p(\boldsymbol{b}) d\boldsymbol{b} = \sum_{j=1}^{T} \frac{A_{t_j}}{\hat{d}} \sum_{|\kappa|=d} c_k^{t_j} \tag{4.31}$$

4.4. Physical Spline Theory

Simplex B-splines can approximate functions to an arbitrary degree. To achieve higher approximation power the spline degree is increased. This also introduces many difficult to interpret terms, in contrast to the linear polynomial models common in the aerospace industry, that include control and stability derivatives with a clear physical meaning. Some of these higher order terms might improve the model fit, however the opposite might also be true for others. This can introduce issues related to overfitting. The low interpretability is also a consequence of the Barycentric coordinate system. Namely, the B-coefficients do not have a physical meaning.

Huisman et al. came up with a method to improve interpretability of the spline model, allow model structure selection and introduce the possibility to incorporate a-priori knowledge of the physical system [37]. The method proposed is a transformation from the Barycentric coordinates to physical coordinates. This way the abstract B-coefficients can be related to familiar parameters such as C_{m_α}, but also to less familiar, yet easier to interpret parameters like $C_{m_{\alpha^3\beta^2}}$. As mentioned, the method also makes it possible to perform model structure selection, so that it can be assessed fairly whether a parameter such as $C_{m_{\alpha^3\beta^2}}$ provides added value to the model or that it only leads to overfitting. In case the latter is concluded, constraints can be imposed such that the Barycentric B-coefficients are estimated in such a way that their physical counterpart will have a 0 value for the redundant parameter.

The transformation from B-coefficients in Barycentric space $\boldsymbol{c}_b$ to the physical counterpart $\boldsymbol{c}_p$ is given by Equation 4.32

$$\begin{aligned} \boldsymbol{c}_p &= \begin{bmatrix} \boldsymbol{\Lambda}_{t_1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Lambda}_{t_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Lambda}_{t_T} \end{bmatrix} \boldsymbol{c}_b \\ &= \boldsymbol{\Lambda}\boldsymbol{c}_b \end{aligned} \tag{4.32}$$

The $\boldsymbol{\Lambda}_{t_j}$ is the Barycentric to Cartesian transformation matrix for simplex t_j. The exact derivation can be found in [37].

The parameter estimation is still done in Barycentric space to keep the benefits of the local coordinate system. However the default set of constraints (Equation 4.20) is extended to a set including constraints in Cartesian space. This new set consists of four parts, as listed in Equation 4.33.

$$\begin{aligned} &\boldsymbol{H}(\boldsymbol{c}_b) = \boldsymbol{h}_e && \text{Barycentric equality constraints} \\ &\boldsymbol{G}(\boldsymbol{c}_p) = \boldsymbol{g}_e && \text{Physical equality constraints} \\ &\boldsymbol{G}_u(\boldsymbol{c}_p) \leq \boldsymbol{g}_u && \text{Physical inequality constraints, upper bound} \\ &\boldsymbol{G}_l(\boldsymbol{c}_p) \geq \boldsymbol{g}_l && \text{Physical inequality constraints, lower bound} \end{aligned} \tag{4.33}$$

The first constraint in Equation 4.33 is the same as the one from Equation 4.20. The second constraint is used to force physical model parameters to zero when they are deemed redundant to the model. The MOF-algorithm, that was already introduced in Section 2.4, is used to select the useful terms. the first order terms are frozen into the model as a default. The last 2 physical inequality constraints are used to inject information about the physical stability derivatives of the Citation II. This information is shown in Equation 4.34. Note that the these stability derivatives are not model parameters of the C_m model in physical space, but the derivatives of C_m with respect to α, q and δ_e and thus are a linear combination of physical coefficients themselves.

$$-2 \leq C_{m_\alpha} \leq 0, \quad -22 \leq C_{m_q} \leq -1.5, \quad -3 \leq C_{m_\delta} \leq -0.5, \tag{4.34}$$

Huisman et al. found that restricting the solution space too much by imposing many constraints, tends to yield infeasible solutions. Therefore he proposes to use constraints on constraints: set a maximum number of constraints. A good rule is a limit of $d - r - 1$ constraints. Furthermore it was found that stability derivatives C_{m_q} and $C_{m_{\delta_e}}$ were well bounded on all simplices, leading to big improvements compared to the unbounded case. However, the C_{m_q} estimates appeared to lie close to the upper and lower bound of the boundaries, implying underlying problems. Lastly it should be noted that that in this study no stall data set was used, but data from nominal flight. To illustrate this, it was found that the C_m model is highly linear, revealing a lack of high angle of attack data.

4.5. Conclusions

Multivariate Simplex B-splines is one of various multivariate spline models. However, the simplex B-spline is the only real candidate for a nonlinear global system identification framework. That is due to its ability to fit scattered data sets, local polynomial basis and access to efficient parameter estimation methods. The simplex B-spline are piecewise Bernstein polynomials defined on a geometric structure called a triangulation. These splines generally have an arbitrarily high approximation power, only with sparse data sets do they not live up to their potential. Sparse data sets tend to lead to inflated B-coefficients values and variances. Physical splines are an adaptation to the simplex B-spline. They combat overfitting by removing and setting boundaries on certain parameters in the physical space. Multivariate Simplex B-splines can likely be used for global aerodynamic system identification if the issue with sparse data sets can be resolved.

5

Ill-Posed Regression

system identification models rely on data to create a model that can predict the real world. In some cases the information content in that data is insufficient for the desired accuracy. If this is the case, model complexity needs to be reduced, otherwise an ill-posed problem results. Reducing model complexity is undesirable because it goes at the cost of approximation power. This chapter focuses on the cause, diagnostics and remedies of ill-posed regression. In Section 5.1 the concepts of over- and underfitting are discussed. In Section 5.2 the root cause of ill-posed regression is explained. Section 5.3 and Section 5.4 give a short overview of possible diagnostics and remedies respectively. A conclusion on what has been found on these topics is provided in Section 5.5.

5.1. Bias-Variance Tradeoff

In regression problems such as spline regression, a very important considerations is the tradeoff between bias and variance. Bias and variance are two different sources of error that arise during the development of a predictive model. [38]

Bias errors arise when the model is underfitted. In this scenario the model is not complex enough to accurately describe the system of interest. To reduce bias, the model complexity needs to be increased. In case of B-splines, this can be done by increasing the spline degree. On the other hand, errors due to variance occur when the model is overfitted. By continuously increasing the model complexity, the model is capturing more and more variation in the data. In the case of system identification of aerodynamic systems, the model is capturing increasingly more complex dynamics. At some point the model complexity will exceed the information content present in the data. At this point, the model is so complex that it starts to capture the noise in the measurements, and the model will start to replicate this noise. This error will especially become apparent when testing the model on unseen data. In other words, the model does not generalize well to new data. To reduce the variance errors, model complexity must be decreased. Decreasing bias and variance errors are conflicting objectives and therefore the best model is obtained by balancing the two.

Figure 5.1 depicts the trade-off. In this figure bias and variance are monotonically decreasing and increasing functions of complexity, respectively. The total error is U-shaped and has a minimum that is to be found. Note that only the test errors are shown, computed from a validation data set.

5.2. Inverse Problems

Stall models obtained through flight test data, retrieve the stall dynamics from observations of this phenomenon. The process of retrieving the causal factors that led to a set of observations through the observations themselves, is called an inverse problem. A linear inverse problem in the ideal case can be denoted like Equation 5.1.

$$\boldsymbol{c} = \boldsymbol{X}^{-1}\boldsymbol{Y} \tag{5.1}$$

When a linear system is over- or underdetermined, or severe noise is present, Equation 5.1 does not hold. Then the inverse problem can for example be posed like a least-squares problem, such as in Equation 4.19.

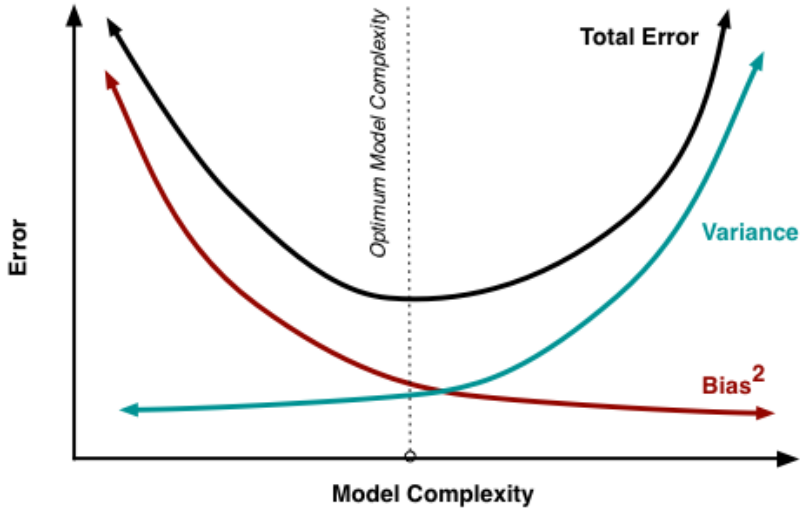


Figure 5.1: Illustrative bias, variance and total error curves as a function of model complexity, taken from [38].

Here the quality of the solution to the problem is strongly affected by the quality of the observations. In case measurements are too noisy or too limited to obtain a satisfactory model, the inverse problem is called ill-posed. To be more precise, an inverse problem is ill-posed when it fails to meet one of the Hadamard criteria [39]:

- The solution exists
- The solution is unique
- The solution is stable

In other words, ill-posed Inverse problems are characterized by inconsistency and sensitivity to noise. The solution to the problem is either non-unique or small changes in the input data can lead to large changes in the output, if a solution exists at all.

The presence of ill-posedness often becomes apparent through numerical issues with the inverse of the matrix $\boldsymbol{X}$. It is exactly what simplex splines suffer from. The cause of the numerical problems is often caused by a lack of information in the data. collinearity or locally missing data lead to small singular values of matrix $\boldsymbol{X}$. As was illustrated before, stall data is especially prone to locally missing data, due to difficulties in covering the edges of the flight envelope. These data deficiencies lead to overfitting and consequently numerical issues. By looking deeper into the numerical issues, a cure for this overfitting might be found.

The inverse of $\boldsymbol{X}$ has a big contribution in the B-coefficient parameter variance (Equation 4.24). The numerical issues with the inverse of $\boldsymbol{X}$ can be better studied by looking at the Singular Value Decomposition (SVD) of $\boldsymbol{X}$. It is given by Equation 5.2 [40].

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{5.2}$$

$\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices, and are known as the left and right singular vector matrices, respectively. If $\boldsymbol{X}$ is the matrix from Equation 4.15 in $\mathbb{R}^{N \times T \cdot \hat{d}}$, then $\boldsymbol{U}$ is of size $N \times N$, $\boldsymbol{\Sigma}$ of size $N \times T \cdot \hat{d}$ and $\boldsymbol{V}$ of size $T \cdot \hat{d} \times T \cdot \hat{d}$. Matrix $\boldsymbol{\Sigma}$ has the singular values of $\boldsymbol{X}$ on its diagonal. The inverse of $\boldsymbol{X}$ can be composed as a sum of the vectors of $\boldsymbol{U}$ and $\boldsymbol{V}$, and the singular values in $\boldsymbol{\Sigma}$. If we now look at the inverse problem Equation 5.1, it can be rewritten to Equation 5.3

$$\hat{\boldsymbol{c}} = \boldsymbol{X}^{-1}\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T\boldsymbol{Y} = \sum_{i=1}^{r} \frac{v_i \boldsymbol{Y}}{\sigma_i} u_i^T \tag{5.3}$$

With r the rank of matrix $\boldsymbol{X}$. Especially the summation in Equation 5.3 is interesting. Each term is divided by a singular value σ_i of $\boldsymbol{X}$. If only one of the σ_i has a very small value, the entire expression blows up. In other words, if the chosen model complexity is too high in only one simplex, the entire model suffers. This is exactly what happens in ill-posed problems and spline estimation with sparse or collinear data.

5.3. Diagnostics

To pinpoint the source of ill-conditioning or to be able to say something about its severity, it is important to have some diagnostic tools to disposal. In this section three ill-conditioning diagnostics are discussed.

One very useful diagnostic for identifying trouble in regression problems are the Variance Inflation Factors (VIF) [41]. The VIF of a model parameter with index i is given by Equation 5.4.

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \tag{5.4}$$

Here R_i is the coefficient of determination. The VIFs represent the increase of a model parameter's variance compared to the overall residual variance σ^2. In practise, VIFs over 5 to 10 indicate a badly estimated model coefficient.

The VIF approach requires the model parameters to be estimated. In case of severe singularities, estimating a model in the first place can be very difficult, if not impossible. For this reason, it is also useful to have a diagnostic tool that can be deployed without making the model. The condition number is such a tool [41]. The condition number is defined as the ratio between the largest and smallest eigenvalue of the matrix $\boldsymbol{X}^T\boldsymbol{X}$. Condition numbers up to a 100 mean a fairly well-conditioned problem. Condition numbers between 100-1000 imply strong ill-conditioning and more than a 1000 severe ill-conditioning.

Lastly, the determinant of $\boldsymbol{X}^T\boldsymbol{X}$ is a very clarifying diagnostic. The determinant is one in case of a fully orthogonal matrix and zero in the presence of a linear dependency and thus ill-conditioning. [41]

5.4. Remedies

It turns out that simplex B-splines have great potential to be a great tool for aerodynamic system identification, but the model estimation often turns out to be an ill-posed problem. There are methods available to counter this ill-posedness. The most important are discussed here.

Augmenting the data set with additional data is perhaps the most obvious, and also the most effective remedy [41]. Especially covering the sparse regions near the edges of the flight envelope is of great benefit. Unfortunately, this is expensive and often unpractical in the context of flight tests, as it is difficult and dangerous to keep the aircraft at the extremes of the flight envelope.

Decreasing the spline degree and thus the model complexity is another option. As explained in Section 5.1, this will reduce the model variance, however also increase bias. Reducing variance while keeping model complexity high is thus preferred. Other than decreasing spline degree, increasing spline continuity is beneficial too for decreasing parameter variance [30]. The same goes for other constraints. Differential constraints have shown to be effective at reducing B-coefficient variance [34]. Also the constraining physical splines by Huisman et al. have shown to be useful in combating overfitting, without reducing model order [37].

A high-quality triangulation is helpful in preventing ill-conditioning, in particular for dimensions exceeding two. Here a standard Delaunay triangulation is likely to result in near-empty simplices [11]. A better triangulation method is an adaptation of Kuhn triangulations [31, 32]. This method is elaborated on in [10].

Lastly, regularization has been proven to be a very powerful tool in regression applications. Many regularization methods exist, that can be almost effortlessly implemented in most conventional linear regression schemes. The idea of regularization is to introduce a small amount of bias into the system, while reducing the variance significantly. One regularization method that has been successfully applied to simplex B-splines by Steiner et al. [12], is Tikhonov regularization [42]. This is essentially done by adding prior information on the B-coefficient values.

5.5. Conclusions

A model complexity that is too low leads to a low approximation power, and gives rise to bias errors. A model complexity that is too high leads to overfitting and variance errors. The right model complexity is found somewhere in the middle of these two extremes. Splines fitted to sparse or collinear data tend to be overfitting and this leads to ill-posed regression problems. The regression matrix in ill-posed problems is characterized by small singular values, leading to inflated parameter estimates. Ill-posedness can be detected with VIFs, condition numbers and the determinant of $\boldsymbol{X}^T\boldsymbol{X}$. Possible remedies are adding data points, decreasing the spline degree, improving the triangulation and most importantly regularization.

6

Regularization

The simplex B-spline has one major shortcoming: its inability to accurately fit sparse and collinear data sets. Attempting to do this often leads to an ill-posed regression problem. Regularization is one potentially powerful remedy. In this chapter the most well known class of regularization methods, and in particular one member, is discussed. This class is spectral regularization and the class member Tikhonov regularization. In Section 6.1 and Section 6.2 spectral and Tikhonov regularization are covered, respectively. In Section 6.3 the implementation of Tikhonov regularization in various spline models is talked about. In Section 6.4 and Section 6.5 tuning methods for regularization parameters are reviewed. Regularization parameters govern the amount of regularization introduced in the model. A conclusion on the findings in this chapter is found in Section 6.6.

6.1. Spectral Regularization

It was explained in Chapter 5 that data collinearity or simply a lack of data can lead to numerical issues in regression problems. This can make it very difficult to obtain a good model from a set of observations. Models obtained through ill-posed inverse problems will often only capture noise in the data. These numerical issues become apparent through the singular values of regression matrix $\boldsymbol{X}$. Small singular values cause inflated parameter standard deviation. If one would be able to improve these singular value, without changing the problem so much that the solution becomes useless, more information could be retrieved from the measurements.

The class of algorithms that attempts to combat ill-posedness by targeting the numerically unstable singular values of the regression matrix is called spectral regularization. Each of these algorithms is characterized by filter factors f. The regularized solution to the inverse problem given by Equation 6.1 [43].

$$\hat{\boldsymbol{c}}_\mu = \sum_{i=1}^{n} f_i \frac{v_i \boldsymbol{Y}}{\sigma_i} u_i^T \tag{6.1}$$

With μ the regularization parameter. The most straightforward spectral regularization algorithm has filter factors valued either zero or one [43]. This algorithm is called Truncated Singular Value Decomposition (TSVD). The components that cause issues are fully removed and the others are used. There are many more algorithms with more complicated filter factors. They have one thing in common: the filter factors act as a high pass filter on the singular value spectrum. The higher values ought to be untouched while the problematic lower values are filtered out. μ is used to control the range of singular values that are mostly affected by the filter factors. In case of TSVD, μ can be considered the cut-off point of the singular values kept in the solution.

When the right amount of regularization is applied, the problem becomes well posed and the variance of the model is reduced. Consequently, the model becomes more robust to noise and overfitting. However, this goes at the cost of some bias. The model has been slightly altered and as a result, it might not be able to capture all the system dynamics perfectly.

6.2. Tikhonov Regularization

One of the most well known spectral regularization algorithms is named Tikhonov regularization [42]. Especially in regression problems it is a popular method due to its ease of implementation and strong performance in combatting ill-posedness, without changing the model structure or removing data [44]. It turns out that many other common regularization methods practically yield the same results as Tikhonov, given that suitable regularization parameters are chosen [43]. This means that the choice of a suitable regularization method is not so much a matter of performance, but more one of convenience. Especially the ease of integration in classical regression makes Tikhonov regularization a very interesting option to implement in Simplex B-splines.

The Tikhonov filter factors are given by Equation 6.2 [45]. Note that when $\sigma_i << \mu$, the filter factors are approximately zero, while they are close to one if $\sigma_i >> \mu$. As explained in Section 6.1, these are the properties expected of a good regularization method.

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \mu} \tag{6.2}$$

Tikhonov regularization is a whole class of algorithms on itself. The method is implemented by adding a penalty term to the least squares problem. Each Tikhonov regularization method has its own characteristic penalty matrix. The general implementation is a natural extension of least squares, explaining the popularity in regression. The extension of Equation 4.19 for Tikhonov is given by Equation 6.3.

$$\hat{\boldsymbol{c}} = \underset{\boldsymbol{c}}{\operatorname{argmin}} \left\| \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{c} \right\|_2^2 + \mu \left\| \boldsymbol{\Gamma}\boldsymbol{c} \right\|_2^2 \tag{6.3}$$

The first term in Equation 6.3 ensures a good fit to the data while the second term applies a penalty to the parameter vector. Regularization parameter μ regulates the trade-off between the two objectives. The type of penalty is decided through the Tikhonov Matrix Γ. Some of the penalty types include constraints on the norm of the parameter vector and difference operators. By expanding the latter term in Equation 4.19, the penalty matrix $\boldsymbol{P}$ is obtained. This is shown in Equation 6.4.

$$\left\| \boldsymbol{\Gamma}\boldsymbol{c} \right\|_2^2 = \boldsymbol{c}^\top \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma} \boldsymbol{c} = \boldsymbol{c}^\top \boldsymbol{P} \boldsymbol{c} \tag{6.4}$$

Ridge regression is one of the algorithms that is part of the Tikhonov regularization class. It was first introduced by Hoerl and Kennard [44] and it adds a squared norm constraint. Its associated penalty matrix is the identity matrix: $\boldsymbol{P} = \boldsymbol{I}$. In the remainder of this subsection the findings by Hoerl and Kennard will be presented to illustrate the strength of Tikhonov.

First the distance between the unknown true parameter vector $\boldsymbol{c}$ and the OLS parameter estimate $\hat{\boldsymbol{c}}$ is defined as L_1, its square is given by Equation 6.5. Obviously it is desirable to have this quantity as low as possible.

$$L_1^2 = \left(\hat{\boldsymbol{c}} - \boldsymbol{c}\right)^T \left(\hat{\boldsymbol{c}} - \boldsymbol{c}\right) \tag{6.5}$$

The expectation and variance of L_1^2 is given by Equation 6.6 and Equation 6.7 if we consider only one simplex. Once more, it is desired to have these quantities as small as possible.

$$E\left[L_1^2\right] = \sigma^2 \sum_{i=1}^{\hat{d}} \frac{1}{\lambda_i} \tag{6.6}$$

$$VAR\left[L_1^2\right] = 2\sigma^4 \sum_{i=1}^{\hat{d}} \frac{1}{\lambda_i}^2 \tag{6.7}$$

The λ_i in Equation 6.6 and Equation 6.7 are the eigenvalues of $\boldsymbol{X}^T\boldsymbol{X}$. They are equal to the squared singular values of $\boldsymbol{X}$. Once more it shows that small numerically troublesome eigenvalues will lead to poor

parameter estimation. Only one eigenvalue near zero can blow up both the precision of the solution, as well as the uncertainty in the solution.

The solution to the Regularized Least Squares (RLS) solution with the ridge penalty is given by Equation 6.8. It is also possible to express this solution in terms of the OLS solution as shown in Equation 6.9.

$$\hat{\boldsymbol{c}}_{ridge} = \left[\boldsymbol{X}^T\boldsymbol{X} + \mu\boldsymbol{I}\right]\boldsymbol{X}^T\boldsymbol{Y} \tag{6.8}$$

$$\hat{\boldsymbol{c}}_{ridge} = \left[\boldsymbol{I} + \mu\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right]^{-1}\hat{\boldsymbol{c}}_{OLS} \tag{6.9}$$

Further, L_1^2 can be reexpressed, now as the distance between $\boldsymbol{c}$ and $\hat{\boldsymbol{c}}_{ridge}$. This gives Equation 6.10.

$$\begin{aligned} E\left[L_1^2(\mu)\right] &= E\left[\left(\hat{\boldsymbol{c}}_{ridge} - \boldsymbol{c}\right)^T\left(\hat{\boldsymbol{c}}_{ridge} - \boldsymbol{c}\right)\right] \\ &= \sigma^2\sum_{i=1}^{\hat{d}}\frac{\lambda_i}{\lambda_i+\mu}^2 + \mu^2\boldsymbol{c}^T\left(\boldsymbol{X}^T\boldsymbol{X} + \mu\boldsymbol{X}\right)^{-2}\boldsymbol{c} \\ &= \gamma_1(\mu) + \gamma_2(\mu) \end{aligned} \tag{6.10}$$

$\gamma_1(\mu)$ is the total variance of the parameter estimates, while $\gamma_2(\mu)$ is a measure for the square of the bias introduced when the ridge estimate is used. It can be shown that $\gamma_1(\mu)$ and $\gamma_2(\mu)$ are monotonically decreasing and increasing function of μ respectively. Also the values of these two quantities near zero is known, they are the limits in Equation 6.11 and 6.12.

$$\lim_{\mu\to 0^+}\left(\frac{\mathrm{d}\gamma_1}{\mathrm{d}\mu}\right) = 2\sigma^2\sum_{i=1}^{\hat{d}}\frac{1}{\lambda_i}^2 \tag{6.11}$$

$$\lim_{\mu\to 0^+}\left(\frac{\mathrm{d}\gamma_2}{\mathrm{d}\mu}\right) = 0 \tag{6.12}$$

According to these limits, the variance decreases notably for small values of μ, while the bias remains unaltered. Consequently, one can significantly decrease the variance in the solution estimate, at neglible additional bias. This concept is visualized in Figure 6.1. Hoerl and Kennard coined it the ridge trace. It clearly shows that the variance will continue to decrease for increasing μ, while the bias does the opposite. Indeed for small amounts of regularization does the variance decrease significantly, at little cost of bias. Furthermore, it can be shown that the optimal μ creeps closer towards zero for a larger sized true parameter vector norm $\boldsymbol{c}^T\boldsymbol{c}$. Obviously, the true parameter vector is unknown. Therefore regularization parameter tuning methods are needed. Methods to determine the optimal value of μ are discussed in Section 6.4 and 6.5.

Finally, Hoerl and Kennard also provide a proof that there exists some $\mu > 0$ such that $E\left[L_1^2(\mu)\right] < E\left[L_1^2(\mu = 0)\right]$. This theorem is known as the Existence Theorem. In other words, by using good μ tuning methods an improved solution is guaranteed.

6.3. Regularization and Splines

Regularization has great potential to solve the issues related to aircraft system identification with simplex splines. Especially the Tikhonov regularization method is promising as it is a natural extension of least-squares regression. In this section an array of spline researches (univariate and multivariate) with Tikhonov regularization is discussed.

6.3.1. P-splines

Approximation power of a univariate spline can be improved by increasing the number of knots [28]. This practise also induces overfitting problems. Reinsch pioneered a method to still achieve a smooth function

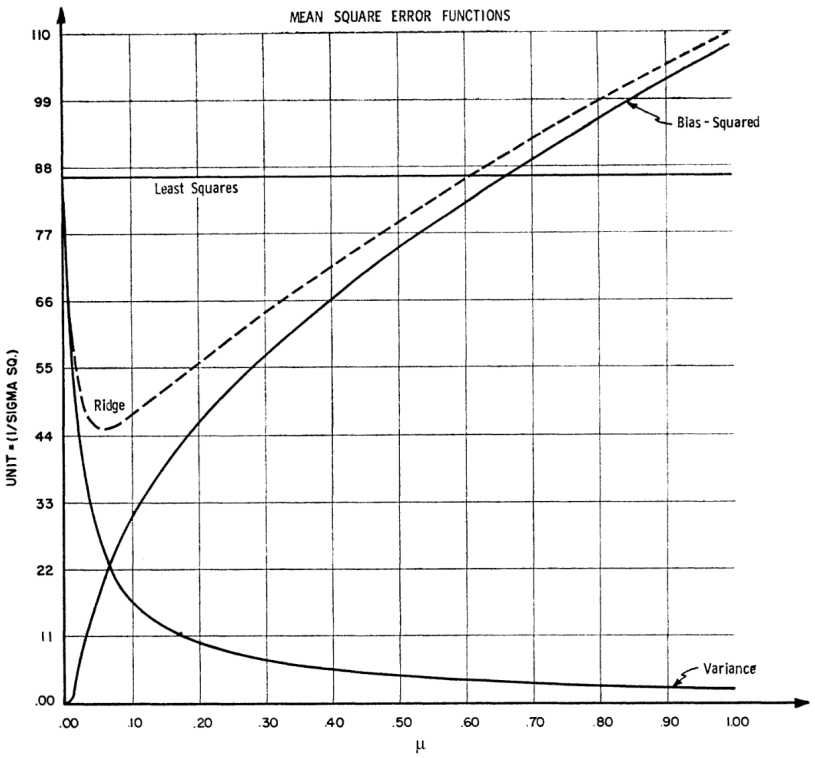


Figure 6.1: Ridge trace, taken from [44].

[46]. He added a smoothness penalty based on the integral of the second derivative of the fitted curve. This can be interpreted as the average curvature of the spline. By doing this the flexibility of the spline is restricted. The resulting spline turns out to be a cubic spline. Later, O'Sullivan introduced a spline with a similar second derivative penalty [47]. In fact, the spline by O'sullivan is a direct generalization of the smoothing spline [48]. He posed the parameter estimation procedure as a regularization method.

Although such methods are mathematically very pleasing, the practicality of implementation leave something to be desired. Marx and Eilers came up with a generalization of O'Sullivan's method and called it the P-spline [29, 49]. The P-splines is in essence a B-spline with difference penalties on the B-coefficients. They have some desirable properties such as no boundary effects, straightforward implementation in regression, and conservation of the mean and variance in the data.

The P-spline estimation problem can be formulated like the Tikhonov estimator from Equation 6.3. Γ is the discrete differential operator D_k. The first order operator has elements of $\begin{bmatrix}-1 & 1\end{bmatrix}$ on the diagonal and everything else zero. The second order operator has elements of $\begin{bmatrix}-1 & 2 & -1\end{bmatrix}$ on the diagonal and everything else zero. This pattern can be extended to any desirable difference order k. The amount of constraints totals to $m-k$, with m being the number of model parameters. The corresponding penalty matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are shown in Equation 6.13.

$$P_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots \\ -1 & 2 & -1 & 0 & \cdots \\ 0 & -1 & 2 & -1 & \cdots \\ 0 & 0 & -1 & 2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad P_2 = \begin{bmatrix} -1 & 2 & -1 & 0 & 0 & 0 & \cdots \\ -2 & 5 & -4 & 1 & 0 & 0 & \cdots \\ 1 & -4 & 6 & -4 & 1 & 0 & \cdots \\ 0 & 1 & -4 & 6 & -4 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{6.13}$$

The interpolating and extrapolating properties of the P-spline are well known. When interpolating, the B-coefficients will form a polynomial of order $2k-1$, and when extrapolating a polynomial of order $k-1$.[49]

For example, a second order difference penalty will have third order interpolation (just like smoothing splines) and linear extrapolation. By forcing the B-coefficients to form a smooth curve, the fitted curve will also be smooth. The typical large variations in B-coefficients estimated from ill-posed problems are prevented.

Next, some of the result achieved using P-splines are shown. In Figure 6.2 a P-spline fitted to a sparse data set is shown. Even though no data is present in the middle of the domain, a smooth curve is fitted that can well approximate the data that is available. Not a single data point is present between many knots, yet a stable solution is found. note that one is extrapolating the curve when evaluating it inside the data empty region. Without data or a-priori knowledge there is no reason to believe this is the correct fit. However, at least the curve is stable and consequently the curve can be evaluated in the data rich parts of the domain.

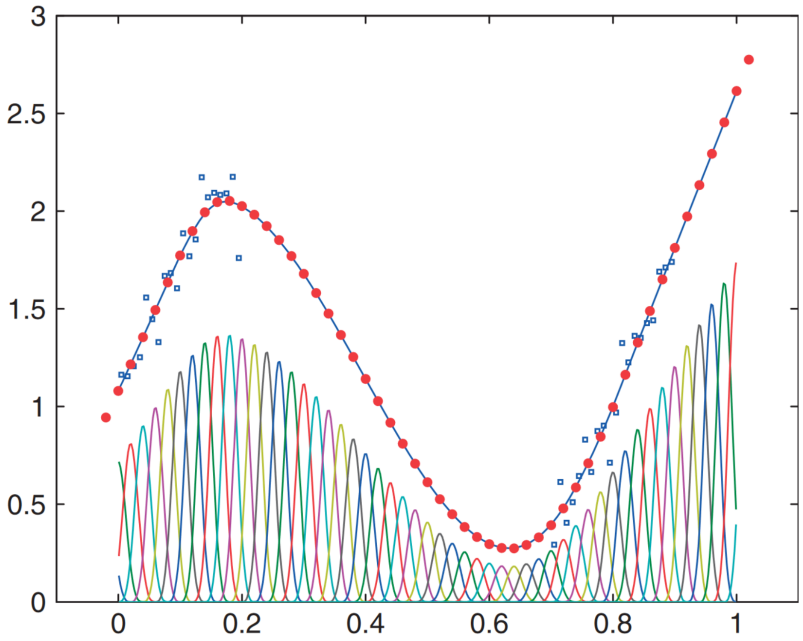


Figure 6.2: Third order P-spline with second order difference penalty, taken from [49]. The individual basis functions are shown on the bottom and their corresponding B-coefficient are shown as red dots.

After the introduction of the P-spline in 1986, many extensions to the original method have been proposed. A multivariate P-spline using tensor splines is one of them [50, 51]. This method only allows fitting of grid data and not scattered data. Furthermore, new penalty types were proposed different to the difference penalty [52]. They include asymmetric penalties (not all coefficients receive same penalty) to enforce shape constraints, the sign of the curve can be steered, and combinations of penalties can be made. For example a higher order for good approximation in the main body of data, and a lower order to get behaviour at the boundaries.

Lastly, one interesting development is the use of multiple penalties to account for different smoothness demand in the data [53, 54]. It was shown by Wand in 2000 that splines with multiple smoothing parameters are superior to those with only one [55]. Ruppert and Carroll find regularization parameters for subsections of their data set independently to later compile them into one set [53]. Currie and Durban split their problem into three, resulting in a regression matrix $\boldsymbol{B}$ and a penalty matrix $\boldsymbol{P}$ as in Equation 6.14 [54].

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & 0 & 0 \\ 0 & \boldsymbol{B}_2 & 0 \\ 0 & 0 & \boldsymbol{B}_3 \end{bmatrix}, \quad \boldsymbol{\lambda P} = \begin{bmatrix} \lambda_1 \boldsymbol{P}_1 & 0 & 0 \\ 0 & \lambda_2 \boldsymbol{P}_2 & 0 \\ 0 & 0 & \lambda_3 \boldsymbol{P}_3 \end{bmatrix} \tag{6.14}$$

In the subsequent, the terms regularization and smoothing parameter will be used interchangeably in the context of regularization with difference operator penalties.

6.3.2. O-splines

The P-spline was inspired by the spline introduced by O'Sullivan in 1986 [47]. The arguments of using the difference penalty instead of using the integral of the analytical second derivative is that the latter penalty is very impractical [29]. In response Wand and Ormerod provide an exact matrix expression in their 2008 paper [56] and name it the O-spline. The (simplified) penalty matrix they provide is given by Equation 6.15.

$$\boldsymbol{P} = (\boldsymbol{B}'')^T \boldsymbol{W} (\boldsymbol{B}'') \tag{6.15}$$

The penalty matrix derived is only for third degree splines with a second order penalty. $\boldsymbol{B}''$ the univariate spline basis belonging to the second derivative of a cubic spline. It is a $\boldsymbol{W}$ is a diagonal weight matrix with the diagonal elements found by applying Simpson's rule to each of the inter-knot differences.

When comparing their O-spline to a degree P-spline with order two difference penalty, they conclude that P-splines frequently stray from the typical boundary behavior of smoothing splines, while this is not the case for O-splines. The O-splines by Wand and Ormerod seem to behave slightly better at the boundaries, however the P-splines are still a lot more flexible. Their O-splines do not have the flexibility of mixing and matching different spline degrees with different penalty orders. Later, Wood derived a method to obtain any combination of spline degree with penalty order for the O-spline [48]. With this result the P-spline and O-spline are equal in terms of simplicity and flexibility. In terms of performance their is no clear evidence which method is to be preferred.

6.3.3. Thin-Plate Energy Penalty Splines

The O-spline is a direct generalization of smoothing splines, however it is only valid in univariate space. A multivariate extension of smoothing splines also exists, and it goes by the name of the thin-plate spline [56]. The thin-plate spline is not suitable itself for large scale system identification as it has a non-local basis, and consequently inefficient parameter estimation. The optimization has a computational efficiency of $\mathcal{O}(N^3)$ [11]. Yet, the thin plate can potentially be a useful tool for smoothing. Pfeifle and Seidel use a DMS spline (polyhedral spline) with a thin-plate energy penalty to fit scattered data with irregularities [57]. The DMS spline is a bivariate spline. The thin plate energy functional used as a smoothing term is given by Equation 6.16. Note that this is not a penalty matrix, Pfeiffle and Seidel only provide a mathematical description of their smoothing method.

$$J = \int_{\Omega} F_{xx}^2 + 2F_{xy}^2 + Fyy^2 \, \mathrm{d}x\mathrm{d}y \tag{6.16}$$

The penalty can be applied to a subset of the domain Ω. F is the fitted spline surface. Unfortunately the DMS spline is again mostly mathematically pleasing, and not so much practically. It does have some striking differences with simplex splines, however. It is defined on a triangulation and it has B-spline basis functions. On bivariate, small scale problems the effectiveness of the method is also demonstrated by Pfeifle and Seidel. In Figure 6.3 a result from their paper is shown.

In the top row a complete data set is used to fit a DMS spline. In the bottom row an incomplete data set is used to fit a DMS split with thin-plate energy smoothing . In the bottom quarter of the domain no data is present, yet a stable spline surface is manufactured.

An efficient parameter estimation algorithm for the DMS spline does not exist and it is therefore not feasible to extend it past the bivariate domain. However a n-dimensional thin-plate energy functional does exist that can potentially be exploited for smoothing, provided by Wood in 2003 [58]: Equation 6.17.

$$\int \cdots \int_{\mathrm{Re}^n} = \sum_{\nu_1! \ldots \nu_n! = m} \frac{m!}{\nu_1! \ldots \nu_n!} \left(\frac{\partial^m F}{\partial x_1^{\nu_1} \ldots x_n^{\nu_1}} \right)^2 \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{6.17}$$

With n the dimension of the data set, m the thin-plate penalty order (Pfeifle and Seidel used order two) and ν a multi-index.

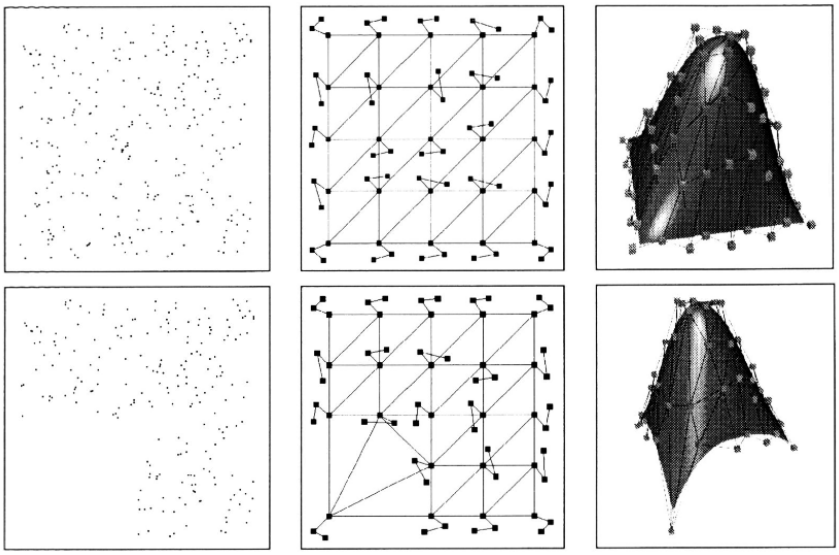


Figure 6.3: Top row: from left to right, scattered data points, triangulation and corresponding DMS spline surface. Bottom row: from left right, scattered data points with data removed in bottom left corner, triangulation and corresponding DMS spline surface with thin-plate energy smoothing. Taken from [57].

6.3.4. Simplex P-splines

Many examples of regularization applied to univariate splines exist, as discussed already in this chapter. Some exist for multivariate splines other than simplex splines, although those only restrict to non-scattered data fitting or lack efficient parameter estimation methods. The results achieved in the light of fitting sparse data sets are promising and have potential to solve issues related to the fitting of flight data with simplex B-splines. Still, it is not a trivial step to extend the concepts of well known regularized splines to simplex B-splines. Yet, it was successfully achieved by Steiner et al. in 2023 [12], when he applied a ridge and first order difference penalty to simplex B-splines to fit an aerodynamic data set with strong collinearities. This spline is referred to as the simplex P-spline.

The method by Steiner was inspired by the P-splines from Eilers and Marx, with a penalty matrix added to the ordinary spline regression problem. Additionally, Steiner implemented a ridge penalty. The implementation of the ridge penalty is rather straightforward. The identity matrix is used as penalty matrix like before, i.e. $\boldsymbol{P} = \boldsymbol{I}$. Steiner used a global regularization parameter μ, meaning that it was constant over all the simplices. It is however not impossible to have local regularization parameter throughout the triangulation. The complete penalty term including regularization parameter is then given by Equation 6.18. The ridge penalized simplex spline is named simplex RP-spline.

$$\boldsymbol{\mu P} = \text{diag}(\mu_1, \mu_2, \ldots, \mu_n)\boldsymbol{I} \tag{6.18}$$

The penalty matrix for the first order difference is much less straightforward to obtain. unlike P-splines, it is not possible to constrain the difference between each 2 subsequent B-coefficient as they are not structured in a linear fashion. Instead the B-coefficients in simplex splines are geometrically structured in the B-net, and consequently each B-coefficient has multiple neighbours. To obtain the difference equations, a forward difference is used. The forward difference is used over central or backwards differences for better geometrical interpretation. The order k forward difference at B-coefficient c_i in the direction u is given by Equation 6.19.

$$\begin{aligned}\Delta_{i,u}^{k} &= \Delta_{i+1,u}^{k-1} - \Delta_{i,u}^{k-1} \\ &= \sum_{j=0}^{k} (-1)^j \binom{k}{j} c_{\kappa_i + (k-j)\zeta_u} v \\ &= \boldsymbol{D}_u^k c_{\kappa_i + (k-j)\zeta_u}\end{aligned} \tag{6.19}$$

The difference operator $\boldsymbol{D}_u^k$ holds a difference constraint in a single direction. ζ is a multi-index that holds a set of difference offsets for all the coefficients involved. A difference matrix can be constructed for each coefficients by assembling the $\boldsymbol{D}_u^k$ in all directions, as in Equation 6.20.

$$\boldsymbol{D}_{c_i} = \begin{bmatrix} \boldsymbol{D}_1^k & \boldsymbol{D}_2^k & \dots & \boldsymbol{D}_{n+1}^k \end{bmatrix}^T \tag{6.20}$$

Similarly, a difference matrix exists for each simplex:

$$\boldsymbol{D}_{t_j} = \begin{bmatrix} \boldsymbol{D}_{c_1}^k & \boldsymbol{D}_{c_2}^k & \dots & \boldsymbol{D}_{c_{\hat{\delta}}}^k \end{bmatrix}^T \tag{6.21}$$

Finally, the global difference matrix $\boldsymbol{D}_g$ is given by Equation 6.22. The penalty matrix then is obtained straightforwardly with Equation 6.23. The difference constrained spline is called simplex DP-spline.

$$\boldsymbol{D}_g = \text{diag}\begin{bmatrix} \boldsymbol{D}_{t_1}^k & \boldsymbol{D}_{t_2}^k & \dots & \boldsymbol{D}_{t_J}^k \end{bmatrix} \tag{6.22}$$

$$\boldsymbol{P} = \boldsymbol{D}_g^T \boldsymbol{D}_g \tag{6.23}$$

The parameter estimation procedure remains mostly intact. Equation 4.21 is augmented with the penalty matrix, shown in Equation 6.24. Then the iterative solver from [35] and [36] is used again.

$$\begin{bmatrix} \hat{\boldsymbol{c}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q} + \mu \boldsymbol{P} & \boldsymbol{G}^\top \\ \boldsymbol{G} & 0 \end{bmatrix}^+ \cdot \boldsymbol{X}^\top \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \boldsymbol{C}_3 & \boldsymbol{C}_4 \end{bmatrix}^+ \cdot \boldsymbol{X}^\top \boldsymbol{Y} \tag{6.24}$$

The effectiveness of the P-spline is illustrated by looking at the variance plot of a P-spline. This plot shows the parameter standard deviation of each B-coefficients at its respective position in the B-net. For a practically unpenalized and optimally penalized RP-spline these surface are shown in Figure 6.4 and 6.5. Looking at the vertical axis, it can be observed that the parameter variance is drastically improved. While the spline from Figure 6.4 is practically useless because it offers no confidence at all that it will reproduce good results on data that is is not trained on, the opposite is true for Figure 6.5. The parameter surface is very stable and has a small mean. Note the vertical axes.

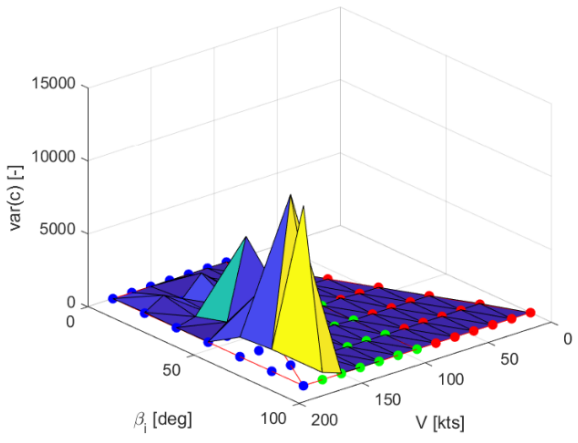


Figure 6.4: Variance surface of practically unpenalized simplex spline of degree 6, taken from [12].

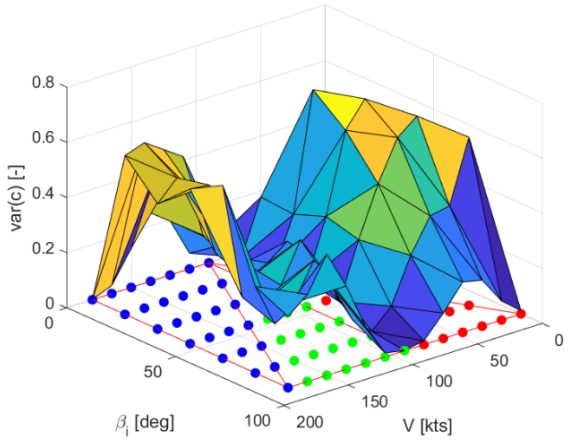


Figure 6.5: Variance surface of simplex spline of degree 6 with ridge penalty, taken from [12].

6.4. Univariate Regularizaton Parameter Tuning

The selection of an appropriate regularization parameter is a tedious task. When chosen too small, the parameter standard deviation might not have been reduced enough. When chosen too big a large amount of bias is introduced into the model. This trade-off was illustrated by the ridge trace Figure 6.1. It shows that parameter variance and bias are monotonically decreasing and increasing functions of the regularization parameter, respectively.

To the knowledge of the author, no optimal regularization parameter optimization methods exists in literature. However, a number of algorithms exist that use a heuristic approach to ensure a good level of smoothness. These methods each have a solid track record in literature and have shown to provide very good estimates of an optimal solution. The algorithms in the remainder of this section can only be used to tune a single regularization parameter.

6.4.1. Information Criteria

A straight forward approach is to vary μ over a range of feasible values in a grid search and pick the one that gives the 'best' result. The 'best' result can be quantified by some scoring metric. A common method in literature, is the use of an information criterion, that is a function of μ. Eilers and Marx use the Akaike Information Criterion (AIC) in their first paper on P-splines [29]. The best μ is deemed the one that minimizes the AIC. The AIC penalizes the number of model parameters and therefore a low AIC score means a lower complexity model that generalizes better to unseen data. It also takes into account the deviance from the data mean to ensure a close fit to the data.

Many more information criteria exist but they are typically not used in the context of regularization and splines. However, one interesting criterion worth mentioning is called Mallows C_p [59]. This method was found to be the best regularization method in a study that compared many selection methods for a spectroscopy application (no splines involved) featuring a least squares problem with ridge and difference penalties [60]. The study also included methods that are well established in the P-spline literature and that will be discussed in the subsequent. Mallow's C_p minimizes an estimate of the model's so called recovery error.

Information criteria generally have the unfavourable property that they require knowledge of the variance in the residuals, as well as that they assume uncorrelated noise in the data. Residual variance can be difficult to obtain in ill-posed problems.

6.4.2. Generalized Cross Validation Method

One scoring algorithm that can be used to evaluate the performance of a trial μ in a grid search, but that does not require knowledge of the noise level in the data, is generalized cross validation (GCV). It was first developed by Golub, Heath and Wahba [61] for choosing the μ in ridge regression problems. Later, both Eilers and Marx used it in their P-spline [29, 49], and Steiner et al. in their simplex P-spline [12], where they show that GCV is also effective in combination with difference penalties. The GCV score V_g is shown in Equation 6.25.

$$V_g(\mu) = \frac{N \left\| \boldsymbol{Y} - \boldsymbol{H}\boldsymbol{Y} \right\|^2}{\left[\text{tr} \left(\boldsymbol{I} - \boldsymbol{H} \right) \right]^2} \tag{6.25}$$

$\boldsymbol{Y}$ is the observation vector, N is the number of data points and $\boldsymbol{H}$ is the hat-matrix. For splines, $\boldsymbol{H}$ can be obtained with Equation 6.26.

$$\boldsymbol{H} = \text{trace}(\boldsymbol{G}) = \boldsymbol{X}\boldsymbol{C}_\mathbf{1}\boldsymbol{X}^T \tag{6.26}$$

For a more general expression, the work by Kagerer can be consulted [62]. The hat-matrix projects the observations $\boldsymbol{Y}$ to the model estimates $\hat{\boldsymbol{Y}}$, as in Equation 6.27 [61].

$$\hat{\boldsymbol{Y}} = \boldsymbol{G}\boldsymbol{Y} = \boldsymbol{B}\hat{\boldsymbol{c}} \tag{6.27}$$

The trace of $\boldsymbol{H}$ is the effective dimension of the model. If no continuity, constraints or regularization are used, the effective dimension is equal to the number of model parameters. If they are used, the effective dimension decreases.

GCV banks on the concept of leave one out cross validation. By evaluating the model at every data point with a model that has been trained on all data except that one data point, a good assessment can be made on the performance at a certain level of regularization. The beauty of this method is that there is no need to train N models per trial μ, instead the hat-matrix is used to scale the evaluation of each data point while it has been trained on all data. One shortcoming of GCV is that it can not always cope well with correlated noise [63].

6.4.3. L-Curve Method

The final means for regularization parameter optimization discussed in this section are the L- and V-curve methods. Originally, the L-curve was introduced by Hansen [43]. The outcome of his work is a graphical tool that can be used to seek out a good regularization parameter. His work suggests to plot on a log-log scale the solution semi-norm $||\boldsymbol{L}\boldsymbol{c}_\lambda||$ versus the residual vector norm $||\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{c}||$, each evaluated at a range of μ. The matrix $\boldsymbol{L}$ is a full rank matrix and in application often taken as the Tikhonov matrix $\boldsymbol{\Gamma}$. The resulting plot has an L-shape and it turns out that the optimal degree of regularization occurs at the point of maximum curvature, that is in the corner of the L. In Figure 6.6, an example of the L-curve is displayed. Some of the points that make up this curve have their corresponding regularization parameter mentioned. The L-shape can be clearly distinguished. To the left of the corner, the solution norm decreases steeply with the residual norm increasing insignificantly. After the corner has been rounded, the opposite is true. A watertight explanation on what makes the corner so special does not exist. Yet, it is clear that the corner strikes a nice balance between a feasible solution (small solution norm) and a close fit to the data (small residual norm).

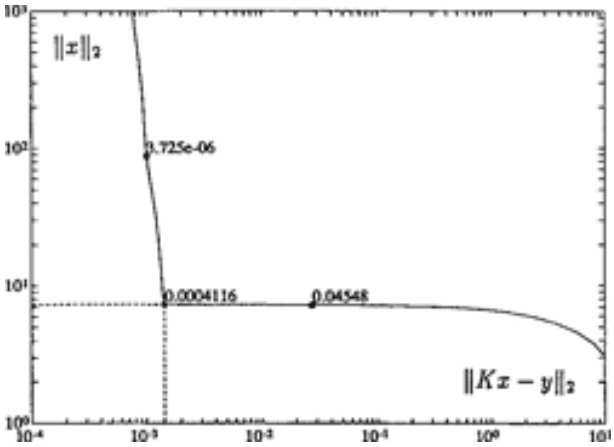


Figure 6.6: L-curve generated from a data set with correlated noise, taken from [43]. The value of μ of 0.0004116 gives the point in the corner of the L.

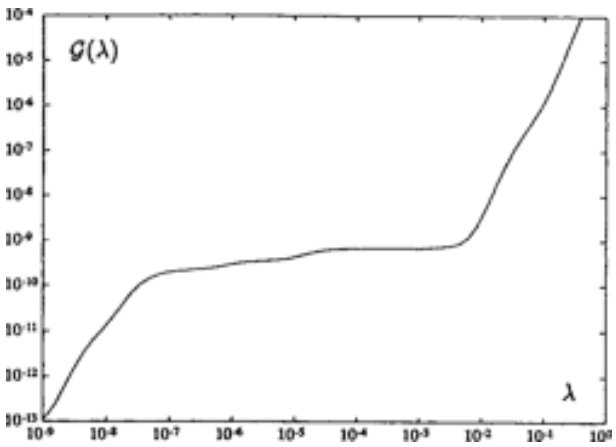


Figure 6.7: GCV gurve generated from the same dataset as Figure 6.6, taken from [43].

Contrary to the GCV method, one of the strengths of the L-curve is the ability to deal well with correlated errors. Whereas Hansen found that the solution to both methods is very similar in the presence of purely white noise, the opposite is true for correlated noise. In Figure 6.7, the GCV as a function of μ is shown for the exact same problem with correlated noise as the one from Figure 6.6. The minimum found by the GCV method reaches its minimum outside the order of machine precision. Furthermore, one thing to note about the L-curve is that the sharpness of the corner levels off when difference penalties are being used instead of a ridge penalty [63]. This makes the optimal smoothing parameter selection more arbitrary.

Frasso and Eilers found it to be difficult in general to pin point the corner of the L is case of a discretized curve [63]. It is also not easy to find a smoothing parameter automatically, since there is not some score one can minimize as a function of μ. For easier and automatic parameter selection with the L-curve they introduced the V-curve. They observed that subsequent points on an equidistant grid that make up the L-curve are spaced closer near the corner. The V-curve plots the euclidean distance between two

subsequent points on the L-curve versus the mean of their μs. This gives a V-shaped curve, that has a minimum near the point that gives the corner of the L.

6.5. Multivariate Regularizaton Parameter Tuning

Using the methods from Section 6.4, a single regularization parameter can be found. In the context of simplex splines this might not be sufficient. The degree of ill-posedness will likely vary between simplices, and this asks for locally varying regularization parameters. Wand found that P-splines with multiple smoothing parameters are superior to those with only one [55].

The approach used for univariate penalties boils down to evaluating some information some metric at a range of possible μ. This is typically done in a grid search. A grid search is computationally unfeasible for tuning many parameters. The problem size for an M-dimensional grid and n regularization parameters will require M^n evaluations. This section describes three algorithms that are capable of reducing the problem size and potentially generate good regularization parameters. These methods are a spatially adaptive method, a Gauss-Newton method for GCV and a fixed point algorithm for the L-curve, described in Section 6.5.1-6.5.3, respectively.

6.5.1. Spattially Adaptive Method

There is one contribution in the P-spline literature that uses locally varying smoothing parameters, and tunes the parameters with more sophisticated means than a brute grid search. Ruppert and Carroll introduced univariate P-splines with spatially adaptive penalties [53].

This method allows the smoothing parameter to vary between knots. At first they select a subset of the knots at which the parameter changes, so that initially the penalty does not change at each and every knot. Then they linearly interpolate between the 'subknots' to give unique penalties between the original knots.

In this method, first a global smoothing parameter is to be determined. Next, the local smoothing parameters are found one by one by performing a one-dimensional grid search between the relevant subknots. This grid search can for example be done in combination with GCV or the L-curve. During this process, the parameters are updated subsequently. Although this method does not guarantee finding a global minimum, Ruppert and Carroll found that it produced satisfactory results.

They also tried to add more iterations to the algorithm. At new iterations the initial μs are not the global μ but those from the previous iteration. Interestingly, they found that more iterations do not results in significant improvements. Ruppert and Carroll concluded that locally varying penalties are as effective as a global one when there is little spatial variability in the data. On the other hand, when there is a lot of variability, their method is far superior.

6.5.2. Newton method for GCV

Gu and Wahba published a paper in 1991 on an efficient algorithm for optimizing GCV scores with multiple smoothing parameters [64]. This work was continued by Wood in a range of papers [65, 66, 67].

The method initially proposed by Wood aims to tune smoothing parameters in a weighted least squares problem. Wood improved his own method in 2004 [66]. He mentions that his method from 2000 and Gu and Wahba's could fail in some rank deficient problems. In this paper he tries to solve the objective in Equation 6.28.

$$\hat{\boldsymbol{c}} = \underset{\boldsymbol{c}}{\operatorname{argmin}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{c}\|^2 + \boldsymbol{c}^T\boldsymbol{S}\boldsymbol{c} + \sum_{i=1}^{n} \mu_i \boldsymbol{c}^T \boldsymbol{P}_i \boldsymbol{c} \tag{6.28}$$

Equation 6.28 is the extension of Equation 6.3 to multiple regularization parameters, with the addition of a global penalty matrix $\boldsymbol{S}$. The term $\boldsymbol{c}^T\boldsymbol{S}\boldsymbol{c}$ can be used to add for example a global ridge penalty in addition to the local penalties. Alternatively, it can be used to fix some smoothing parameters while estimating the remainder, or impose a lower bound on the smoothing parameters.

To optimize the GCV score, Wood uses Newton updates. Equation 6.29 shows this optimization scheme for a function f that is minimized with respect to a variable x [68]. This equation requires the gradient and Hessian of f, which are g_f and H_f respectively.

$$x_{k+1} = x_k - [H_f(x_k)]^{-1} \cdot g_f(x_k) \tag{6.29}$$

$$\frac{\partial V_g}{\partial \eta_i} = \frac{N}{\delta}\frac{\partial \alpha}{\partial \eta_i} - \frac{2n\alpha}{\delta^3}\frac{\partial \delta}{\partial \eta_i} \tag{6.30}$$

The function f is in this case obviously the GCV score V_g. The required derivatives are derived by Wood in Equation 6.31.

$$\begin{aligned}\frac{\partial^2 V_g}{\partial \eta_i \partial \eta_j} =& -\frac{2N}{\delta^3}\frac{\partial \delta}{\partial \eta_j}\frac{\partial \alpha}{\partial \eta_i} + \frac{N}{\delta^2}\frac{\partial^2}{\partial \eta_i \partial \eta_j} \\ &- \frac{2N}{\delta^3}\frac{\partial \alpha}{\partial \eta_j}\frac{\partial \delta}{\partial \eta_i} + \frac{6N\alpha}{\delta^4}\frac{\partial \delta}{\partial \eta_j}\frac{\partial \delta}{\partial \eta_i} - \frac{2n\alpha}{\delta^3}\frac{\partial \delta^2}{\partial \eta_i \partial \eta_j}\end{aligned} \tag{6.31}$$

In Equation 6.30 and Equation 6.31 some notation is introduced. $\eta = \log(\mu)$, $\alpha = \|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{Y}\|^2$, $\delta = N - \text{tr}(\boldsymbol{H})$. The derivatives of α and δ with respect to η can be found in Wood's 2004 paper. The problem that is solved in this way, does not put any constraints on the model parameter estimate $\boldsymbol{c}$. However, the constraints $\boldsymbol{C}\boldsymbol{c} = \boldsymbol{0}$ can be imposed, with $\boldsymbol{C}$ a constraint matrix. Wood does this with a null space projection of $\boldsymbol{C}$ on $\boldsymbol{X}$, $\boldsymbol{S}$ and the $\boldsymbol{P}_i$, as he shows in his paper.

Wood uses penalized iteratively reweighted least squares (P–IRLS) to solve Equation 6.28, a solver similar to the MIS used for B-splines. The smoothing parameter selection can be done inner or outer loop to the iterative solver used. Inner loop means that at every iteration of the solver the smoothing parameters are update using Newton. Outer loop means that the smoothing parameters are updated upon convergence of the solver. In this case one would simply evaluate the GCV score at convergence, calculate the change of the GCV score with respect to the change in smoothing parameters by means of a finite difference scheme, and use a quasi-Newton method to update the parameters [66]. The inner loop mechanism is considerably more efficient and is also referred to as the "performance iteration" method. The outer loop one can be more accurate however. Also, the outer loop method does not require the calculation of the derivatives in Equation 6.30 and Equation 6.31.

There is some limitations to this method. Wood notices in his 2008 paper that his 2004 Newton method can diverge or cycle, most frequently for binary data [67]. If this happens, the 2008 paper provides an improved method. Also, Wood only derives his method for a hat-matrix with zero constraints on the model parameters. This means that the inner-loop method does not allow non-zero equality constraints such as $\boldsymbol{G}\boldsymbol{c} = \boldsymbol{g}$, but the outer loop method does. Lastly, Wood does not specify how many parameters can be realistically estimated with his method, nor does he speak of the effect of the size of data sets on the computational cost.

6.5.3. Fixed Point L-curve Method

Belge et al. introduced a generalization of the L-curve to more dimensions in 1998, allowing the selection of multiple regularization parameters [69]. This curve they called the L-hypersurface. The L-hypersurface is characterized by a corner, akin the L-curve, that represents the point of maximum curvature in the curve, see Figure 6.8. Belge et al. determined this corner by maximizing the Gaussian Curvature. Unfortunately, this approach is computationally problematic. Additionaly, the L-hypersurface bulges of local extrema, making gradient-descent like optimization schemes troublesome. In a 2002 follow up paper the authors propose a new method that is computationally more realistic [70].

In the 2002 paper, they replace the Gaussian curvature with a surrogate function, the Minimum Distance Function (MDF). The MDF is maximized by a fixed point algorithm. The algorithm is initialized by picking a origin point somewhere in the hypersurface space, that spans a dimension of one more than the number of regularization parameters, but not on the surface. More precisely, the point needs to be somewhere below the surface. Then an expanding 'bubble' is formed from the origin. The first point on the L-hypersurface this bubble will touch, is exactly or near the corner.

Belge et al. prove that for a single regularization parameter scenario under some 'mild assumptions' their method will converge to a critical point of the L-hypersurface. They also mention that it is difficult

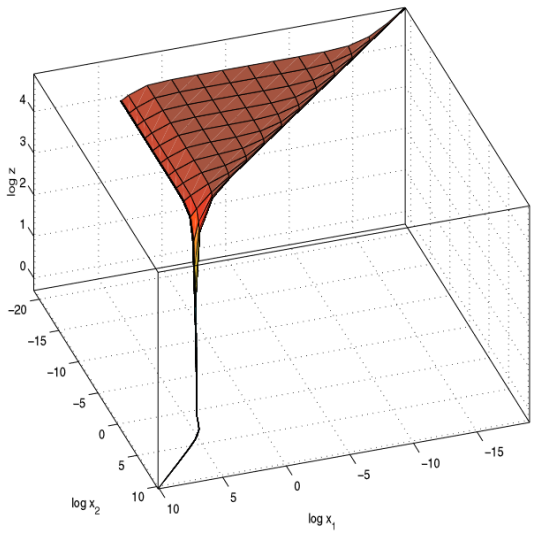


Figure 6.8: A typical L-hypersurface. On the axes in the horizontal plane the logs of the solution semi-norms of two regularization parameters are plotted, on the vertical axis the residual norm. Taken from [69].

to guarantee convergence for the multidimensional case. However, they do mention that in practise the method will converge, if the origin is chosen carefully. Guidelines on selecting an appropriate origin point are provided.

6.6. Conclusions

The class of spectral regularization methods can alter the contribution that each singular value has on the parameter estimates. A good regularization method acts as a high-pass filter on the singular value spectrum. The cut-off point is determined by the regularization parameter. It turns out that available spectral regularization methods practically yield the same solutions. Tikhonov regularization is one of those methods, and it stands out between them because its implementation is rooted in classical linear regression. This makes it an ideal candidate to be implemented in the simplex B-spline framework. Tikhonov regularization essentially adds some additional a-priori information to the problem. For example, the knowledge that the solution to physical systems generally have a small norm and a smooth surface. Furthermore, Tikhonov regularization has been implemented successfully by Eilers and Marx in univariate splines, the P-splines. They have demonstrated that it is even possible to fit a stable spline with less data points than to be estimated parameters. Steiner et al. also implemented Tikhonov regularization in the simplex B-splines. They only found one global regularization parameter. Both methods for finding a single global regularization parameter, and for finding many local parameters are available in literature.

7

Conclusion and Research Plan

This chapter provides conclusions as far as possible after the preliminary thesis phase. In Section 7.1 the current conclusions to the research questions (Section 1.2). Furthermore, a research plan is proposed as well in Section 7.2 that will be leading during the remaining thesis work, in which answers are sought to the unanswered research questions.

7.1. Conclusion

This thesis kicked off with the goal to develop a global aerodynamic model, with the hypothesis that the Holy Grail would involve multivariate simplex B-splines. Some of the research questions that were composed to realize his goal can be (partially) answered from reviewing literature. An answer to the main research questions is not possible at the current stage, but here follows the findings on the sub-questions.

1. What are the advantages and disadvantages of using Multivariate Simplex B-splines for identifying and validating a full flight envelope model from sparse flight-test data of the Cessna Citation II aircraft?

It has been found from the literature that multivariate simplex B-splines is the best global modeling method for developing and validating a full flight envelope model from sparse flight-test-data of the Cessna Citation II aircraft. These splines trump other available global modeling methods such as neural networks, kernel methods and polynomial methods based on a number of characteristics. First of all, simplex B-splines do not suffer from the inherent lack of transparency that characterizes neural networks. Furthermore, neural networks and kernel methods do not have local basis functions and therefore very inefficient model estimation and evaluation characteristics. Then for polynomial models it becomes clear that they are most suitable for local modeling. Combining local polynomial models into a global one is possible, but it is impractical. On the other hand, splines have an inherent continuity over the entire domain. Lastly, the simplex B-splines are also great for model validation: local parameter variance is a great indication of model quality and the stability of the spline surface is easy to inspect by looking at the B-coefficients.

Simplex B-splines have one drawback: the parameter estimation problem tends to be ill-posed when fitting to sparse data. This can be circumvented using Tikhonov regularization. Many Tikhonov penalty matrices are available from literature, such as ridge, finite difference operators and thin-plate energy penalties. Ridge and first order difference penalties have already shown to be effective in splines, other penalty types are promising but need to be worked out. Regularization parameter tuning methods are also widely available, both for local and global parameter tuning.

Whether multivariate simplex B-splines will really be a suitable method for the purpose of full flight envelope modeling still needs to be shown in practise, as they have not yet been (succesfully) applied to stall data sets. However, the perspective is hopeful as the advantages of splines seem to be numerous, and the disadvantages surmountable,

2. What flight test maneuvers are suitable for identifying both steady and unsteady, and both longitudinal and lateral aerodynamics?

For the purpose of developing a full flight envelope longitudinal and lateral model, the *Global* data set was assembled from flight tests with the Cessna Citation II PH-LAB. Based on previous experience test flights accelerated stalls with QR inputs in the stall were recorded. Furthermore, based on the recommendations

by van Horssen et al., van Ingen et al. and Morelli, a dynamic maneuver was recorded before the stall to capture steady dynamics. The stalls were flown at a range of altitudes, resulting in less stalls per flight condition. Whether this test flight design will be sufficient for identifying steady and unsteady, longitudinal and lateral aerodynamics, or that a different flight test design should be proposed, needs to become clear from the remaining research.

3. Which geometric and polynomial model structure can best be used for steady and unsteady aerodynamic modeling with multivariate simplex B-splines?

The geometric model structure selection consists of two parts. Firstly, the model structure can be determined automatically using the MOF algorithm and secondly the triangulation can best be generated using a Kuhn triangulation. The results by the MOF algorithm should be reviewed critically based on engineering insight and validation results. The Kuhn triangulation can also be further optimized using engineering insight. It might be a good idea to split the triangulation at some critical angle of attack, and apply different amounts of regularization and/or physical parameter elimination in both sections. This idea is based on the global model by Wang et al. who merged two polynomial models based on a critical angle of attack. The polynomial model structure selection is to be decided during the next phase.

4. How does a multivariate simplex B-spline model compare to the local stall models by van Ingen et al. [8] and de Fuijk et al. [9]?

In this research question the performance by the to be developed model will be compared to the longitudinal model by van Ingen et al. and the lateral model by de Fuijk et al.. There is nothing that can be said about this at the moment.

5. How can local regularization enable simplex B-spline modelling of sparse and ill-conditioned flight datasets?

Hoerl and Kennard's ridge trace teach us that the optimal amount of regularization is strongly affected by the degree of ill-posedness in a regression problem. It is therefore to be expected that there will be a varying demand for regularization between simplices on ill-conditioned data sets. Additionaly, Wand concluded that splines with multiple smoothing parameters trumped those with only one global parameter.

Global regularization parameter optimization parameters and local regularization parameter optimization are extensively described in literature. The L-curve and GCV methods are the most prominent. Information criteria like AIC and Mallow's C_p also surface. For global methods these can be employed with a simple grid search. For local methods that is not computationally feasible. Newton updates or Fixed point algorithms are more efficient. Also the method from the spatially adaptive P-splines by Ruppert and Carroll is promising. what method is the most suitable to be used in a full flight envelope spline model is to be determined.

Other conclusions

The point of flow separation variable X cannot be measured directly. The four X-parameters a_1, α^*, τ_1 and τ_2 need to be estimated first in a non-linear optimization for X to be reconstructed. The objective function is a function of X through either Kirchoff's description of C_L or C_l. De Fuijk et al. suggest to estimate X through C_L and asymmetric separation variables X_L and X_R through C_l. Fuijk et al. and van Ingen et al. did this from arbitrarily chosen data windows including all phases of stall. Brill et al. introduced a slice based method in which the parameters can be estimated from more optimal data windows. Wang et al. suggested optimal data windows in more general terms. They propose to first estimate a_1 and α^* from pre-stall data, since these parameters are only relevant for describing steady flow separation. He then proposes to estimate τ_1 and τ_2 from unsteady data windows where $\dot{\alpha}$ and $\frac{\mathrm{d}X}{\mathrm{d}t}$ have significant variation, while keeping a_1 and α^* fixed at the previously found values. By splitting the parameter estimation into two, smaller parameter standard deviation can be expected. The method by Wang et al. can potentially yield an increase in performance of the already existing nonlinear optimization framework.

7.2. Research Plan

The research plan for the remaining thesis time is layed out in the flow chart in Figure 7.1. Note that the first step, collecting data and applying FPR, has already been completed and is therefore not shown. Apart from initialization, there is 3 main steps. That is the X-parameter estimation, that is split in a longitudinal and lateral path. Then comes the model structure selection. With the model structure completed, the

dimension of the data set is also know and the final phase can start. This is the spline model estimation. Once the final spline model is finished, it can be compared to the models by de Fuijk and van Ingen to answer sub-question 4.

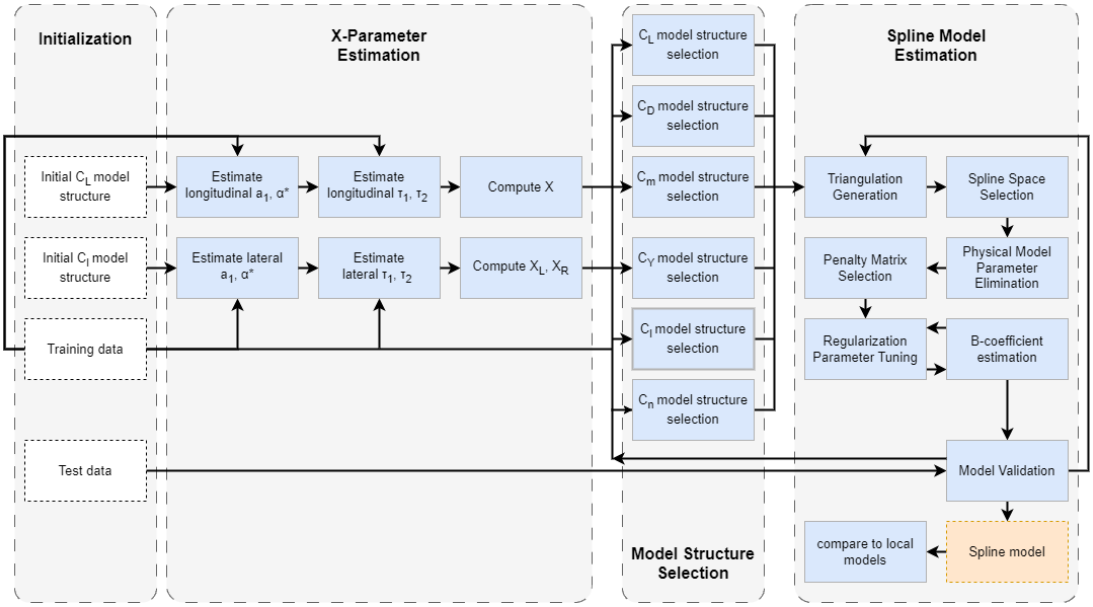


Figure 7.1: Flow chart of project plan for the remainder of the thesis.

Part IV

Closure

Conclusion

The main findings of the research conducted during this thesis can be found in Part II. The conclusions to the research questions in Part I are found below, in addition to recommendations for future work.

Research Questions

Research Question 1

What are the advantages and disadvantages of using Multivariate Simplex B-splines for identifying and validating a full flight envelope pitching moment model from sparse flight-test data of the Cessna Citation II aircraft?

The original hypothesis for this thesis was that Multivariate Simplex B-splines were the best candidate for realising a full flight envelope aerodynamic model. After encountering problems with identification of aerodynamic damping, this hypothesis was put on hold. Based on the research of this thesis, RQ 1 can not be answered.

Yet, it can be noted that splines have the potential of accurately capturing local nonlinear changes in aerodynamic damping. On the other hand, it will be challenging, if not impossible, to capture time varying effects similar to the model presented in this thesis. Namely, it would require the use of lag states of the angle of attack. In the context of Multivariate Simplex B-splines, this will inevitably give rise to numerical issues.

Research Question 2

What flight test maneuvers are suitable for identifying both steady and unsteady pitching moment aerodynamics?

The flight test maneuvers used in this research are suitable in the sense that each results in a sequence of stall cycles, where the aircraft oscillates in and out of stall, every time with a different angle of attack range and stall degree. This is important since those criteria strongly affect the observed aerodynamic loads, meaning a varied stall data base is key. However, the random input design seemingly is not as effective as desired. There appears to be no causal relationship between the elevator wiggles on the one hand, and the angle of attack or pitch rate time traces on the other hand. The observed time traces of the flow states rather seem to be the result of bias applied to the control column, which serves the purpose of keeping the aircraft in the stall. A logical explanation of this is that the input frequency is too high, meaning the aircraft is not granted enough time to respond to the input, given its inertia. Consequently, the quasi-random inputs provide only little excitation. An alternative and better suited input type could be orthogonal phase-optimized multisines. Such inputs are in essence a sum of sinusoidal inputs, with frequencies in the range of the eigenfrequencies of the test aircraft, such that both the short period and phugoid eigenmodes can be excited simultaneously. If this indeed allows data collection with higher information content, the current model can be improved by including higher order regressors, or the parameter standard deviations can be reduced, yielding more confidence in the estimated parameters.

Furthermore, it was found that a data set of combined 1.4g accelerated and wings-level stalls is suitable for identifying both steady and unsteady contributions to the pitch moment model. The accelerated stalls provide a decorrelation between pitch and angle of attack rate, while the wings-level maneuvers can reach higher angles of attack.

Research Question 3

Which model structure can best be used for steady and unsteady aerodynamic modeling of the pitching moment?

it was found that the conventional control and stability derivative model structure of the nominal flight envelope can be retained, but it should be augmented with correction terms to account for nonlinear and unsteady phenomena. These correction terms include lag states of the angle of attack to capture time varying effects caused by the simultaneous loss of wing lift and moving center of pressure, and univariate splines to model changes in pitch and downwash lag damping at high angles of attack. Additionally, a thrust coefficient parameter and a correction on the elevator effectiveness were found to be beneficial for model fit.

This yields a model with high physical interpretability, because the nominal parameters directly reflect the linear steady aerodynamics, while the correction terms do the same for nonlinear and unsteady aerodynamics. For the field of stall modeling, this finding means an alternative to the Kirchoff based model approach, such that the nonlinear optimization required to reconstruct Kirchoff's flow separation parameter X can be circumvented, and that model parameters found through simple linear analysis can be retained in the model structure.

Research Question 4

How does a global model compare to the local stall model by van Ingen et al. [8]?

Overall, the global model presented in the paper in Part II yields a reduction of 55.9% in validation MSE. This is a significant improvement, that is largely due to its global modeling capacities. A large gain stems from data that resides outside the stall regime. However, the new model also outperforms the model by van Ingen et al. on data segments inside the stall regime, although to a smaller extent.

Additionally, the global model approach has a much higher degree of physical interpretability. The Kirchoff based method is phenomenological in nature, whereas the control and stability method is based on the underlying physics that drives the generation of the data. As a result, a far more profound understanding of the global model and its mechanisms can be acquired, which is vital for verification and validation. For example, the three lag states of the angle of attack can be physically explained by the oscillatory nature of the identification data. The frequency of these oscillations is directly reflected in the model behaviour, which can be compared to examples from wind tunnel tests and fundamental aerodynamic research on stall dynamics and vortex formation.

Research Question 5

What modifications to Kirchoff's flow separation based model are necessary to enable the effective modeling of aerodynamic damping?

The aerodynamic damping can be modeled by using the pitch damping from the nominal flight model as a baseline. For global modeling capacity of the Cessna Citation II, it should be accompanied by an additional pitch damping term that is activated at angles of attack above 5° and a downwash lag damping term for angles of attack over 8°. This structure means that during high angle of attack maneuvers the aerodynamic damping is altered, while it is restored to the nominal at low angles of attack, guaranteeing dynamic stability of the model.

The main modification to Kirchoff based models necessary to accommodate the introduction of aerodynamic damping model terms, is simply the removal of X related contributions. To capture time varying phenomena, lag states of the angle of attack can be employed. It is difficult for Kirchoff and spline based damping coefficients to coexist, because the former has little physical meaning concerning the pitching moment, while the latter does. The X variable is reconstructed through the lift coefficient, and therefore captures a large hysteresis component of the lift. Consequently, Kirchoff model terms inherently predict hysteresis, also in

the pitching moment. However, the hysteresis of the pitching moment is mostly due to the aerodynamic damping. Hence, it is likely that the identification of the two interfere.

Recommendations

The following list of recommendations are proposed:

- Increasing the model order of the candidate regressors using regularization techniques. During the current research it was found that increasing model order leads to excessive parameter standard deviation, likely due to lack of information content in the data. Regularization might resolve the issue.
- A flight test experiment design using orthogonal phase-optimized multisines. As explained in the conclusion to RQ 2, such an experiment has the potential of improving the information content in the data.
- Using the method presented in Part II to identify a complete 6 degrees of freedom full envelope flight model.

References

[1] United States. National Transportation Safety Board. Aircraft Accident Report : Loss of Control on Approach Colgan Air, Inc. Operating as Continental Connection Flight 3407 Bombardier DHC-8-400, N200WQ Clarence Center, New York February 12, 2009, for sale by the National Technical Information Service, 2010.

[2] IATA. Loss of control in-flight accident analysis report, 2019.

[3] E. Morelli, K. Cunningham, and M. Hill. “Global Aerodynamic Modeling for Stall/Upset Recovery Training Using Efficient Piloted Flight Test Techniques”. In: Aug. 2013. DOI: 10.2514/6.2013-4976.

[4] Federal Aviation Administration. Advisory Circular 120-109A - Stall Prevention and Recovery Training.Tech. rep. 2015. URL: https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/1028646.

[5] EASA. Certification specifications for aeroplane flight simulation training devices ’csfstd(a)’, 2018.

[6] S. Advani and J. Field. “Upset Prevention and Recovery Training in Flight Simulators”. In: Portland, US, Aug. 2011. DOI: 10.2514/6.2011-6698.

[7] L. Van Horssen, C. C. De Visser, and D. M. Pool. “Aerodynamic Stall and Buffet Modeling for the Cessna Citation II Based on Flight Test Data”. en. In: *2018 AIAA Modeling and Simulation Technologies Conference*. Kissimmee, Florida: American Institute of Aeronautics and Astronautics, 2018. DOI: 10.2514/6.2018-1167. URL: https://arc.aiaa.org/doi/10.2514/6.2018-1167 (visited on 11/01/2023).

[8] J. van Ingen, C. C. de Visser, and D. M. Pool. “Stall Model Identification of a Cessna Citation II from Flight Test Data Using Orthogonal Model Structure Selection”. In: *AIAA Scitech 2021 Forum*. American Institute of Aeronautics and Astronautics, 2021. URL: https://arc.aiaa.org/doi/abs/10.2514/6.2021-1725 (visited on 10/26/2023).

[9] D. de Fuijk, D. Pool (mentor), and C. De Visser (mentor). “Asymmetric Cessna Citation II Stall Model Identification using a Roll moment-based Kirchhoff method”. MA thesis. Delft University of Technology, 2023.

[10] C. C. de Visser and D. M. Pool. “Stalls and Splines: Current Trends in Flight Testing and Aerodynamic Model Identification”. en. In: *Journal of Aircraft* (Aug. 2023), pp. 1–23. DOI: 10.2514/1.C037283. URL: https://arc.aiaa.org/doi/10.2514/1.C037283 (visited on 09/07/2023).

[11] C. C. de Visser. “Global Nonlinear Model Identification with Multivariate Splines”. doctoral thesis. Delft University of Technology, 2011. URL: https://repository.tudelft.nl/islandora/object/uuid:6bc0134a-0715-4829-903d-6479c5735913?collection=research.

[12] L. Steiner and C. de Visser (mentor). “Towards a Multivariate Spline-Based qLPV Flight Dynamics Model of Tiltrotor Aircraft”. en. MA thesis. Delft University of Technology, 2023. URL: https://repository.tudelft.nl/islandora/object/uuid%3Acda96de4-4ed5-4b07-9478-a18cefa968f0 (visited on 09/08/2023).

[13] J. D. Anderson. *Fundamentals of aerodynamics*. sixth edition. New York: McGraw-Hill Education, 2017.

[14] M. Goman and A. Khrabrov. “State-space representation of aerodynamic characteristics of an aircraft at high angles of attack”. en. In: *Journal of Aircraft* (May 2012). DOI: 10.2514/3.46618. URL: https://arc.aiaa.org/doi/10.2514/3.46618 (visited on 11/01/2023).

[15] D. Fischenberg. “Identification of an unsteady aerodynamic stall model from flight test data”. In: *20th Atmospheric Flight Mechanics Conference*. Guidance, Navigation, and Control and Co-located

Conferences. American Institute of Aeronautics and Astronautics, 1995. DOI: 10.2514/6.1995-3438. URL: https://arc.aiaa.org/doi/10.2514/6.1995-3438 (visited on 10/26/2023).

[16] "aerodynamics of flight". In: *Pilot's Handbook of Aeronautical Knowledge*. FAA-H-8083-25C. Federal Aviation Administration, 2023. URL: https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/phak.

[17] L. E. Ericsson and J. P. Reding. "Fluid mechanics of dynamic stall part I. Unsteady flow concepts". In: *Journal of Fluids and Structures* 2.1 (Jan. 1988), pp. 1–33. DOI: 10.1016/S0889-9746(88)90116-8. URL: https://www.sciencedirect.com/science/article/pii/S0889974688901168 (visited on 11/03/2023).

[18] D. R. Gingras, J. N. Ralston, R. Oltman, C. Wilkening, R. Watts, and P. Derochers. "Flight Simulator Augmentation for Stall and Upset Training". In: *AIAA Modeling and Simulation Technologies Conference*. AIAA SciTech Forum. American Institute of Aeronautics and Astronautics, Jan. 2014. DOI: 10.2514/6.2014-1003. URL: https://arc.aiaa.org/doi/10.2514/6.2014-1003 (visited on 11/03/2023).

[19] V. Klein and E. Morelli. *Aircraft System Identification: Theory and Practice*. Williamsburg, Virginia, July 2016.

[20] P. A. R. Brill, D. M. Pool (mentor), and C. de Visser (mentor). "Improving Stall Model Accuracy through Optimal Data Slicing by Analyzing Kirchhoff Stall Parameter Estimate Behaviour". MA thesis. Delft University of Technology, 2023.

[21] A. Ly, M. Marsman, J. Verhagen, R. P. P. P. Grasman, and E.-J. Wagenmakers. "A Tutorial on Fisher information". In: *Journal of Mathematical Psychology* 80 (Oct. 2017), pp. 40–55. DOI: 10.1016/j.jmp.2017.05.006. URL: https://www.sciencedirect.com/science/article/pii/S0022249617301396.

[22] Federal Aviation Administration. 14 CFR Part 60 - Flight Simulation Training Device Initial and Continuing Qualification and Use. Tech. rep. 2016. URL: https://www.ecfr.gov/current/title14/chapter-I/subchapter-D/part-60.

[23] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D (1960), pp. 35–45.

[24] E. Wan and R. Van Der Merwe. "The unscented Kalman filter for nonlinear estimation". In: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*. Oct. 2000, pp. 153–158. DOI: 10.1109/ASSPCC.2000.882463. URL: https://ieeexplore.ieee.org/document/882463 (visited on 01/28/2024).

[25] L. van Horssen, D. Pool (mentor), and C. de Visser (mentor). "Aerodynamic Stall Modeling for the Cessna Citation II". en. MA thesis. Delft University of Technology, 2016.

[26] K. Hornik, M. Stinchcombe, and H. White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (Jan. 1989), pp. 359–366. DOI: 10.1016/0893-6080(89)90020-8. URL: https://www.sciencedirect.com/science/article/pii/0893608089900208.

[27] C. de Boor. "B-form basics". In: *Geometric modeling: Algorithms and new trends* (1987). Edited by G. Farin, pp. 131–148.

[28] P. Dierckx. "Curve and surface fitting with splines". In: *Monographs on numerical analysis*. 1994. URL: https://api.semanticscholar.org/CorpusID:122918752.

[29] P. H. C. Eilers and B. D. Marx. "Flexible Smoothing with B-splines and Penalties". In: *Statistical Science* 11.2 (1996). Publisher: Institute of Mathematical Statistics, pp. 89–102. URL: https://www.jstor.org/stable/2246049 (visited on 09/20/2023).

[30] C. C. de Visser, Q. P. Chu, and J. A. Mulder. "A new approach to linear regression with multivariate splines". In: *Automatica* 45.12 (Dec. 2009), pp. 2903–2909. DOI: 10.1016/j.automatica.2009.09.017. URL: https://www.sciencedirect.com/science/article/pii/S0005109809004270.

[31] H. W. Kuhn. “Some Combinatorial Lemmas in Topology”. In: *IBM Journal of Research and Development* 4.5 (Nov. 1960). Conference Name: IBM Journal of Research and Development, pp. 518–524. DOI: 10.1147/rd.45.0518. URL: https://ieeexplore.ieee.org/document/5392466?denied= (visited on 11/16/2023).

[32] J. Bey. “Simplicial grid refinement: on Freudenthal’s algorithm and the optimal number of congruence classes”. en. In: *Numerische Mathematik* 85.1 (Mar. 2000), pp. 1–29. DOI: 10.1007/s002110050475. URL: https://doi.org/10.1007/s002110050475 (visited on 11/16/2023).

[33] F. Boudaoud, F. Caruso, and M.-F. Roy. “Certificates of Positivity in the Bernstein Basis”. In: *Discrete and Computational Geometry* 39 (June 2008). DOI: 10.1007/s00454-007-9042-x.

[34] C. C. de Visser, Q. P. Chu, and J. A. Mulder. “Differential constraints for bounded recursive identification with multivariate splines”. In: *Automatica* 47.9 (Sept. 2011), pp. 2059–2066. DOI: 10.1016/j.automatica.2011.06.011. URL: https://www.sciencedirect.com/science/article/pii/S0005109811003104 (visited on 09/20/2023).

[35] G. M. Awanou and M. J. Lai. “On convergence rate of the augmented Lagrangian algorithm for nonsymmetric saddle point problems”. In: *Applied Numerical Mathematics*. 6th IMACS 54.2 (July 2005), pp. 122–134. DOI: 10.1016/j.apnum.2004.09.020. URL: https://www.sciencedirect.com/science/article/pii/S0168927404001771 (visited on 11/16/2023).

[36] G. Awanou, M. Lai, and P. Wenston. “The Multivariate Spline Method for Scattered Data Fitting and Numerical Solutions of Partial Differential Equations”. In: Athens, 2005, pp. 24–75. URL: https://www.semanticscholar.org/paper/The-Multivariate-Spline-Method-for-Scattered-Data-Awanou-Lai/29aefe70f085612254eb2b7d4e77562e8dc02757 (visited on 11/16/2023).

[37] F. Huisman and C. C. de Visser (mentor). “Full Flight Envelope Aerodynamic Modelling of the Cessna Citation II using Physical Splines”. MA thesis. Delft University of Technology, 2017. URL: https://repository.tudelft.nl/islandora/object/uuid%3Af9926c24-e86b-4eee-8986-d427fb4bc667.

[38] K. Q. Weinberger. *Machine Learning for Intelligent Systems Lecture 12: Bias-Variance Tradeoff*. Lecture notes. Cornell Univerisity, 2018.

[39] J. Hadamard. “Sur les problemes aux derivees partielles et leur signification physique”. fr. In: *Princeton University Bulletin* (1902), pp. 49–52.

[40] G. H. Golub and C. F. V. Loan. *Matrix Computations*. English. 3rd edition. Baltimore: Johns Hopkins University Press, Oct. 1996.

[41] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Linear Regression Analysis*. 5. Hoboken, NJ: Wiley, 2012.

[42] A. Tikhonov and V. Arsenin. *Solutions of Ill-posed Problems*. New York: Wiley, 1977.

[43] P. C. Hansen. “Analysis of Discrete Ill-Posed Problems by Means of the L-Curve”. In: *SIAM Review* 34.4 (1992). Publisher: Society for Industrial and Applied Mathematics, pp. 561–580. URL: https://www.jstor.org/stable/2132628 (visited on 09/27/2023).

[44] A. E. Hoerl and R. W. Kennard. “Ridge Regression: Biased Estimation for Nonorthogonal Problems”. In: *Technometrics* 12.1 (1970). Publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], pp. 55–67. DOI: 10.2307/1267351. URL: https://www.jstor.org/stable/1267351 (visited on 09/19/2023).

[45] P. C. Hansen. “The L-curve and its use in the numerical treatment of inverse problems”. en. In: *Computational Inverse Problems in Electrocardiology*. Vol. 4. WIT Press, 2001, pp. 119–142.

[46] C. H. Reinsch. “Smoothing by spline functions”. In: *Numerische Mathematik* 10.3 (Oct. 1967), pp. 177–183. DOI: 10.1007/BF02162161. URL: http://link.springer.com/10.1007/BF02162161.

[47] F. O’Sullivan. “A Statistical Perspective on Ill-Posed Inverse Problems”. In: *Statistical Science* 1.4 (1986). Publisher: Institute of Mathematical Statistics, pp. 502–518. URL: https://www.jstor.org/stable/2245801 (visited on 09/28/2023).

[48] S. N. Wood. “P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data”. en. In: *Statistics and Computing* 27.4 (July 2017), pp. 985–989. DOI: 10.1007/s11222-016-9666-x. URL: https://doi.org/10.1007/s11222-016-9666-x (visited on 10/04/2023).

[49] P. H. C. Eilers and B. D. Marx. “Splines, knots, and penalties”. en. In: *WIREs Computational Statistics* 2.6 (2010). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wics.125, pp. 637–653. DOI: 10.1002/wics.125. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.125 (visited on 09/27/2023).

[50] P. H. C. Eilers and B. D. Marx. “Multivariate calibration with temperature interaction using two-dimensional penalized signal regression”. In: *Chemometrics and Intelligent Laboratory Systems* 66.2 (June 2003), pp. 159–174. DOI: 10.1016/S0169-7439(03)00029-7. URL: https://www.sciencedirect.com/science/article/pii/S0169743903000297 (visited on 10/10/2023).

[51] P. Eilers, B. Marx, and M. Durbán. “Twenty years of P-splines”. In: *SORT (Statistics and Operations Research Transactions)* 39 (Jan. 2015), pp. 149–186.

[52] P. H. C. Eilers. “Uncommon penalties for common problems”. en. In: *Journal of Chemometrics* 31.4 (Apr. 2017), e2878. DOI: 10.1002/cem.2878. URL: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.2878 (visited on 10/09/2023).

[53] D. Ruppert and R. J. Carroll. “Theory & Methods: Spatially-adaptive Penalties for Spline Fitting”. en. In: *Australian & New Zealand Journal of Statistics* 42.2 (2000). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-842X.00119, pp. 205–223. DOI: 10.1111/1467-842X.00119. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-842X.00119 (visited on 10/03/2023).

[54] I. Currie and M. Durbán. “Flexible smoothing with P-splines: A unified approach”. In: *Statistical Modelling* 2 (Dec. 2002), pp. 333–349. DOI: 10.1191/1471082x02st039ob.

[55] M. P. Wand. “A Comparison of Regression Spline Smoothing Procedures”. en. In: *Computational Statistics* 15.4 (Dec. 2000), pp. 443–462. DOI: 10.1007/s001800000047. URL: https://doi.org/10.1007/s001800000047 (visited on 11/15/2023).

[56] M. P. Wand and J. T. Ormerod. “On semiparametric regression with O’Sullivan penalized splines”. In: *Australian & New Zealand Journal of Statistics* 50.2 (2008). arXiv:0707.0143 [stat], pp. 179–198. DOI: https://doi.org/10.1111/j.1467-842X.2008.00507.x.

[57] R. Pfeifle and H.-P. Seidel. “Fitting Triangular B-Splines to Functional Scattered Data”. en. In: *Computer Graphics Forum* 15.1 (Feb. 1996), pp. 15–23. DOI: 10.1111/1467-8659.1510015. URL: https://onlinelibrary.wiley.com/doi/10.1111/1467-8659.1510015 (visited on 09/14/2023).

[58] S. N. Wood. “Thin Plate Regression Splines”. In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 65.1 (2003). Publisher: [Royal Statistical Society, Wiley], pp. 95–114. URL: https://www.jstor.org/stable/3088828 (visited on 10/16/2023).

[59] C. L. Mallows. “Some comments on CP”. In: *Technometrics* 15.4 (1972), pp. 661–675. DOI: 10.2307/1267380. URL: https://www.jstor.org/stable/1267380.

[60] T. H. Edwards and S. Stoll. “Optimal Tikhonov regularization for DEER spectroscopy”. In: *Journal of Magnetic Resonance* 288 (Mar. 2018), pp. 58–68. DOI: 10.1016/j.jmr.2018.01.021. URL: https://www.sciencedirect.com/science/article/pii/S1090780718300442 (visited on 10/03/2023).

[61] G. H. Golub, M. Heath, and G. Wahba. “Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter”. In: *Technometrics* 21.2 (1979). Publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], pp. 215–223. DOI: 10.2307/1268518. URL: https://www.jstor.org/stable/1268518 (visited on 10/17/2023).

[62] K. Kagerer. “A hat matrix for monotonicity constrained B-spline and P-spline regression”. In: Mar. 2015. URL: https://www.semanticscholar.org/paper/A-hat-matrix-for-monotonicity-constrained-B-spline-Kagerer/522163d4c3ed72ee3c8c5934368a9312e871f75d (visited on 09/27/2023).

[63] G. Frasso and P. H. Eilers. “L- and V-curves for optimal smoothing”. In: *Statistical Modelling* 15.1 (Feb. 2015). Publisher: SAGE Publications India, pp. 91–111. DOI: 10.1177/1471082X14549288. URL: https://doi.org/10.1177/1471082X14549288 (visited on 09/27/2023).

[64] C. Gu and G. Wahba. “Minimizing GCV/GML Scores with Multiple Smoothing Parameters via the Newton Method”. en. In: *SIAM Journal on Scientific and Statistical Computing* 12.2 (Mar. 1991), pp. 383–398. DOI: 10.1137/0912021. URL: http://epubs.siam.org/doi/10.1137/0912021 (visited on 10/12/2023).

[65] S. N. Wood. “Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties”. In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 62.2 (2000). Publisher: [Royal Statistical Society, Wiley], pp. 413–428. URL: https://www.jstor.org/stable/3088868 (visited on 10/12/2023).

[66] S. N. Wood. “Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models”. In: *Journal of the American Statistical Association* 99.467 (Sept. 2004). Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/016214504000000980, pp. 673–686. DOI: 10.1198/016214504000000980. URL: https://doi.org/10.1198/016214504000000980 (visited on 10/12/2023).

[67] S. Wood. “Fast Stable Direct Fitting and Smoothness Selection for Generalized Additive Models”. In: *Journal of the Royal Statistical Society Series B* 70 (July 2008), pp. 495–518. DOI: 10.1111/j.1467-9868.2007.00646.x.

[68] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. San Diego: Academic Press, INC., 1981.

[69] M. Belge, M. Kilmer, and E. Miller. “Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet transform domain”. In: *Proceedings of SPIE - The International Society for Optical Engineering* 3459 (Aug. 1998). DOI: 10.1117/12.323812.

[70] M. Belge, M. Kilmer, and E. Miller. “Efficient determination of multiple regularization parameters in a generalized L-curve framework”. English. In: *Inverse Problems* 18.4 (2002), pp. 1161–1183. DOI: 10.1088/0266-5611/18/4/314.