Data-driven Boundary Layer Loss Model for Organic Rankine Cycle Turbomachines

Technische Universiteit Delft

K. Johri

Data-driven Boundary Layer Loss Model for Organic Rankine Cycle Turbomachines

by

K. Johri

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday November 26, 2020 at 09:30 AM.

Student number:	4363051	
Project duration:	July, 2019 - November, 2020	
Supervisors:	Dr. Ir. M. Pini,	TU Delft
	Ir. F. Tosto,	TU Delft
Thesis Committee:	Prof. Dr. Ir. P. Colonna	TU Delft
	Dr. Ir. M. Pini,	TU Delft
	Dr. Ir. A. H. van Zuijlen,	TU Delft
	Ir. F. Tosto,	TU Delft

Cover Image Courtesy: GE Reports [1]

Acknowledgement

My journey of working on my Master's thesis has been long, tumultuous as well as rewarding. I would like to thank the following people for their support and guidance.

My daily supervisor Ir. Francesco Tosto, whose room door (and Slack) was always open for all my doubts and queries regarding my research. His invaluable insights steered me in the right direction and inspired me to hone my critical thinking and enhance the quality of my work.

My supervisor Dr Ir Matteo Pini for firstly bringing the entire Turbomachinery Group closer by promoting open communication and for consistently pushing us all to challenge ourselves as well as the future.

Simulating several thousands of blades seemed easy at first but the learning curve has been steep. I am glad I worked on this challenging project which helped me push my boundaries. Working from home during the pandemic led to phases of demotivation and frustration but the unwavering support and guidance offered by my parents, sister and BIL kept me going.

I would like to thank all my friends in Delft, especially Ankit, Marta, Hiera for being there to make my student life in university a memorable and fun experience. A special thanks to Seb and Claudia who have been my family for the past five years. Lastly, a shout out to Devyani, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research.

K. Johri
Delft, November 2020

Abstract

There is a requirement for localised efficient electricity generation systems that increase the efficiency of power plants and reduce wasted heat from other engineering applications such as cement manufacturing units and brick kilns. Organic Rankine Cycle (ORC) power systems use organic fluids and low-grade heat sources to accomplish this. Lack of preliminary loss models for turbomachines system components, such as turbine and compressor blades, operating in organic fluids prevents the estimation of losses at the preliminary design phase. The design procedure of ORC power system components relies on correlations developed for gas turbines operating with air or steam. The physics-based loss model of Denton is based on assumptions and empirical data to estimate losses [2]. One drawback of this model is the constant dissipation coefficient ($C_{d,blade}$) value of 0.002. The estimation of $C_{d,blade}$ requires parameters of the boundary layer that are not available at the preliminary design phase. There is thus a requirement for a model to estimate $C_{d,blade}$ as a function of preliminary design parameters such as duty coefficients, volumetric expansion ratio, fluid, operating thermodynamic conditions and blade geometrical parameters. The objective of the present work is to develop a loss model capable of predicting $C_{d,blade}$ based on a database of results from numerical Reynolds-Averaged Navier-Stokes equation simulations for 2D axial stator blades. The developed model is a machine learnt model that is then implemented in an in-house meanline design loss estimation tool, TurboSim, to estimate changes from the current implementation in profile losses and stage efficiency losses. Results from a parametric analysis show that the duty coefficients and working fluid have the highest influence on $C_{d,blade}$. The implementation of the machine-learnt model in TurboSim shows that using a constant $C_{d,blade} = 0.002$ leads to a slight underestimation of the profile loss for all cases with the exception of complex molecules operating close to the critical conditions. Although the lack of validation data prevents the estimation of the accuracy of the findings, the results highlight two aspects. First, the need to change the existing assumption of $C_{d,blade} = 0.002$ for ORC applications. Second, the potential application of data-driven models for loss estimation.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

CFD	Computational Fluid Dynamics
D6	Dodecamethylcyclohexasiloxane
MDM	Octamethyltrisiloxane
NICFD	Non-Ideal Compressible Fluid Dyanamics
ORC	Organic Rankine Cycle
RANS	Reynolds Averaged Navier Stokes

Greek Symbols

α	Flow angle from axial direction	[m]
α_{tt}	Volumetric expansion ratio	[-]
β	Pressure expansion ratio	[-]
η	Efficiency	[-]
Γ	Fundamental derivative of gas dynamic	[-]
γ	Specific heat ratio	[-]
γ_{pv}	Polytropic exponent	[-]
ρ	Density	[-]
ρ	Density	[kg/m^3]
τ	Shear	[kg/m^3]
θ	Parameter vector	
θ_n	Feature weight	[-]
ζ	Entropy loss coefficient	[J/kg]

Latin characters

$\dot{S}$	Rate of entropy generation	[J/K/s]
c	Speed of sound	[m/s]
C_d	Dissipation coefficient	[-]
C_p	Heat capacity at constant pressure	[J/kg/K]
C_v	Heat capacity at constant voluem	[J/kg/K]

h	Static enthalpy	[J/kg]
k	Constant	[-]
P	Pressure	[bar]
P_r	Reduced pressure	[-]
q	Specific heat	[J/kg]
R_{gas}	Specific gas constant	[J/ kg/K]
s	Specifc entropy	[J/kg/K]
s	Specific entropy	[J/K/kg]
T	Temperature	[K]
T_r	Reduced temperature	[-]
V	Velocity	[m/s]
v	Volume	[kg/m^3]
w	Specific work	[J/kg]
Lasso	Least Absolute Shrinkage and Selection Operator Regression	
MSE	Mean Square Error	
RMSE	Root Mean Square Error	
SVR	Support Vector Regression	
Z	Compressibility factor	[-]

Subscripts

δ_e	Edge of the boundary layer	
$\hat{y}$	Predicted value	[-]
$\hat{y}$	Predicted value	[-]
a	Area	
$comp$	Compressor	
$crit$	Critical conditions	
h_θ	Hypothesis function	[-]
$ideal$	Ideal gas conditions	
$irrev$	Irreversible	
o	Total condition	
rev	Reversible	

s	Static condition	
tt	Total-to-total	
$turb$	Turbine	
x	x plane	
x_n	Feature	[-]
yx	y-x plane	
$\mathbf{x}$	Feature vector	[-]

1

Introduction

According to the projections made by the U.S. Energy Information Administration the global energy consumption is expected to increase by almost 50% by the year 2050 assuming economic, infrastructural and technological growth continues its course of the recent decades [3]. A similar projection has been made by the International Energy Agency as well [4]. Historically, this energy demand has been fulfilled by energy systems that use either coal, oil or gas. However, with the current state of CO_2 emissions and greenhouse gases - the main drivers of climate change, there is a strong push towards usage of renewable sources of energy and development of new sustainable technologies to meet the future energy demands. Figure 1.1 show the growing trend towards renewable sources over the past few years.

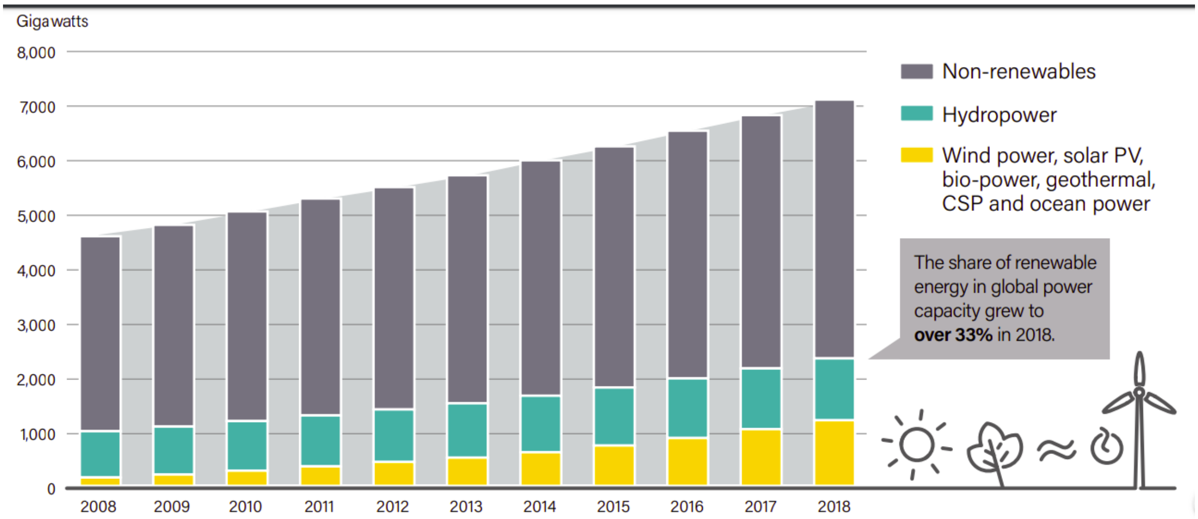


Figure 1.1: Global power generating capacity, by source, from 2008 to 2018 [5].

As of 2018, the share of installed global power generation capacity from renewable sources of energy was 33%. Of this, hydro, solar and wind together contributed almost 90%, whereas biomass, geothermal and ocean power contributed the remaining 10%. While solar, wind and hydro power remain the dominant contributors, there lies potential in the other sources such as geothermal reservoirs, biomass combustion and waste heat recovery from industrial processes. Contrary to the traditional setup of centralised power generation, these sources can be used to both stand-alone plants as well as a grid of small-to-medium scale power plants. The advantage is that it allows one to reduce expenditures on retail power and sometimes even sell power back to the grid. The downside is that these sources provide limited thermal power (from 10kW to 70MW) [6]. As a result it is very important to use efficient technologies that can convert thermal power into electricity.

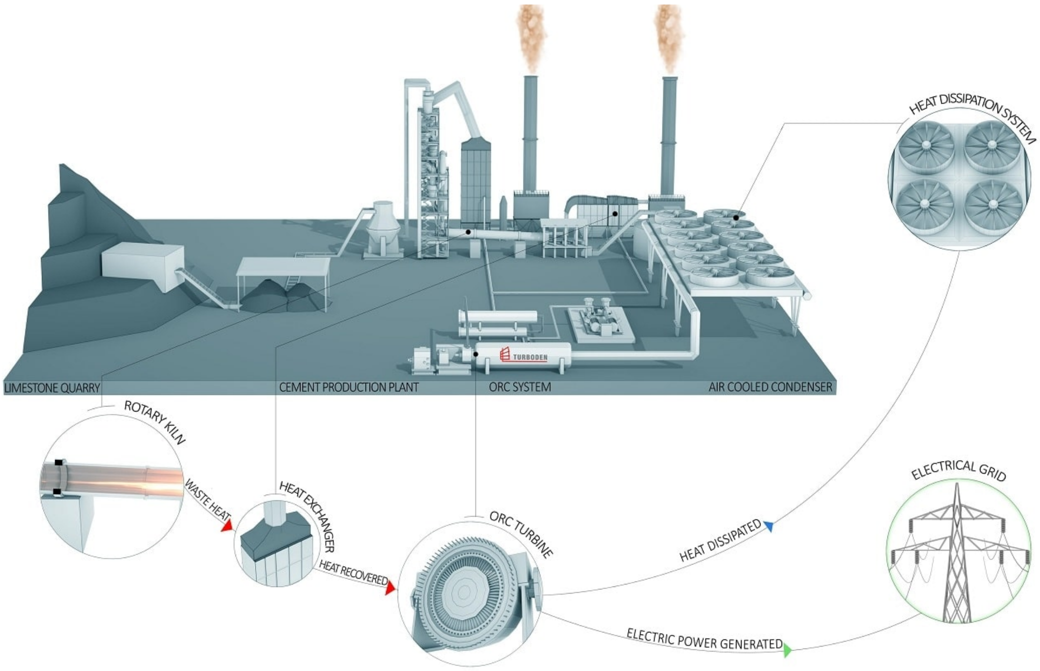


Figure 1.2: The setup of an ORC turbine in a cement production plant. The wasted heat recovered from this plant is used by the turbine to generate electricity [10].

One such technology is the Organic Rankine Cycle (ORC). Figure 1.2 shows a localised ORC power system application in a cement production plant. Besides generating electricity the heat dissipation system can be used to increase the overall efficiency of the power system. An ORC power system makes use of the same principle as that of the Rankine Cycle but uses an organic fluid. An organic fluid is characterised by higher molecular mass than water and generally has a boiling point temperature lower than water so that low grade heat can be recovered with minimum losses. Several fluids such as hydrocarbons (forms of alkanes and alkynes), refrigerants or siloxanes meet these general requirements. Therefore, one of the first challenges faced when designing an ORC, is selecting an appropriate working fluid. Additionally, unlike conventional gas turbines there is no single fluid that works with all ORC power systems. Literature shows that different working fluids are optimal for different applications. For instance, Honeywell lists the 19 different refrigerants that they use based on the application [7]. Research performed by Drescher and Brüggemann shows that alkylbenzenes provide the highest efficiencies for small sized biomass power and heat plants [8]. The selection of the most appropriate fluid is based on different criteria like critical temperatures and pressures, molecular complexity, suitable molecular mass, environmental impact, hazardousness, availability and costs [9].

Eventually, efficient electricity generation has relied and continues to rely on enhanced turbomachinery designs to extract power from the fluid in use. The traditional turbomachinery engineering design practices developed, for gas turbines, over the past decades are now being applied to develop turbomachinery components of ORC power systems. Organic fluids pose a challenge to the design of turbomachines in use ,namely, compressors, turbines, fans or pumps. Additionally, organic fluids are a relatively newer field of study and fall into the branch of fluid mechanics called Non-Ideal Compressible Fluid Dynamics (NICFD). Consequently, there is a shortage of reliable empirical loss correlations and efficiency prediction methods that ease the preliminary design phase of ORC turbines, making the design process of these turbines more challenging.

The design guidelines used for modern day gas turbines have been ever-evolving since 1940s. These guidelines are based on experiments performed in laboratories and are in the form of correlational diagrams, performance charts and loss models. Some examples are Balje and Cordier diagram, Smith chart, Ainley and Mathieson, Soderberg and Traupel loss models [11][12]. Given recent advances in computational optimisation techniques and Computational Fluid Dynamics(CFD) one might question the need for preliminary design at all. However, preliminary design choices dictate decisions such as the machine layout, size based on number of stages and coefficients (such as flow and work), and preliminary estimations of losses. A good preliminary design narrows down the design space and reduces the computational load and time, that would otherwise be required for the detailed design phase. Most importantly, it gives an insight into the loss breakdown of the turbomachinery component design per category, namely, profile, endwall and tip leakage. The aforementioned correlations and models are used to develop preliminary designs for ORC power system components but are inaccurate. The flow behaviour of organic fluids in ORC power systems deviates from conventional fluids like air or steam and thus alters the dissipation from different loss sources.

Knowledge Gap

Recent research performed by Pini and Giuffre led to the development of a loss model for ORC power systems that resorts to physics models and real gas thermodynamics to estimate profile, endwall and tip leakage losses [13][14]. Profile losses are further segregated into boundary layer, shock-wave and mixing losses. The model was used to draw guidelines and performance maps for the conceptual design phase for axial machines and provide a physical understanding of the loss mechanisms. The two main outcomes of this study are that the optimal duty coefficients are dependent on the the fluid and the turbine expansion ratio. The two influence dissipation due to mixing and shock losses respectively. Though the model is verified with 3D Reynolds-averaged Navier-Stokes, certain assumptions used in the development of the model require further research. A fixed value of dimensionless dissipation coefficient (C_d) equal to 0.002 is one such example. C_d is a non-dimensional measure of rate of entropy generation in a boundary layer and is expressed as a function of the momentum thickness Reynold's number (Re_θ) by Schlichting [15]. As most turbomachines operate in order of Re_θ 1000, Denton suggests 0.002 can be used for most turbomachine blades [2]. However, recent research has revealed that fluid molecular complexity might influence the value of C_d [16] [17].

In light of all the considerations mentioned, there is a need to further explore the modelling of loss mechanisms for the preliminary design phase of unconventional turbomachines. *It is realised that there remains a knowledge gap in determining a value of C_d to predict boundary layer losses for axial turbines operating with organic fluids. Additionally, there lies a gap regarding the parameters that affect the value of C_d.* Covering this gap will i) aide in understanding the influence of different parameters like the working fluid, its thermodynamic properties and blade shape parameters on the boundary layer of ORC power system turbines and ii) allow better boundary layer loss predictions for these turbines. Lastly, covering this gap might help in enhancing the current design guidelines.

As C_d is a function of the momentum thickness based Reynolds number the challenge is to predict a value for C_d without knowing the details about the state of the boundary layer. So C_d needs to be predicted from and expressed as a function of parameters that are known at the preliminary design phase. The conventional approach is to use a theoretical physics based approach to derive this function. Another is to perform experiments or CFD simulations and use the results to develop correlations using a data-driven approach or machine learning. Machine learning utilises algorithms to build mathematical models capable enough to receive and analyse input data to predict a certain output within a certain error margin. Recently, machine learning algorithms have been used to significantly improve and fine tune different model coefficients for Computational Fluid Dynamics (CFD) Reynolds Averaged Navier Stokes (RANS). The benefit of using machine learning for the present work is that a model between multiple parameters can be built with ease and it can be extended to a large variety of fluids, thermodynamic variations and operating conditions. Additionally, this study plays a role in assessing utility and applicability of machine learning algorithms to develop performance trends.

Motivation and Research Question

Given the problem and the available knowledge, *the motivation of the present work is to update the calculation of boundary layer loss in the physics-based model by developing a new data-driven loss model based on results from CFD simulations for the conceptual design phase for unconventional turbomachines.* In doing so the present work seeks to answer the following research question:

> **What are the relevant parameters that affect the blade boundary layer loss model for axial turbines of unconventional turbomachines and how can available models be improved using a data-driven loss model based on results from 2D RANS simulations?**

From this follow a few sub-questions that this thesis will tackle as well:

1. **What is the effect of flow compressibility, working fluid and blade geometrical parameters on C$_d$ for boundary layers of axial stator blades?**
   (a) Which are the most relevant parameters that should be considered while determining C$_d$?
   (b) Which parameters strongly influence the value of C$_d$?
2. **Which 2D blade profiles will define the design space being studied?**
   (a) What are the key design parameters that will define the blade profile?
3. **What strategy will be used to develop the trained data-driven loss model from 2D-CFD RANS simulations?**
   (a) What is the minimum dimension of the database that can be used to provide reasonable results for this study?
   (b) What is the accuracy, in terms of R-square and RSME values, of the machine learnt model ?

Points of originality

The present work is a novel approach to improve a physics loss model using a data-driven approach. The product of this work is a framework consisting of open-source tools to generate numerous blade configurations, perform CFD simulations under different conditions and post-process the results to evaluate and calculate C$_d$. The following points are considered to be original and included in the report:

- Investigation of the relation between working fluid, expansion ratio, compressibility factor and blade design parameters with C$_d$ for unconventional axial turbines.
- Development of performance maps imposed on the Smith chart to determine C$_d$ values for unconventional axial turbines.

- A sensitivity study of C_d with respect to blade design parameters, duty coefficients and inlet and operating conditions.
- Development of a framework to generate a database consisting of working fluid, expansion ratio, compressibility factor, blade design parameters and C_d.
- A machine learnt model capable of predicting C_d based on parameters known at the preliminary design phase and thereby enhancing the accuracy of boundary layer losses.

Thesis Layout

The literature review of topics such as ORC, NICFD, loss mechanisms in turbomachines, especially blade boundary layer losses and machine learning are presented in Chapter 2. The key take-aways and applicability of the literature discussed is the concluding section of this chapter. Chapter 3 describes the flow and details of the framework devised to generate blades, simulate them, extract boundary layer parameters to calculate C_d and eventually perform machine learning on the generated database. The results obtained from flow simulations show a trend upon varying different parameters values. The influence of each parameter and the change in the position of the optimal duty coefficients on the Smith Chart is discussed in Chapter 4. The novel approach of using machine learning to predict C_d from preliminary parameters is discussed in Chapter 5. Thereafter, the machine learnt model is validated and it is plugged into the physics-based loss model. The results obtained from the physics-based model prior to and post the addition of the machine learnt C_d model are also compared in this chapter. Chapter 6 concludes the findings of the present work, answers the research questions and provides recommendations for future work.

2
Background

This chapter presents the literature on Organic Rankine Cycles that aims to explain why organic fluids are better than conventional fluids from a thermodynamic perspective. This is followed by which organic fluids are applicable to ORCs. Subsequently, the study of organic fluids in high velocity flows, NICFD, is presented. As the focus of this thesis is on quantifying loss, the following sections present preliminary design parameters and procedure that are used to design blades and losses in blade rows of turbomachines. Emphasis is given to blade boundary layer losses along with findings of the most recent studies. Finally, since the results generated from this study are used to develop a machine learnt tool to ease preliminary conceptual design phase, literature on machine learning and its usage in engineering applications is provided as well.

Each of the aforementioned topics are an entire study on their own and therefore providing detailed background to every concept is out of scope of this paper. Where required, a brief background is provided to help understand the results that are presented in this study.

2.1. Organic Rankine Cycle

The steam power plant is used globally for electricity generation [18]. The heat required to convert water to steam in these power plants is often obtained by the combustion of fossil fuels such as coal, fuel oil, and fuel gas. Though there are sufficient reserves of these fuels to provide power globally for the next decades, the combustion of the fuels has a significant negative impact on the environment. Another drawback of steam turbines is that even after using the latest technologies they have low efficiencies of about 34% [19]. Lastly, they generate heat that is wasted and is discarded into the atmosphere. As a matter of fact there are several engineering applications that generate waste heat. This can be in the form of heat lost through conduction, convection or radiation, heat discharged from combustion or reactive processes [20]. For instance more than 35% of heat is discharged into the atmosphere from a cement plant [21].

The heat lost from these processes is categorised into three groups, high temperature, medium temperature and low temperature grades. When the temperature of wasted heat is greater than 400°C, it corresponds to high temperature grade. 100°C - 400°C is the medium temperature grade and values below 100°C is low temperature grade[22]. As a typical steam plant requires an inlet steam temperature of 700–720°C [23], temperature ranges from lost heat are insufficient to be utilised by steam plants for generate electricity. When the thermal power available is limited prime movers such as Organic Rankine Cycle (ORC) turbomachines are utilised.

Organic Rankine Cycles follow the exact same principle as a Rankine Cycle. Both are a closed loop cycles that consists of 6 steps which are visualised on a temperature-entropy (T-s) diagram in Figure 2.1. First, the working fluid is compressed using pumps (Step 1-2). This fluid is then heated using a boiler to compressed liquid state (Step 2-3), followed by a two-phase mixture (Step 3-4) at constant pressure. This two-phase mixture is then superheated using superheaters (Step 4-5), after which it is passed through a turbine (Step 5-6). Eventually, the working fluid, which is in a vapour state, is condensed at constant pressure and temperature to a saturated liquid (Step 6-1). The difference between the two cycles is the operating fluid; ORCs use organic fluids.

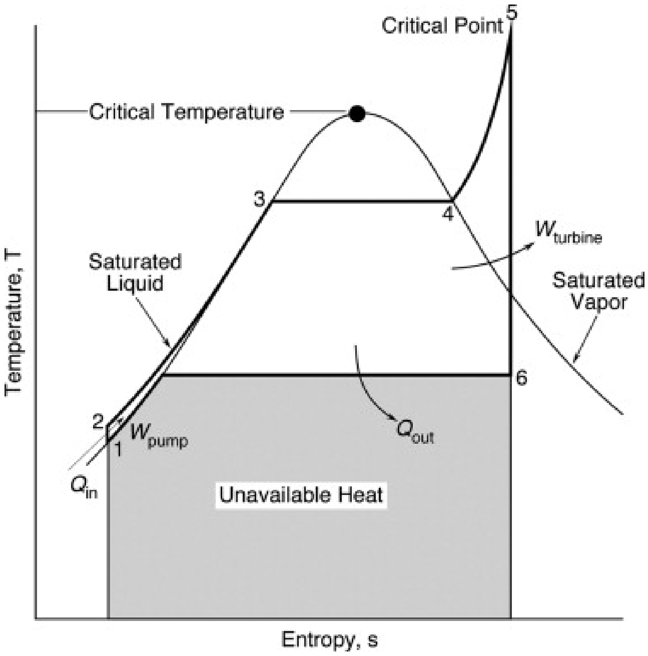


Figure 2.1: An ideal Rankine Cycle on a temperature-entropy curve. The 6 steps show compression of the working fluid (Step 1-2), heating working fluid to a compressed liquid state (Step 2-3) then two-phase mixture (Step 3-4) and finally the supercritical state (Step 4-5), expanding the fluid isentropically (Step 5-6) and finally condensation to saturated liquid state (Step 6-1). [18]

2.1.1. Characteristics of Organic Fluids

Fluids are characterised and differentiated from one another by their unique combination of elements that form their molecules. The elements that form the molecule and how they are attached to one another is what causes each substance to have unique properties. For instance water consists of two hydrogen atoms attached to an oxygen, whereas typical organic fluids such as alkanes, alkynes, siloxanes, refrigerants and alcohols consist of other elements and are arranged in a much more complex manner. One measure of molecular complexity is the acentric factor, that is defined as the magnitude by which the molecule shape deviates from being spherical. Water has an acentric factor of 0.344, whereas that of Dodecamethylcyclohexasiloxane (D6) is 0.736 [24]. Furthermore, the acentric factor is used as a correction factor for thermodynamic models based on simple molecules. As organic fluids contain a larger number of chemical bonds than water their acentric factor is higher. Molecular complexity is directly correlated to the specific molar heat (Cp or Cv), that in turn influences the slope of the saturation vapour curve on a T-s diagram. This is evident from Figure 2.2. This figure shows that the slope of the saturation curve for heavier fluids MDM and Benzene is more positive than water. Complex fluids tend to have a high specific molar heat and the $\gamma = Cp/Cv$ ratio tends to unity. Thus complex fluids have a positive slope of the upper limit curve. Simpler fluids such as water have modest specific heat values and have a less positive slope of the upper limit curve [25]. More positive slopes prove to be advantage for organic fluids in extraction of low grade heat as compared to water.

2.1.2. Organic Fluids Compared to Steam

The merits of organic fluids over water are demonstrated using the T-s diagram shown in Figure 2.2. This figure is used to compare the performance of the three different working fluids, namely water, Benzene and Octamethyltrisiloxane (MDM), using one Rankine cycle. It is assumed that the cycle is limited by the heat source and consequently, does not consider the superheating step shown in Figure 2.1. To make a fair comparison the certain limits are held constant across different fluids so the cycle operates within the same heat sink and source temperatures. Additionally, the condensation and evaporation temperature is assumed to be the same as well. Lastly. it is assumed that the component performance for the three cycles is the same. Based on this the following observations can be made.

- Expansion of water with a temperature limit dictated by low energy power cycles occurs in the wet steam region of the T-s diagram, implying that the vapour quality is low.
- Thermodynamic properties of steam make it a capital-intensive fluid to work with as the formation of vapours in the expander poses higher chances of blade corrosion.
- As a consequence water is a suitable choice for Rankine cycles that operate with high temperature sources, such as nuclear or coal fired power stations.
- For Benzene and MDM the expansion takes place in the steam region of the T-s diagram. This improves the vapour quality which in turn prevents the corrosion of the turbine blades.
- The cycle efficiency of MDM can be improved by introducing a recuperator, however, this drives up the costs and might negate the benefits of a compact machine. The comparison of Benzene and MDM highlights the importance of selecting the correct working fluid for a given limit on the heat source for the ORC turbomachine.

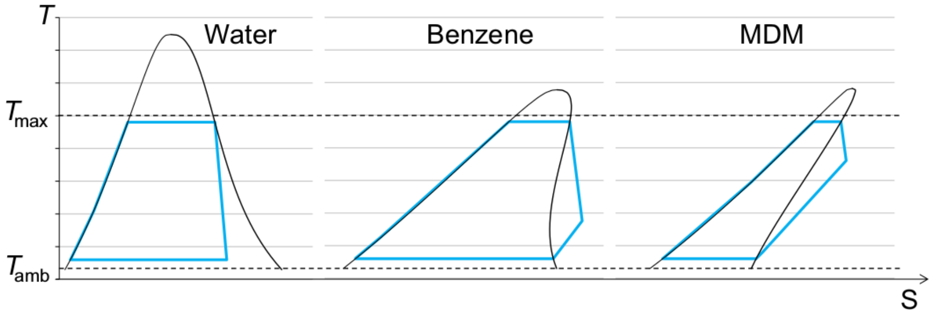


Figure 2.2: A Rankine cycle operating with three working fluids in increasing molecular complexity, on the temperature-entropy diagram. The operating conditions (evaporation and condensation temperature limits and component performance) of the cycle are same for all fluids to enable a fair comparison between the working fluids [9].

The initial arguments that highlight the benefits of organic fluids than water can be further substantiated through further studies. Yamamoto et al compare HCFC-123 to water as a working fluid for an ORC and finds that HCFC-123 provides a better cycle efficiency [26]. Colonna et al design and optimise an energy system operating with water and siloxane MDM, provided the boundary conditions and energy source remains the same. The results showed that cycle efficiency increases when operating with a complex molecule [27].

Selecting the correct working fluid for an ORC becomes a priority. Several studies comparing the performance of different organic fluids for different applications (small/big scale and geothermal/biomass) can be found in literature. The fluid for an application must have a suitable critical temperature, and pressure. This is because the expansion of working fluids of ORC turbomachines occurs close to critical temperatures and pressures. The critical temperature should permit evaporation at an appropriate temperature and allows condensation pressures to be higher than the atmospheric pressure [9].

2.2. Non-Ideal Compressible Fluid Dynamics

The combination of temperatures and pressures above the critical values lie in the supercritical region. The distinction between liquid and gaseous phase cannot be made for fluids in this region. Non-Ideal Compressible Fluid Dynamics (NICFD) is a branch of gas dynamics that deals with fluids whose thermodynamic properties largely depart from the ones predicted by the ideal gas model. As a matter of fact, the ideal gas law, which assumes the there is no inter-molecular attraction, proves not to be reliable in correctly determining the trend of the thermodynamic variables such as the Mach number [28].

2.2.1. Quantification of Non-Ideal Behaviour

To determine which equation of state is to be used to predict flow behaviour for which combinations of temperatures and pressures, the deviation from gas ideality is quantified. Non-dimensional parameters such as compressibility factor, polytropic exponent and fundamental derivative of gas dynamics are commonly used.

Compressibility Factor

A parameter to quantify the deviation of the behaviour of real gas from ideal gas is the compressibility factor (Z), shown in Equation (2.1) [29].

$$Z = \frac{Pv}{R_{gas}T} = \frac{v}{v_{ideal}} \tag{2.1}$$

This equation shows that the compressibility factor is a ratio of the volume of the gas (v) at a given temperature (T) and pressure (P) to the volume of the gas if it were an ideal gas (v_{ideal}) at the same temperature and pressure. Z is a function of the temperature, pressure and composition of the gas that is specified using (R_{gas}). Consequently, the compressibility factor is determined from an equation of state that considers compound specific empirical constants. However, the compressibility factor can also be determined using compressibility charts as shown in Figure 2.3. This figure relates Z with the reduced pressure (P_r = P/P_{crit}) for a constant reduced temperature T_r = T/T_{crit}.

Figure 2.3 shows that Z can have a value ranging from 0 to greater than 1. The value of unity implies ideal behaviour, Z less than unity indicates non-ideal behaviour and Z greater than unity implies processes such as dissociation or ionisation of molecules. This is true for all fluids. As per the theorem of corresponding states, if two fluids have the same T_r and P_r then Z will be similar for both fluids. Therefore the compressibility chart shown in Figure 2.3 can be used for determining Z for real gases as well [30].

Polytropic Exponent

While the compressibility factor is useful in quantifying the deviation from ideal gas, a parameter is required to evaluate the performance of a turbomachine in non-ideal conditions. For this the polytropic exponent (γ_{pv}) is used.

Consider the expansion of the working fluid in a turbine that was show in Figure 2.1. The drop from stage 5 to 6 is vertical on the T-s diagram, implying that the expansion is isentropic.

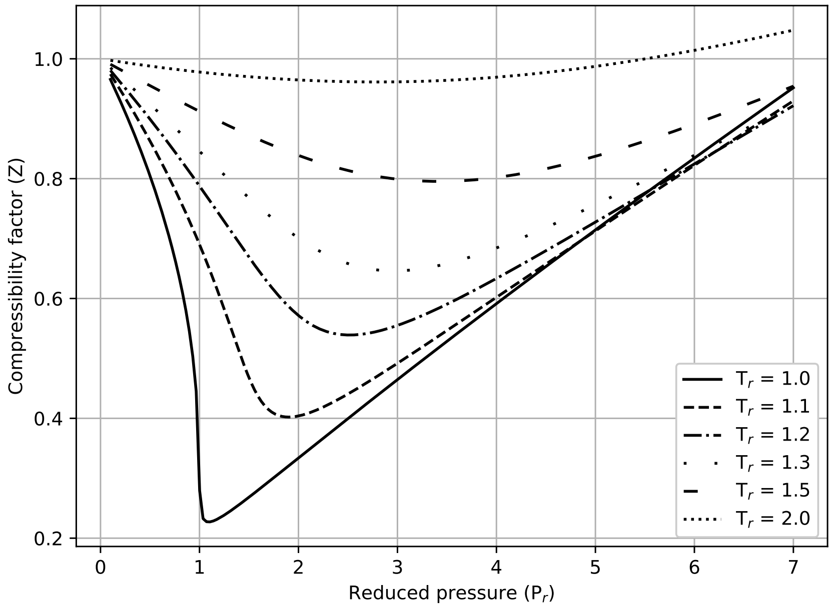


Figure 2.3: Compressibility factor chart for Air.

Equation (2.2) describes the variation of thermodynamic variables pressure and volume in dense gas flows where k is a constant and γ is the specific heat ratio.

$$[ht]Pv^{\gamma} = k \tag{2.2}$$

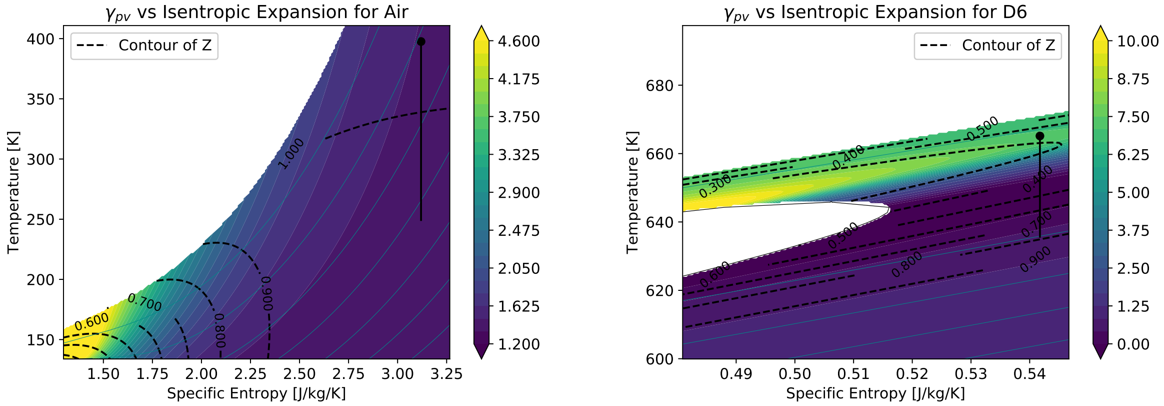
However, this relation is valid only under ideal gas assumption. For non-ideal conditions, γ value is replaced by γ_{pv}, known as the polytropic exponent. γ_{pv} accounts for the non-ideality of the flow. It is the equivalent of γ in ideal gas. Mathematically, it is expressed in Equation (2.3).

$$\gamma_{pv} = -\frac{c_p}{c_v}\frac{v}{P}\left(\frac{\partial P}{\partial v}\right)_T \tag{2.3}$$

For application, this exponent can be determined using a thermodynamic library or an equation of state. The accuracy of this exponent relies on the equation of state being used to relate the state variables [31] [32].

Figure 2.4 illustrate the application and relation between γ_{pv} and Z. Both these figures map the total-to-total isentropic pressure expansion ratio (β_{tt}) on the iso-contours of Z and γ_{pv}. Two unique inlet conditions on two different fluids, namely Air and D6, are used. Figure 2.4a shows that inlet conditions P$_r$ = 1.0 and T$_r$ = 3.0 correspond to Z 1 and γ_{pv} = 1.4. Z 1 signifies that the ideal gas law can be used to determine local thermodynamic variables, hence γ_{pv} = γ. The expansion process shows almost no variation in either of the two contour levels. On the other hand, Figure 2.4b shows the dramatic change in values of Z and γ_{pv} along the expansion. Therefore, signifies the importance of determining the local conditions when away from Z « 1. The change in γ_{pv} for two fluids with same β_{tt} influences volumetric expansion ratio α_{tt}. Therefore, the influence of γ_{pv} in turbomachinery applications is of importance.

This is substantiated by results from Wheeler et al; they show that for a range of γ_{pv} the loss in a turbine can vary between 20% and 35%, depending on vane exit Mach number. This proves the importance for turbines operating with real-gas in regions close to the critical point, such as ORC turbomachines.



(a) Air, P$_r$ = 1.00, T$_r$ = 3.00.

(b) D6, P$_r$ = 1.80, T$_r$ = 1.03.

Figure 2.4: Isentropic expansion path of β_{tt} = 4 on iso-Z and iso-γ_{pv} contours.

Fundamental Derivative of Gas Dynamics

Another parameter that is used in NICFD is the fundamental derivative of gas dynamics (Γ). This term, expressed in Equation (2.4), measures the variation of the speed of sound with respect to density in isentropic transformations such as in gas-dynamic nozzles [33].

$$\Gamma = 1 + \frac{\rho}{c}\left(\frac{\partial c}{\partial \rho}\right)_s \tag{2.4}$$

It quantifies the gas dynamic behaviour into three categories, namely classical ideal ($\Gamma >$ 1), classical non-ideal ($0 < \Gamma < 1$) and non classical ($\Gamma < 0$). The first suggests that the speed of sound is positively correlated to the density, the second suggests the opposite and the last occurs in close proximity of the critical point for complex molecules and suggests the presence of non classical gas dynamics phenomena such as rarefaction shock waves [28]. Therefore, Γ is also used to study the wave propagation in dense vapours [34]. This use of this parameter is limited in the present work as it is considered out of scope of this work.

The quantification of non ideal behaviour allows one to determine the equation of state to use and allows one to determine the local thermodynamic conditions along the expansion process for a set of inlet P$_r$, T$_r$. Change in γ_{pv} values close to critical conditions influence losses in turbomachines.

2.3. Loss Mechanisms in Turbomachines

While the importance of selecting a working fluid is crucial for assessing performance of an ORC turbomachine, so is the design of the individual components, such as the blades of stators and rotors of expanders. The thermodynamic properties and the operating conditions of fluids influence the design of the turbines, which then contributes to the efficiency of the entire system as a whole.

In order to enhance the performance of these individual components it is important to comprehend the sources and types of losses that occur in turbomachines. Unlike in aircraft aerodynamics, where the drag force is considered to be loss in performance, in rotating flow passages losses are quantified using the rate of entropy generation. There are two main reasons for this. First, the direction of flow is difficult to define and second, the drag force contributes to the pressure increase in compressors, which is positive work, therefore is not a loss [2].

2.3.1. Relation between entropy and loss

Specific entropy is an extensive property and a quantitative measure of microscopic randomness in a macroscopic system [35]. The Second Law of thermodynamics states that entropy of a system can be produced but never destroyed and is constant if and only if all processes are reversible. Greitzer demonstrates how entropy can be used as a measure of loss by considering two cases, reversible and irreversible, in which a unit of mass is transformed from one state to the other [36]. As the state transformation for both cases is the same the change in energy is the same. The First Law of Thermodynamics in equation for both the cases is equated, as shown in Equation (2.5).

$$\begin{aligned} \delta q_{rev} - \delta w_{rev} &= \delta q_{irrev} - \delta w_{irrev} \\ \Longrightarrow \ \delta q_{rev} - \delta q_{irrev} &= \delta w_{rev} - \delta w_{irrev} = \delta w_{loss} \end{aligned} \tag{2.5}$$

In this equation δq_{rev} and δw_{rev} are reversible heat and work respectively. δq_{irrev} and δw_{irrev} are irreversible heat and work respectively. The difference between reversible work and work that has been lost due to irreversibility is associated with loss (δw_{loss}). Substituting the definition of entropy in Equation (2.5) leads to Equation (2.6).

$$ds = \frac{\delta q}{T} + ds_{irrev} = \frac{\delta q}{T} + \frac{\delta w_{loss}}{T} \tag{2.6}$$

According to this equation the change of entropy (ds) is the sum of the entropy change due to heat transfer ($\delta q/T$) and irreversibility (ds_{irrev}). If the process is reversible the irreversible entropy will be zero, otherwise, the irreversible entropy is lost work (w_{loss}). Furthermore entropy is a function of any two thermodynamic properties such as temperature and pressure, and, it is independent of static or stagnation conditions because the change from one to the other is isentropic.

2.3.2. Loss coefficient

The performance of a machine is characterised by the isentropic efficiency, which is the ratio of actual work done by the machine to the isentropic work [12]. Any deviation in isentropic flow implies presence of irreversibilities and lost work. The loss coefficient is a measure of the net increase in entropy to the available total energy in the flow. If the pressure iso-lines on the enthalpy-entropy chart is assumed to be equal to the local temperature, the entropy loss coefficient for a turbine can be expressed as in Equation (2.7).

$$\zeta_{turb} = \frac{T_2(s_2 - s_{2s})}{h_1 - h_{2s}} \tag{2.7}$$

Where T_2 is the outlet temperature, s_2 is the specific entropy at exit of turbine, s_2s is the isentropic specific entropy at the exit of the turbine, h_1 is the specific enthalpy at the inlet of the turbine and h_2s is the isentropic specific enthalpy at the exit of the turbine.

2.3.3. Sources of irreversibility

The flow in turbomachines is mostly adiabatic thus the heat transfer term in Equation (2.6) is almost always negligible. This implies that the main source of entropy generation is due to irreversibilities. The creation of entropy in a flow is attributed to the following fluid dynamic processes:

1. Viscous dissipation. This includes dissipation in boundary layers or due to mixing.

2. Heat transfer across finite temperature differences.
3. Shock waves causing sudden expansion.

Each of these processes will be used to explain the sources of losses in turbomachines.

2.3.4. Loss breakdown

Due to the complex fluid dynamics, loss mechanisms in turbomachines are seldom independent. Three main mechanisms are recognised and contribute to the loss coefficient almost equally [2]. These are:

1. **Profile losses** entail losses due to the blade boundary layers, mixing losses at the trailing edge, losses due to separation and shock waves. Profile losses can be calculated using 2D assumptions and as a result are mostly calculated away from the hub and the casing.
2. **Endwall losses** generate at the hub and the casing due to the boundary layer and secondary flows in and out of blade rows. These losses occur due to a combination of multiple factors and is different to differentiate from the other two losses.
3. **Tip leakage losses** occur when the flow from the pressure side leaks to the suction side at the tips of the blades over the clearance gap. The size of the clearance gap and blade type are detrimental to efficiency.

As the objective of the present work is to develop a data-driven loss model for only losses due to blade boundary layer, the other two sources of losses are not discussed.

2.3.5. Profile losses

Blade boundary layer, trailing edge and shock wave losses are the three contributors to profile losses. The flow over a turbine or compressor blade forms a boundary layer close to the surface due to viscous effects. After traversing a certain distance along the chord of the blade this flow either separates or remains attached to the flow. Eventually at the end of the chord length the flow from both the surfaces of the blade mix causing mixing losses. Lastly, if the flow is transonic or supersonic, shock waves might occurr; shock losses thus need to be accounted for as well,

Blade boundary layer losses

The flow in a turbomachine is 3D, which makes estimating losses challenging. However, blade boundary layer losses can be simplified to 2D under the assumption that the flow over the blade is analysed away from the casing and the hub. For a 2D boundary layer on an adiabatic surface Denton derives a mathematical formulation for total rate of entropy generated per unit surface area shown in Equation (2.8):

$$\dot{S}_a = \frac{d}{dx}\int_0^{\delta_e} (\rho V_x (s - s_{\delta_e}))dy = \int_0^{\delta_e} \frac{1}{T}\tau_{yx} dV_x \tag{2.8}$$

Equation (2.8) states that the total rate of entropy creation per unit area is the same as the entropy created due to viscous effects in the boundary layer. The dimensionless form of rate of entropy generation is the dissipation coefficient (C$_d$), expressed as Equation (2.9). C$_d$ is a function of the temperature (T), entropy production rate per area ($\dot{S_a}$), density (ρ_{δ_e} and velocity at the edge of the boundary layer (v_{δ_e}).

$$C_d = \frac{T\dot{S}_a}{\rho_{\delta_e} V_{\delta_e}^3} \tag{2.9}$$

There are two types of boundary layers, laminar and turbulent. For most boundary layers the entropy generation is concentrated in the inner layer [37]. This statement is reinforced by Dawes' research, which shows that approximately 90% of all the entropy generated in the boundary layer is from the inner layer [38]. As the exact value of C_d cannot be calculated without knowing details of the flow, separate empirical methods to estimate C_d for laminar and turbulent flows are used, both of which are shown in Figure 2.5.

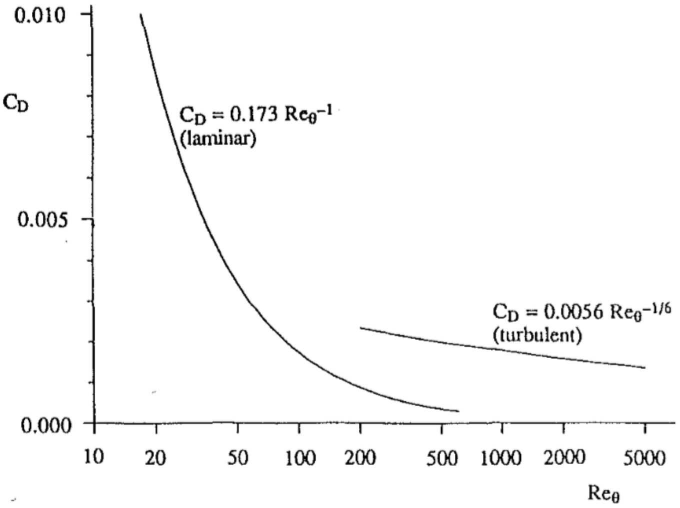


Figure 2.5: Correlation between dissipation coefficient (C_d) and momentum thickness Reynolds number (Re_θ) for laminar and turbulent boundary layers based on works of Truckenbrodt and Schlichting respectively [2].

Schlichting developed the correlation for turbulent boundary layers reported in Figure 2.5. This relationship between C_d and momentum thickness Reynolds number (Re_θ) is based on experiments with air as a fluid. This relationship is applicable for boundary layer shape factors between 1.2 to 2 and momentum thickness Reynolds numbers (Re_θ) between 10^3 and 10^5 [15].Cebeci developed a tool to estimate C_d for boundary layers operating under three different conditions. The first condition is where the flow is diffusing, second of constant pressure and third where the flow is accelerating [37]. Figure 2.6 shows the results obtained from the tool and Schlichting's correlation. Three observations are drawn from this figure that eventually leads to one useful approximation.

1. Both simulation and experiment shows a similar behaviour for constant pressure boundary layer at high Re_θ.
2. C_d for the accelerating boundary layer decreases more than the other two cases with increase in Re_θ. This can perhaps be attributed to the fact that the thickness of accelerating boundary layers is less than that of the other two.
3. The boundary layer for the flow that is diffusing leads to higher C_d than the other boundary layers. This is due to higher entropy generation in the boundary layer due to adverse pressure gradient.

Given all three behaviours are observable in turbomachinery blades, Denton and Cumpsty suggest that for machines operating at an average Re_θ of 1000 a constant C_d value of 0.002 can be assumed [2].

The correlation for laminar boundary layers shown in Figure 2.5 $C_d = 0.173 Re_\theta^{-}1$ is derived by Truckenbrodt. This relation for C_d is inversely proportional to Re_θ and directly to a function

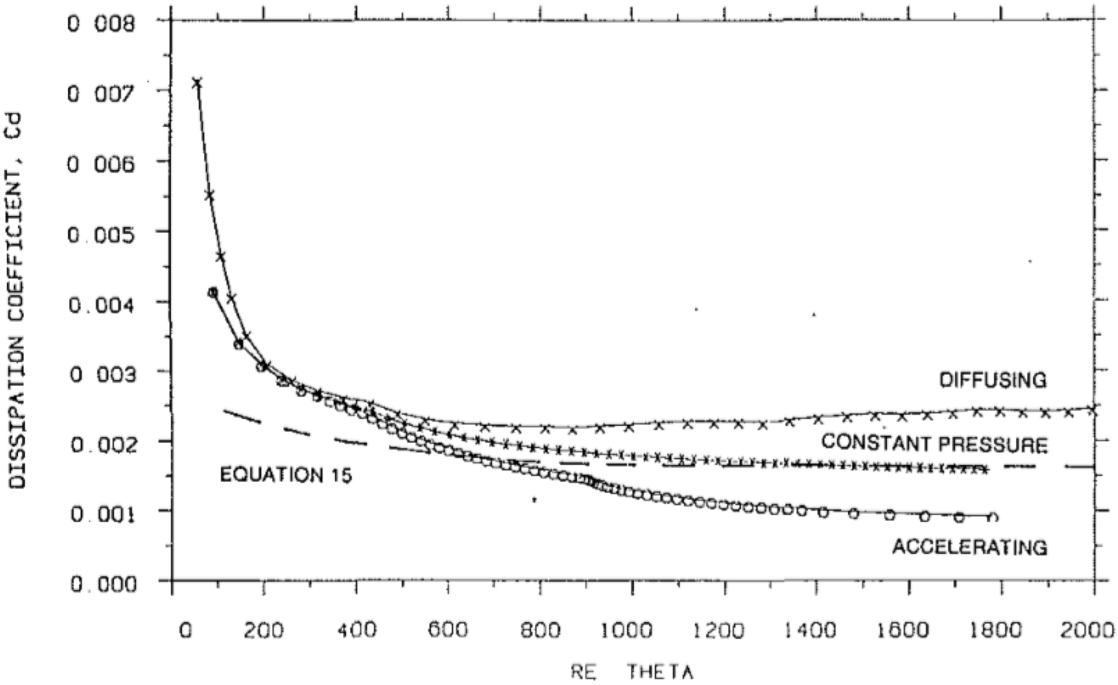


Figure 2.6: Dissipation coefficient determined using emperical equation of Schlichting (Equation 15 in figure) and computer tool by Cebecci [2].

of the shape factor 0.173. However the shape factor has little to none variation and is just a constant for a laminar boundary layer with no pressure gradient. This value is determined from an analytical expression [39].

The region where the overlapping of the two trends occurs in Figure 2.5 is where laminar to turbulent transition takes place. Identifying the transition point is a challenging task on its own but can be crucial in estimating the losses as both trends predict a value of C_d that is different from each other by a factor of 2 to 5.

The total entropy generated due to the boundary layer upstream of a certain point on a surface is determined using Equation (2.10).

$$\dot{S} = \int_0^x \frac{\rho V_{\delta_e}^3 C_d}{T_{\delta_e}} dx \tag{2.10}$$

This equation shows that if the distribution of density, C_d, velocity and temperature at the edge of the boundary layer is known then integration along the surface length (x) produces the rate of entropy generation on that surface length. If the flow regime over a surface is known to be strictly turbulent a constant value of C_d suffices. However, doing the same for transitioning or accelerating boundary layers would lead to an overestimation of losses. Subsequently, if this procedure is repeated for both surfaces of a blade, the entropy generation for a blade would only be overestimated. The present work intends to explore and develop trends or correlations to prevent this overestimation.

To convert the entropy generated due to boundary layer losses to the loss coefficient, the entropy generated is divided by the product of a the dynamic head and the mass flow of the fluid being analysed. The entropy equation for a low speed flow is Equation (2.11).

$$\zeta = 2\sum \frac{c}{p\cos\alpha}\int_0^1 C_d\left(\frac{V_{surf}}{V}\right)^3 d\left(\frac{x}{c}\right) \tag{2.11}$$

From this equation it is observed that the velocity distribution along the surface (V_{surf}/V) and the transition point are two key parameters to determine the loss coefficient. Furthermore, it is known that the upper surface of the blade or the suction side would contribute more to the loss coefficient more than the lower surface. Similarly, a turbulent boundary layer would lead to a higher dissipation coefficient and therefore it is important to keep the boundary layer laminar for as long as possible. Figure 2.5 shows that as the Reynolds number increases and the boundary layer is laminar the C$_d$ decreases. However soon after transition the dissipation coefficient increases significantly and thereby increases the loss coefficient in Equation (2.11). In reality, determining an accurate velocity distribution for a preliminary analysis is a challenge. To solve this Denton assumes C$_d$ as a constant and develops an idealised velocity distribution.

The physics-based loss model developed by Pini and Giuffre called TurboSim uses the same approach to estimate boundary layer losses for axial turbine stages. The model has the option to use two velocity distributions. First is the same as Denton's and the second is an average blade surface velocities. The latter always leads to lower values of blade boundary layer loss when assuming constant C$_d$ values. The results of the study conclude that the if C$_d$ is assumed to vary proportionally with the blade surface velocity along the streamwise direction, as suggested in the predicted blade boundary layer loss could be equal or even higher than the one estimated with the simple rectangular distribution [14]. The present work focuses on implementing a model that would allow for the variation of C$_{d,blade}$.

Finally, Denton extends the idea of loss estimation from a blade to an entire stage by incorporating isentropic velocity but still assuming that C$_d$ stays constant. He concludes that boundary layer losses contribute up to half of the total profile losses in most turbines [2]. Once the fluid flows over the entire length of the chord, the trailing edge losses start to dominate.

Trailing edge losses

The mixing of the boundary layers from the suction and pressure surface just behind the trailing edge generates losses. Research by Mee et al. and Roberts showed that the trailing edge losses contribute approximately one-third of the total profile losses [40][41]. Shear strain of fluids is not limited to boundary layers and it occurs in mainstream flow, wakes, separated flow, vortices and leaking flows. It is known that mixing will continue till the flow becomes uniform far downstream the trailing edge. By calculating the entropy via the control volume analysis one bypasses the physics of the flow in the region and determines the total loss.

Denton demonstrates this statement by calculating the trailing edge loss coefficient on a blunt trailing edge. He introduces a term called base pressure coefficient (C$_{pb}$) that is expressed as Equation (2.12). P$_b$ is the pressure acting on the trailing edge because then C$_{pb}$ becomes a function of the boundary layer, shape of the trailing edge and trailing edge thickness to boundary layer thickness. The value of C$_{pb}$ is used to derive the loss coefficient relation according to Equation (2.13). The formulation consists of three parts, from left to right the different terms of the equation account for the losses due to base pressure, mixed out losses of the boundary layer and the blockage of the trailing edge and boundary layer respectively. The implementation of Equation (2.13) in a model using only theoretical equations remains challenge because the value for C$_{pb}$ is determined through experimental data [2].

$$C_{pb} = \frac{P_b - P}{0.5\rho V^2} \tag{2.12}$$

$$\zeta = -\frac{C_{pb}t}{w} + \frac{2\theta}{w} + \left(\frac{\delta^* + t}{w}\right)^2 \tag{2.13}$$

If flow separation occurs before the trailing edge an increase in the size of vortices and dissipation in the wake occurs. All this would occur in addition to the trailing edge losses therefore suggesting the addition of an extra loss term, expressed as Equation (2.14), to the losses incurred by trailing edge losses [2].

$$\zeta_{sep} = \left(\frac{\delta^{*2} + 2t\delta^*}{w^2}\right) \tag{2.14}$$

Shock losses

The entropy generated due to shock losses can be calculated by Equation (2.15). Expanding this equation shows that the entropy creation is approximately proportional to the cube of (M^2-1). For normal shocks M is the upstream Mach number whereas for oblique shocks it is the value perpendicular to the shock wave. Entropy generated by an oblique wave is less than that of normal one.

$$\frac{\Delta s}{C_v} = \ln\left[\frac{2\gamma M^2 - \gamma + 1}{\gamma + 1}\right] - \gamma \ln\left[\frac{(\gamma + 1)M^2}{(\gamma - 1)M^2 + 2}\right] \tag{2.15}$$

Denton proposes using Equation (2.16) to estimate the entropy generation by weak shocks, irrespective of it being normal or oblique. For transonic turbines shocks help in achieving high stage pressure ratios but for most turbines this is not the case. Unlike shocks forming at the leading edge in compressors shocks in turbines forms at the trailing edge. Consequently it influences the trailing edge losses.

$$\Delta s \approx R\frac{\gamma + 1}{12\gamma^2}\left(\frac{\Delta P}{P}\right)^3 + O\left(\frac{\Delta P}{P}\right)^4 \tag{2.16}$$

Shock wave - boundary layer interaction also affects the rate of entropy generation. The strength of the influence depends on the strength of the shock and type of boundary layer (laminar or turbulent). For instance, a weak shock produces a separation bubble at its foot but if the flow is laminar there is a high probability of the flow to reattach. On the other hand a strong shock might cause boundary layer separation.

Literature from this section shows that the dimensionless blade boundary layer dissipation coefficient C_d is a measure of losses due to boundary layers. It is easier to study compared to other losses because it can be simplified to a 2D plane. To determine C_d for any surface four variables are required, namely temperature, velocity and density at the edge of the boundary layer and the rate of entropy generation. Determining these values at a preliminary design stage of blades is a challenge because the blade profile influences the flow over it and therefore these parameters.

2.4. Design of Blade Profiles

As C_d can be determined from 2D blade profiles, the present work deals with the design of only 2D blade profiles. This study aims to execute an analysis on as many different blade geometries as possible. Simulating the stage is computationally expensive and would limit the number of simulations due to time constraints on the present work. Additionally, this study is an exploratory research. Hence devoting too much emphasis on the design of blade profiles is not considered to be a suitable approach. In light of these considerations, it was decided early on in the literature study phase that a stator cascade of a single stage axial turbine will be used.

2.4.1. Duty Coefficients

There are three duty coefficients, namely flow coefficient ϕ, work coefficient ψ and the degree of reaction $r*$. These three parameters together define the architecture of the stage and thereby the performance [42].

Flow Coefficient

The flow coefficient ϕ is defined as a measure of the flow capacity of the stage. Equation (2.17) shows mathematical formulation, where V_{axial} is the velocity of the flow in the axial direction of the machine and the rotational speed of the rotor speed is U.

$$\phi = \frac{V_{axial}}{U} \tag{2.17}$$

Higher values of ϕ imply high flow capacity of the stage, therefore higher volumetric flow ratio across the stage.

Work Coefficient

Equation (2.18) shows the equation of the work coefficient, where w is the specific work of the stage.

$$\psi = \frac{w}{U^2} \tag{2.18}$$

This parameter measures the work capacity of the stage. Higher values correspond to higher work to be done by the stage and thus more severe the aerodynamic blade loading.

Degree of reaction

The degree of reaction $r*$ is the ratio of the static expansion of the flow across the rotor to the total specific work across the stage. It is expressed as:

$$r* = \frac{\Delta h_R}{w} \tag{2.19}$$

where h_R is the static enthalpy drop in the rotor.

Although these parameters are used for the calculation of thermodynamic parameters across the stage, these parameters dictate the inlet, outlet flow angles of the stator. The relationship between the flow angles is related to the duty coefficients using the following relations:

$$\psi = (\phi tan(\alpha_1) - (\alpha tan(\beta_2)) - 1 \tag{2.20}$$

$$r* = \frac{\psi}{2} - \alpha tan(\beta_1)) - \frac{1}{\psi} \tag{2.21}$$

$$tan(\beta_1) = tan(\alpha_1) - \frac{1}{\phi} \tag{2.22}$$

$$tan(\alpha_2) = tan(\beta_2) + \frac{1}{\phi} \tag{2.23}$$

In these relations α corresponds to absolute angles of the flow with respect to the axial direction of the machine and β corresponds to relative angles of the flow with respect to the axial direction of the machine. Subscripts 1 and 2 correspond to the inlet and outlet of the stator [12].

As the stage is not considered for the present work, the degree of reaction is assumed to be a constant = 0.5 for all blade geometries. Therefore for a given combination of ϕ, ψ the flow angles across the stator can be determined.

2.4.2. Blade Solidity

The outlet angles determined from the duty coefficients are used to determine the solidity for the stator cascade. The solidity is the ratio of the axial chord to pitch ratio and controls the spacing between blades and therefore influence losses. Large values of solidity imply the smaller spacing between blades and therefore losses due to drag dominate. On the other hand, small values of solidity imply larger spacing and the ability of blades to do work reduces [42].

The solidity is determined using the Zweifel criterion, which uses the velocity triangles determined using the duty coefficients [12]. The equation for solidity for the stator is expressed as:

$$\sigma_s = \frac{Zw}{2cos^2(\alpha_2)(tan(\alpha_2) - tan(\alpha_1))} \tag{2.24}$$

The value for the Zweifel coefficient is taken as 0.8 for all blade profiles generated in the present work.

This section presents the parameters that dictate the blade profile that in turn influence the flow over the blade and therefore the boundary layer over it.

2.5. Similarity criteria

The general behaviour of turbomachines is obtained from the dimensional analysis, which is the formal procedure whereby the group of variables representing a physical situation is reduced to a smaller number of dimensionless groups. Similarity parameters are the outcome of the dimensional analysis. The four groups of similarity parameters, shown in Equation (2.25), are Geometric, Dynamic, Kinematic and Thermodynamic. If all the similarity parameters for two or more physically different stages is the same, the performance or the efficiency is the same.

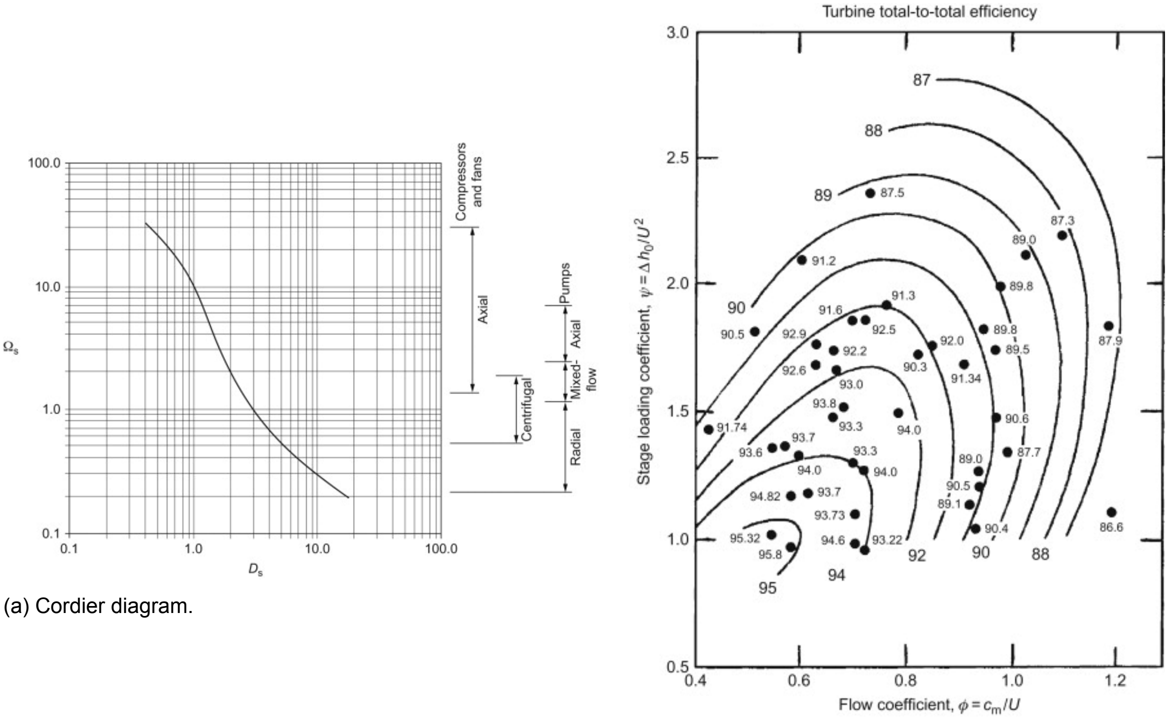
$$\eta = f(\underbrace{\phi, \psi, r*}_{\text{Kinematic similarity}}, \underbrace{Re, M}_{\text{Dynamic similarity}}, \underbrace{P_r, T_r, \alpha, fluid}_{\text{Thermodynamic similarity}}, \underbrace{\sigma}_{\text{Geometric similarity}}) \tag{2.25}$$

The grouping of different dimensional parameters into a dimensionless one is a practical benefit as it allows the representation of performance data using one single curve rather than multiple [12]. For instance the Cordier diagram (shown in Figure 2.7a) correlates the turbomachine size and type as a function of non-dimensional rotational speed and diameter for optimum efficiency. Another widely used example is the Smith Chart that correlates the turbine stage efficiency as a function of the duty coefficients.

For the present work the parameters influencing the dissipation blade boundary coefficient are shown in

$$C_d = f(\phi, \psi, r*, P_r, T_r, Re, \alpha, fluid, \sigma) \tag{2.26}$$

It is assumed that r* = 0.5. Also the inlet T_r and P_r can be expressed as the compressibility factor Z. Additionally, the working fluid, T_r and P_r determine the Re, therefore this parameter can be neglected as well. Consequently, Equation 2.26 reduces to:



(a) Cordier diagram.

(b) Smith Chart.

Figure 2.7: Examples of non-dimensional charts. (a) shows the turbomachine size and type as a function of non-dimensional rotational speed and diameter for optimum efficiency. (b) shows the turbine stage efficiency as a function of the duty coefficients.

$$C_d = f(\phi, \psi, Z, \alpha, fluid, \sigma) \tag{2.27}$$

Equation (2.27) expresses C$_d$ as a function of factors $\phi, \psi, \sigma, Z, \alpha, workingfluid)$. Some of the parameters are similar to those selected by Pini and Carlo in their study of entropy generation in laminar boundary layers [16].

The objective of the present work is to determine i) quantify the influence of each parameter on C$_d$ ii) develop a mathematical equation that factors in the influence of each parameter and is developed using several instances of data. To accomplish the latter machine learning is used.

2.6. Machine Learning

Machine learning is the science of programming computers to learn from data. The machine learns from experience with respect to a particular task and performance measure if its performance on the task improves with experience [43]. Machine learning involves the use of algorithms to dissect the data, learn from it and make predictions or perform a particular task. Larger the data provided to the machine the more it can learn and train to perform the task better [44].

The typical steps involved in developing a machine learnt model are illustrated in Figure 2.8. There are two broad steps, namely Handling data and Selecting a model. The prior deals with manipulating data such that a model capable of predicting accurately is obtained. The latter deals with selecting and training a model using some algorithms. The following sections will describe the sub-steps mentioned in each block.

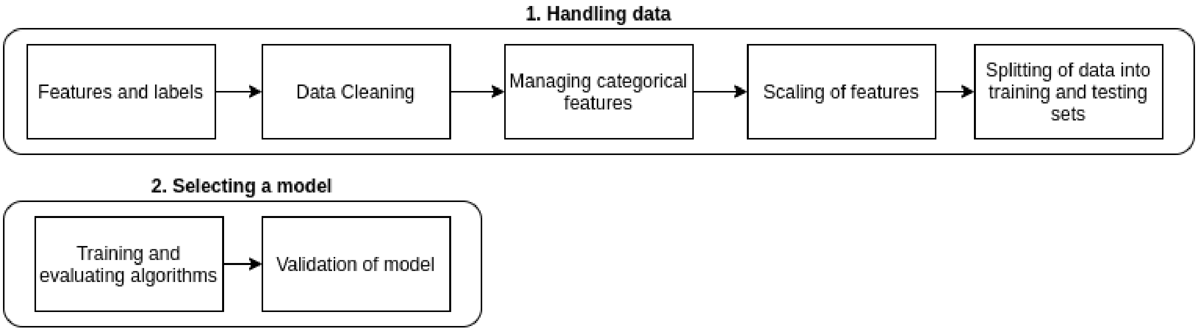


Figure 2.8: Typical steps involved in developing a machine learnt model.

2.6.1. Handling data

The following section presents the steps involved to label data, clean the data and remove outliers or incorrect data, ensure non-numeric or categorical data is treated, scale the features and eventually split the database into a training and testing set.

Features and labels

The data that the machine learning algorithm uses is required to be divided into labels and features. The use of diving the labels into two groups is to inform the algorithm of the values that are being predicted and the values used as inputs to make the prediction.

- **Features** are the descriptive attributes of the data. For instance they are inputs x_1 and x_2 in the linear regression equation $y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$. θ_i is a coefficient that corresponds to input x_i.

  There is no limit to the number of features that can be involved in an analysis for machine learning, however, in practice it increases the complexity and influences the computation time.

- **Labels** are quantities that are being predicted and are the output. It is the 'y' in the aforementioned linear equation.

After the division of data into labels and features the database is checked for incomplete data.

Data Cleaning

The quality of the data in a database might vary depending on the source. There might be instances of data that are incomplete. There are three options to deal with such instances, they are:

- All instances of data with a similar value for that feature are dropped from the database and neglected from the machine learning analysis. The drawback of this approach is that it neglects an entire group of data from the analysis [45]. For example consider a database with instances containing data about the fluid, duty coefficients and C_d. A few instances with incomplete data for one fluids would lead to neglecting all instances of that fluid.

- Only the incomplete instances are dropped from the database and neglected from the machine learning analysis. This is a better approach than that of point 1. However, the drawback of this depends on the number of missing instances. If too many instances of a single feature are incomplete that would lead to introducing a bias in the database.

- The missing values of features are assigned certain values. These certain values could be 0, the mean or median of all the values of that feature.

Once the data is cleaned the data is prepared for the machine learning algorithms. The categorical data is converted to numeric values and numeric data is scaled.

Managing categorical features

Machine learning models are developed on the basis of numeric data. However, in some cases some features will be non-numeric. Such features are called categorical data. Accounting of these features to develop models poses a problem. There are two common approaches to tackle this problem in Sklearn [46], they are:

- Ordinal Encoder: Converts categorical features into integer arrays. Features provided to this transformer are covered to ordinal integers. The results of this is a single column of integers from 0 to ($n_categories$ - 1) per feature. The drawback with this transformer is that values close to each other are considered to similar to each other by the machine learning algorithms.
- OneHot Encoder: This transformer transforms the categorical features using a one-of-K encoding scheme. The result is matrix with a binary column for each category. The algorithms interpret the categorical data as an on/off switch.

After the categorical features are taken care of, the numeric values are scaled to avoid bias in the results.

Scaling of Features

Different features in a database contain different scales of data. For instance a database might contain the features inlet pressure in bar and maximum blade thickness metres. If these two features are not scaled to one single scale, then the machine learning algorithm might give more or less bias to pressure only due to its magnitude. This would cause the output trend to be skewed or have a bias [46] [45]. The two most common approaches to attain the same scale for different features are:

- min-max scaling or normalisation : is the scaling of data between 0 and 1. This is accomplished by using the following equation:

  $$X_{std} = \frac{X - X_{min}}{(X_{max} - X_{min})}$$

  X_{std} is the score, X is an instance of a feature, X_{min} and X_{max} are the minimum and the maximum value respectively out of all the instances of that feature.

- standardisation : is the scaling of features by removing the mean and scaling it to the unit variance. The equation used to calculate the standard score is:

  $$X_s td = \frac{(X - U)}{S}$$

  X_{std} is the score, X is a single instance of a feature, U is the corresponding mean and S the standard deviation. The consequence of this is that the data is converted to a Gaussian distribution with the mean at 0 and unit variance.

The completion of this step means that the features Are ready to be used by machine learning algorithms because categorical features have been quantified and numeric features have been brought to a single scale.

Splitting of Data

The scaled database is split into two sets, namely training and testing set. The data in the prior is used to train the model and the latter is used once a model is finalised. The testing set contains data that the model has never ”seen”. As a result this set is used to assess the performance of the final model to new data. The training to testing data is commonly 8:2 [45], however this ratio can be increased depending on the size of the database.

The segregation of data into the two sets can be executed using many ways. The first is a random shuffle. Random shuffle on the data distributes random instances of data into the two sets. This approach is acceptable if the size of the database is large. For smaller databases the drawback of this method is that a random shuffle might introduce a significant sampling bias. The second type of shuffle is stratified sampling on the basis of a certain feature. This ensures that the percentage distribution of instances in the training and testing set is similar.

The splitting of the database into the two subsets is the end of the first step called Handling data. The training set is used primarily for selecting the model and developing a machine learnt model. In literature machine learning is broadly divided into three main categories, namely Supervised, Unsupervised and Reinforcement learning. The different categories of machine learning along with the usage of their subcategories is visualised in Figure 2.9 and listed as follows:

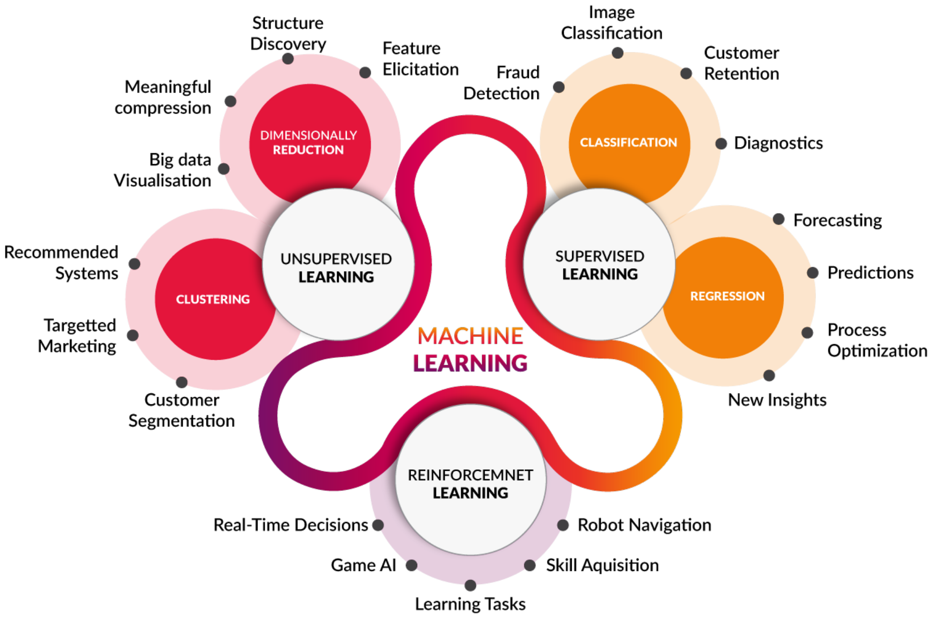


Figure 2.9: An overview of the categories and sub-categories of Machine Learning along with their application in different industries. [47]

- **Supervised learning** uses data that has been segregated into labels and features. The different features are mapped with the label based on a function that is generated, iterated and adjusted until a certain level of accuracy is acquired. This function is generated based on the end utility of the model. either classification, regression or ensembling.
  - Classification based algorithms are used to sort data into categories.
  - Regression is used to make predictions for features either present, similar to or not present in the training data.

- **Unsupervised learning** takes data with labels but no features. Consequently, this algorithm is used to detect unseen trends and relationships in the data set. Two sub categoris are:
  - Clustering algorithms are used to divide and group data based on commonalities.
  - Dimensionality reduction is the process of compressing features in principal values, to convey similar information briefly.
- **Reinforcement learning** is used when the machine is taught to make decisions based on a reward system. It receives information from an environment based on which it takes a decision and each decision is rewarded a negative or positive point. Based on this the machine trains itself by trial and error while assessing the cumulative score.

The objective of the present work is to determine and develop a model capable to *predict or forecast* the feature (C_d) based on a given set of geometrical and thermodynamic labels as shown in Equation (2.27). Classification algorithms can be used to make estimates of C_d, however these algorithms work best for data that is either binary. Unsupervised machine learning can not be used because the label and the features are known. Had the objective been to determine new trends or look for new patterns, this form of machine learning would be appropriate for use. Lastly, Reinforcement learning too can be used, however, the number of data points required to develop a high quality reinforcement learning algorithm is in the order of hundred thousands which is not possible for the thesis and is out of the scope of this these. Consequently, regression algorithms under supervised machine learning is the most suitable choice for the present work.

2.6.2. Regression models for supervised machine learning

Different regression models can be used to make the machine parse and learn data. The most common of these are briefly described below because the models used in this thesis are considered as 'black-boxes', i.e. the input is provided to these models, the mathematics and calculations are trusted, and the output is analysed and assessed.

Linear regression

The linear model draws a linear relationship between the features and the label. It does so by computing the weighted sum of the features along with a bias term, as shown in Equation (2.28). In this equation $\hat{y}$ is the predicted value, θ_0 the bias term, x is the feature and n represents the number of features where θ_n is the feature weight.

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + .. + \theta_n x_n \tag{2.28}$$

It is more convenient to express Equation (2.28) in a vectorised form, which is shown in Equation (2.29), where θ is the parameter vector of the model and **x** the feature vector and h_θ is the hypothesis function.

$$\hat{y} = h_\theta(\mathbf{x}) = \theta^T \cdot \mathbf{x} \tag{2.29}$$

To develop a trained model the values for the parameter vector need to be determined such that the model fits the training set. How well the model fits the training set is determined by a performance parameter, such as the Mean Square Error (MSE). The MSE is calculated as per Equation (2.30), where superscript (i) represents a single row of data in the training data set and m represents the total number of such data entries in the training data set. Thus, the objective is to determine parameter values such that the MSE is as low as possible.

$$MSE(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta^T \cdot \mathbf{x}^{(i)} - y^{(i)}\right) \tag{2.30}$$

Polynomial regression

This model first adds more features by including the power of each feature as a new feature and then train a linear model on this extended set of features. A set of features is transformed into n_{total} number of features, as per Equation (2.31). In this equation n is the number of features and d the degree. While this makes the computation more complex it leads to a more accurate fit compared to linear regression. However, on the other hand it is easy to develop a model that is overfitted. This is resolved by introducing more data to the training set.

$$n_{total} = \frac{(n+d)!}{d!n!} \tag{2.31}$$

Regularization models

Another method that can be used to prevent a model from being overfitted and constraining the model to make it simpler is called regularization. For linear models this means constraining the weights and for polynomial it could be as simple as decreasing the degree. The three forms of regularization that are discussed below are Ridge, Lasso and Elastic Net.

Ridge regression adds a regularization term, the right most term in Equation (2.32), to the cost function Equation (2.30). In this equation α controls the extent to which regularization wants to be added. A value of 0 implies linear regression and higher values force the weights as small as possible.

$$J(\theta) = MSE(\theta) + \alpha\frac{1}{2}\sum_{i=1}^{n}\theta_i^2 \tag{2.32}$$

Lasso (Least Absolute Shrinkage and Selection Operator Regression) regression is another form of regularization. It is similar to Ridge regression except the term that is added is $\alpha\sum_{i=1}^{n}|\theta_i|$. Given that this term is the l_1 norm of the weight vector, the model tends to get rid of the weights of the least important features. So for instance, if polynomial regression is used for data that can be modelled with linear, lasso regression will have a tendency to eliminate the higher degree polynomial features. Basically, it automatically performs feature selection.

Elastic Net is the combination of both Ridge and Lasso regression. This model includes a mix ratio term, r, that controls how much of Ridge or Lasso is to be included. As can be seen from Equation (2.33) when r = 0 or 1 it means purely Ridge or Lasso regression respectively.

$$J(\theta) = MSE(\theta) + r\alpha\sum_{i=1}^{n}|\theta_i| + \frac{1-r}{2}\alpha\frac{1}{2}\sum_{i=1}^{n}\theta_i^2 \tag{2.33}$$

Support Vector Machines - Regression (SVR)

The application and working idea of Support Vector Machines (SVM) can best be explained with the help of Figure 2.10. As can be observed that SVM can be used for classification or regression problems. Support Vector Classification (SVC) works by determining a line in the case shown in the figure or hyperplane in a multi-dimensional problem that separates classes of data (triangles and circles). It attributes either a positive or negative value to either side of the line/hyperplane and uses this value to classify new points in different classes.

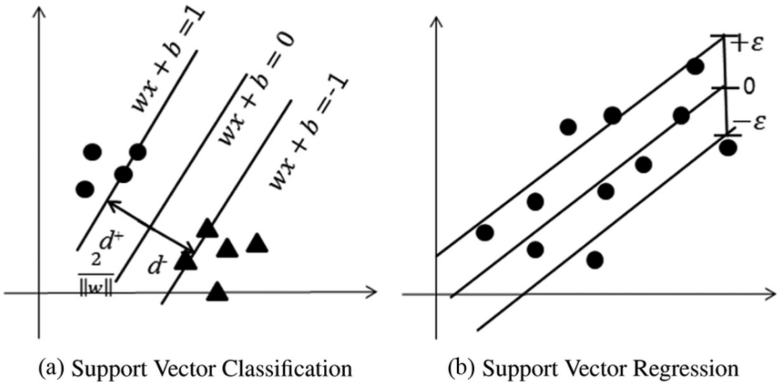


Figure 2.10: Schematic explanation of the sub-categories of Support Vector Machine algorithms. (a) Support Vector Classification. Classification done by determining a line that segregates two classes and assigning positive and negative values to either side. (b) Support Vector Regression. Regression performed by determining a line that includes as many points as possible within the defined acceptable margin of error (ϵ). [48]

The model for Support Vector Regression (SVR) is built like the case for polynomial or linear regression. However, it provides one flexibility to define what margin of error (ϵ) is acceptable. Accordingly it determines a line/hyperplane that attempts to fit as much data as possible in this $\pm\epsilon$ margin. Therefore, the objective is to minimise the l_2 norm of the coefficient vector.

Decision trees

Just like SVM, these models work for both regression and classification problems. The data is first split on the basis of a value or classified into two subsets based on a feature and a threshold. It uses the Classification and Regression Tree (CART) algorithm to do so. It recursively performs this process of picking a new feature and threshold value till it cannot find a split for the remaining training instances. While it might seem that decision trees are apt for binary data, it can work with both discontinuous as well as continuous data. For continuous data the algorithm splits the entire data into multiple regions and takes the average of that region as the predicted value. It adjusts these regions such that maximum training instances can be close to that predicted value.

Neural networks

This model is based on the neural network in the brain, so there are different nodes, each capable of processing data and these nodes are interconnected to other nodes. Additionally the flow of data is per layer. One layer is the simplest architecture called, the perceptron and multiple layers are called multi-layer perceptron architectures. The first layer is the input layer which contains as many nodes as features. The second layer might contain a different number of nodes and the final layer eventually contains either one or multiple output nodes. Each of these nodes within the input and output layer contains a function that computes the weighted sum of each of the input to that node and the goal of the machine learning algorithm is to determine the correct weight.

There are other models such as Logistic regression and k-Nearest Neighbour (kNN), or models that are combinations of the models discussed about such as Random Forests, which is an ensemble of Decision Trees, but are not discussed in the present work as discussing all of them is out of the scope of this work. Additionally, the discussion of some models such as kNN and Logistic regression is considered trivial because even though these can be used for regression but are more suitable and commonly used for classification problems.

2.6.3. Selecting a model

Training and evaluating models

The data prepared using the Handling data steps is used on different algorithms. The performance of each algorithm can be measured using several statistical tools, such as mean deviation, percentage error, R-square parameter, Root Mean Square Error.

The R-square and RMSE metrics are selected for the evaluation of different algorithms in this present work. This is because R-square is a measure of how well the regression model fits the data on the regression curve and the RMSE is the standard deviation of prediction errors.

The comparison of the performance of different algorithms on the training set is done on the basis of the RMSE and R-square values. The model with the highest R-square and lowest RMSE is considered to be the best performing model.

Validation of model

Once a model for application is finalised. It is exposed to the data of the testing set. The same evaluation metrics as that of the training model are used to determine the accuracy in prediction of new data. If the metrics are desirable, such as R-square > 0.9 and RMSE is low, then the machine learnt model can be finalised. If the metrics are low or unsatisfactory then new training models have to be found and all the steps have to be repeated.

2.7. Summary

The focus of the present work is to estimate losses of turbomachines operating with organic fluids close to the critical conditions. More specifically, the emphasis is on estimating boundary layer losses. Laminar and turbulent boundary layers are well understood by simplifying the problem to 2D. This has in-turn eased the development of methods to quantify them relatively accurately. The main challenge in estimating the blade boundary layer loss coefficient is in determining a detailed velocity distribution and transition point apriori. A constant value of 0.002 is used for C_d. While the trends shown in Figure 2.6 for organic fluids might be the same the constant value might be lower. The objective is to use supervised machine learning to develop models that can help predict the C_d distribution, or boundary layer loss coefficient as a function of parameters, namely $\phi, \psi, \gamma\alpha Z$ and working fluid.

3

Methodology

This chapter presents the research method followed in the study to generate a database of boundary layer dissipation coefficient, C$_{d}$, and to develop a machine learnt model. To differentiate between the dissipation coefficient in and the dissipation coefficient for the stator blade, from here on C$_{d}$ is referred to as C$_{d,blade}$. Each instance of data about C$_{d,blade}$ is generated from a unique combination of input parameters, namely duty coefficients, blade geometrical variables , working fluid and boundary conditions for a 2D RANS CFD simulation. The unique combination of input parameters is called a test case in the present study. The range of values of each input parameter together dictate the number of test cases and the size of the database. Table 3.1 shows the range of values used for each input parameter in the present study and the influence of the range on the number of test cases. The reason for selecting the range for each parameter, shown in Table 3.1, is argued in this chapter as well.

Table 3.1: Range of each parameter and its influence on the total number of test cases.

Parameter	Range	Number of test cases
ϕ	0.5 - 0.9 (in steps of 0.1)	5
ψ	0.8 - 2.0 (in steps of 0.1)	65
σ	6 sets	390
$fluid$	N_2, CO_2, Toluene, MM	1560
Z	0.7, 1.0	3120
α	2,3,4	9360

Section 3.1 provides an overview of the methodology devised to generate a single instance of data and the steps involved to develop a machine learnt model. Section 3.2 describes the steps taken to generate a blade geometry, the criteria used to mesh the domain in which this geometry is simulated, the SU2 solver settings used to simulate this geometry and the steps executed to calculate the boundary layer dissipation coefficient for this geometry. The steps described in Section 3.2 are repeated for the each test case listed in Table 3.1. The reason behind selecting the specified range of values in Table 3.1 for each parameter is argued in Section 3.3. Finally, Section 3.4 discusses the research method followed to develop a machine learnt model capable of predicting C$_{d,blade}$ as a function of parameters listed in Table 3.1.

3.1. Overview

Recall that there are two main objectives of the present study. The first is to determine which parameter out of the ones listed in Table 3.1 influences blade boundary layer losses significantly. The second is to build a machine learnt model capable of predicting C$_{d,blade}$. As mentioned in Section 2.6, a machine learnt model uses algorithms to dissect data, learn from it and make predictions. So data and algorithms are two fundamental requirements to build a machine learnt model. The methodology used for the present work is divided on the basis of these two fundamentals.

Figure 3.1 shows the schematic of the methodology in the form of a flowchart. There are two main blocks namely, Data Generation and Machine Learning. The prior block contains the procedure to determine C$_{d,blade}$ from a given set of inputs, namely $\phi, \psi, \sigma, fluid, Z$ and α. The Data Generation block is used recursively with different inputs to generate a database of corresponding C$_{d,blade}$ values. The Machine Learning block utilises this database for algorithms to learn and make predictions from. To explore the most suitable algorithm for the trained model the entire Machine Learning block does not have to be repeated and is used non-recursively. Each step shown within the two blocks are explained in Section 3.2 and Section 3.4.

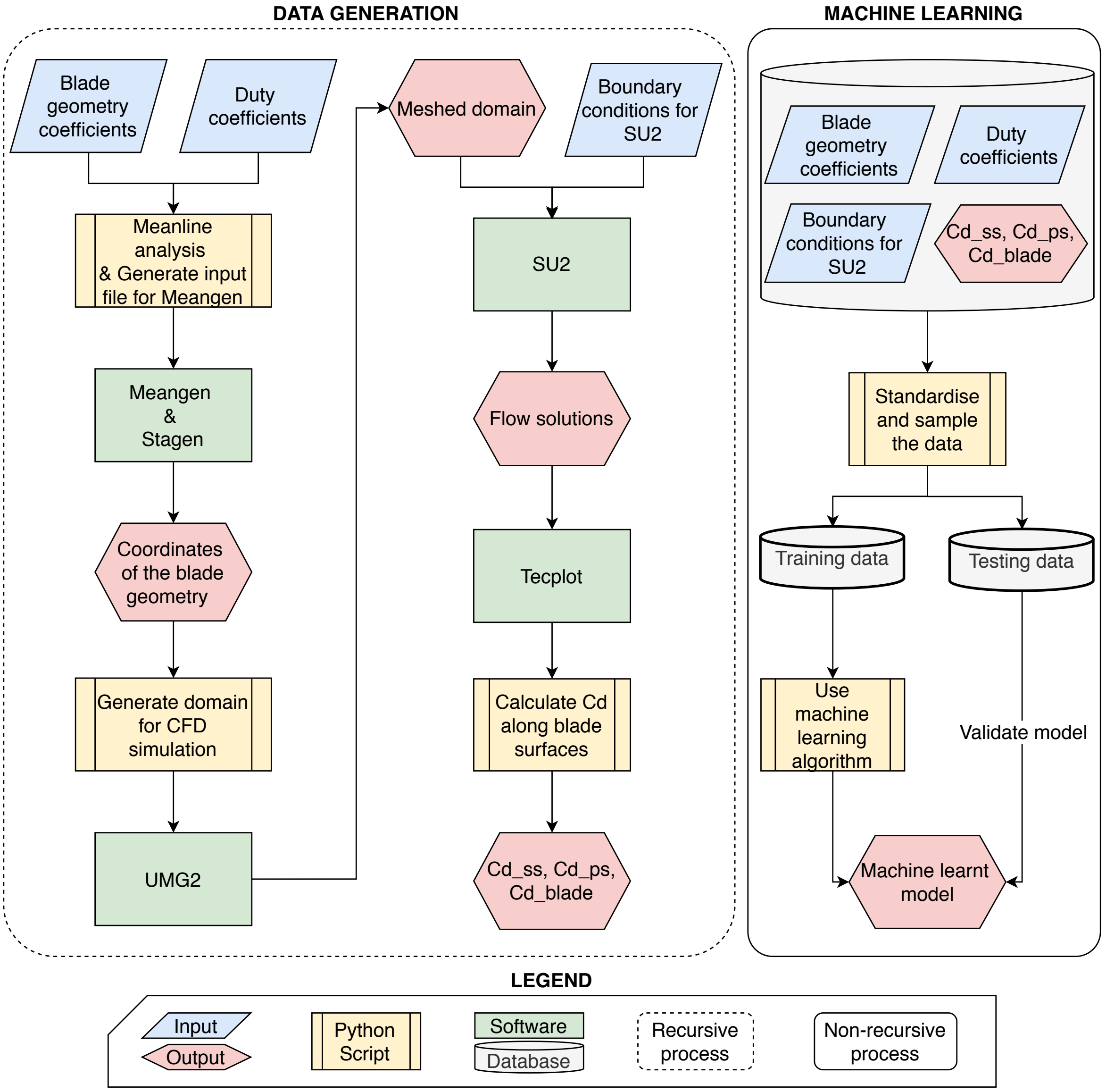


Figure 3.1: Schematic of the methodology required to develop the machine learnt model for the present study. The methodology is split into two blocks, namely Data Generation and Machine Learning. The first block describes the procedure to generate a single instance of data and the entire block is used recursively with different inputs to generate a database. The Machine Learning block uses the database to develop a machine learnt model.

3.2. Data generation procedure

The focus of this section is to explain each step shown in Data Generation block in Figure 3.1. The method adopted to generate a blade geometry, mesh it, simulate it and calculate C$_{d,blade}$ from its flow solution is explained.

3.2.1. Blade geometry generation

To generate blade geometries two inputs, namely duty coefficients and blade geometry coefficients, are required. These coefficients are used to generate a meanline design of the blade using a Python script. Additionally, the Python script creates the input file that is required to initiate the blade geometry generation tool, Meangen and Stagen. The output of Stagen is Cartesian coordinates of a mid-span two dimensional axial stator blade.

Inputs - Blade geometry and Duty coefficients

The blade geometry is expressed as a function of the duty coefficients and other coefficients that describe the blade shape.

The duty coefficients dictate the architecture and the performance of the turbine stage. ϕ indicates the flow capacity of the stage, ψ measures the work capacity of the stage and $r*$ is a measure of the static expansion across the rotor. Additionally, these three determine the inflow and outflow angle of the blades. As the present work considers only first stage stator blades for the analysis, two assumptions regarding duty coefficients are made. First, $r*$ = 0.5 for all blade geometries considered. Second, the inflow angle is assumed to be zero. Therefore, ϕ and ψ are used to calculate the outflow angle of the stator.

The blade is built around a basic camber line, which is a parabolic arc and forms the skeleton of the blade. The blade geometric coefficients are coefficients that help define the blade shape around the camber line. These coefficients influence the flow over the blades and either contribute to or decrease boundary layer losses and therefore are important to consider. Some examples are camber angle, stagger angle, radii at leading an trailing edges, thickness distribution, maximum thickness to chord ratio and camber line shape. Maximum thickness to chord ratio (tc) and thickness distribution (sf) are the two coefficients that are considered for the present analysis. These two coefficients are selected on the basis of a design of experiments study that is described in Section 3.3.2.

Meanline analysis and inputs for Meangen

Table 3.2: The most relevant parameters that are inputs to Meangen software. These inputs are a mix of stage duty coefficients, thermodynamic variables, dimensionless blade geometric coefficients and machine performance parameters.

Parameters			
R_{gas}	Specific gas constant	ω	RPM
$r*$	Degree of reaction	sf	Blade thickness distribution
γ	Specific heat ratio	$\dot{m}$	Mass flow
ΔH	Enthalpy drop across stage	tc	Max. thickness to chord ratio
P_{in}	Inlet pressure	α_0	Absolute inlet flow angle
c_{ax}	Axial chord length	η	Guess isentropic stage efficiency
T_{in}	Inlet temperature	α_1	Absolute outlet flow angle
$\frac{t_{te}}{c_{ax}}$	Trailing edge thickness to axial chord	σ	Pitch to chord ratio

A meanline analysis program is developed in Python for the present work. This program creates a meanline design and calculates intermediate measurements such as enthalpy drop across stage, outflow angles or blade solidity of the blade based on duty coefficients and geometry coefficients. These intermediate measurements along with the parameters tabulated in Table 3.2 consist of the most important inputs required by Meangen. The program arranges these inputs in the required format for Meangen and helps verify the results produced by Meangen and Stagen.

Meangen and Stagen software

The coordinates of a blade are generated using the software suite of Multall. The suite consists 3 programs, namely, Meangen, Stagen and Multall. For the present work only the prior two are used. Meangen is a meanline turbomachinery design program that performs a 1D design to obtain velocity triangles on a specified stream surface. It makes estimations regarding blade shape and blade height based on inlet design parameters mentioned in Table 3.2. The output file of Meangen is the input file for Stagen. Stagen is a blade geometry generation and manipulation program which transforms the initial 1D blade geometry designed by Meangen. It refines the resolution of the blade at the bottom mid and tip span, stacks them together and combines them to generate a 3D design of the blade [49]. The Cartesian coordinates of the 3D blade are outputted and stored in a file called *blade_profiles.dat*.

Output - Coordinates of the blade geometry

The ouput file of Stagen, *blade_profiles.dat* is manipulated such that the coordinates of the stator stacked at mid-span height are extracted and stored in a file called *stator_profile.dat*. This file contains the mid-span slice of the blade that is expressed in 200 2D Cartesian coordinates.

3.2.2. Blade geometry meshing

To simulate free-stream flow in the stator blade passage and capture the interaction of the working fluid with the blade surface, first the blade geometry is captured in a domain. Next, this domain is meshed. Finally, this meshed domain and the defined boundary conditions are used to perform the flow simulations.

Generate domain for CFD simulation

The domain for the blade shapes is generated using a set of rules. These rules are devised to ensure uniformity in the domain generation procedure for all stator blades. These rules are as follows:

- The leading edge of the blade is placed at the origin (0,0).
- The inflow boundary condition is placed one pitch upstream of the leading edge.
- The upper and lower boundaries are the periodic boundaries and are placed half a pitch above and below prior and post the blade.
- The curve of both the periodic boundaries is based on the camber of the blade.
- The outflow boundary condition is placed 1/4th pitch downstream of the trailing edge of the blade.

An example of a domain with different boundaries is shown in Figure 3.2.

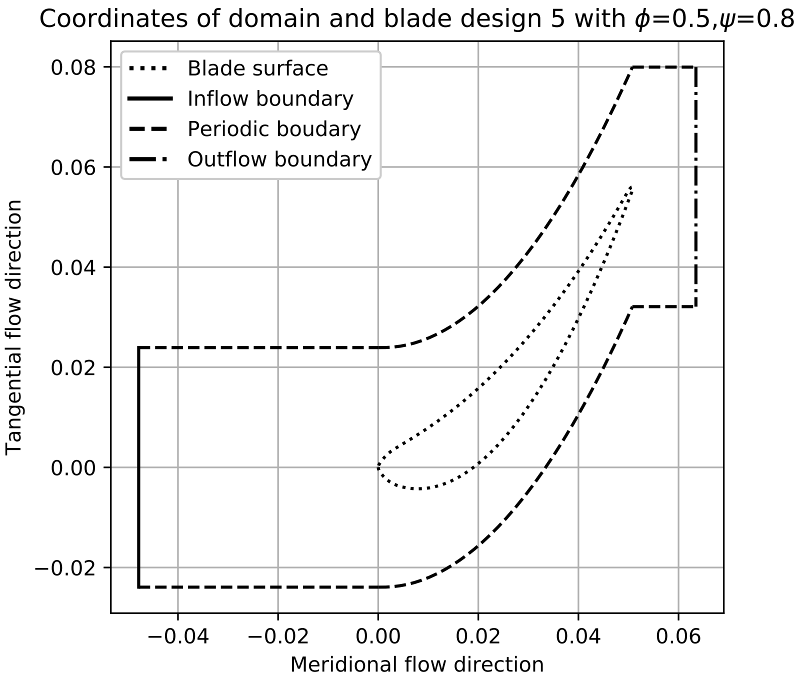


Figure 3.2: An example of the domain for a blade profile with $\phi = 0.5, \psi = 0.8$. Design 5 corresponds to blade geometrical parameters $tc = 0.2$ and $sf = 1.95$.

Meshing using UMG2

Each domain encompassing the blade geometry is generated and meshed using an in-housing meshing program called UMG2 [50]. UMG2 requires three input files. The first file contains the coordinates of the blade walls and domain. The second file labels the inlet, outlet and periodic boundaries of the domain. The third file contains details about the variation of cell type within the domain and the size of the cells.

Table 3.3: Values of the minimum and maximum node size and number of nodes per radius of curvature used in the spacing control input file for UMG2 for all the blade designs.

Boundary	h_{min} **[mm]**	h_{max} **[mm]**	**Nodes/r**$_c$
Inflow	2.9	2.9	5
Outflow	2.9	2.9	5
Blade surface	0.1	0.1	15
Periodic Upper	2.9	2.9	5
Periodic Lower	2.9	2.9	5

The size and skewness of cells influences the rate of convergence, accuracy of the solution and the CPU time required. Higher cell counts give higher accuracy but at the cost of CPU and memory usage. To prevent this non-uniform grids, that cluster cells only where required can be created. Additionally, pertinent flow features such as the boundary layers can be captured using a grid adaptation strategy which adds more cells to resolve the flow field. These properties of the mesh are dictated by providing the minimum and maximum spacing between grid nodes (h$_{min}$ and h$_{max}$) or the grid nodes per radius of curvature ($Nodes/r_c$).

Again, to ensure uniformity in meshing the values of the mesh properties shown in Table 3.3 are selected for all the blade designs. These values are found using a grid convergence study.

These values ensure that there is a higher density of quadrilateral cells clustered close to the blade wall and a y^+ « 1 is maintained. The rest of the domain uses triangular elements.

Output - Meshed domain
Figure 3.3 shows an example of a meshed blade design using the mesh properties discussed in the previous section. The upper left hand corner of the figure shows the fine mesh close to the blade surface. The rest of the domain has a triangular elements which increase in size when further away from the blade surface.

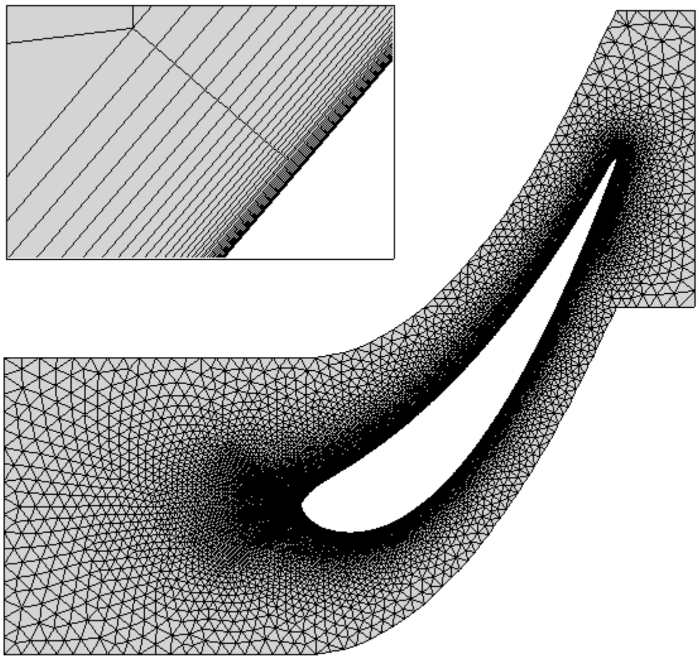


Figure 3.3: Unstructured meshed domain for blade profile with $\phi = 0.5, \psi = 0.8$. Design 5 corresponds to blade with geometrical parameters $tc = 0.2$ and $sf = 1.95$.

3.2.3. Fluid solver execution

To initiate the simulation the physical model regarding equations of fluid motion and boundary conditions are defined.

Input - Boundary conditions for SU2
The boundary conditions define the fluid behaviour and fluid properties at all bounding surfaces of the meshed domain. The inlet reduced temperature and pressure determine the value of the compressibility factor at the inflow boundary, while the total-to-static volumetric flow ratio is used to determine the outlet static pressure at the outflow boundary. The selection of all the boundary conditions used for the present work is discussed in Section 3.3.

Simulation using SU2
The fluid solver used for the present work is the open-source SU2 suite [51]. SU2 requires 2 input files to perform CFD calculations, the first is a file with the meshed domain and the other is a file with the configuration of solver settings. The most relevant solver settings used are listed below.

- SU2 provides Euler and variants of Reynold's Averaged Navier Stokes (RANS) as flow solvers. RANS is considered to be the suitable choice in terms of capturing the effect of the working fluid and the computational cost.

- The 2 options to model the turbulence are the two equation Menter's Shear Stress Transport (SST) and the one equation Spalart-Allmaras (SA). The SA model is used for all the simulations in the present work.

  The selection between SA and SST is made by comparing the C$_d$ values obtained from simulations of a flat plate to the correlation proposed by Schlichting (C$_d$ as a function of Re$_\theta$). Figure 3.4 and Figure 3.5 show the results and errors obtained from SA and SST simulations respectively. The performance of SA improves as the boundary layer develops across the flat plate. The errors are more than 5% for the first three measurements, which represents the start of the flat plate. On the other hand, the SST model has errors less than 5% only for Re$_\theta$ between 750 and 6000. This is substantiated by the findings of the study by Otero and Pecnik. In their study, they state that the SAs model is the most suitable to analyse dense vapour flows [52].

- The two equations of state used for the present work are Ideal Gas and Peng-Robinson. The latter is used to capture flow non-ideality.

- The working fluid is characterised via its thermodynamic properties in SU2. These are R$_{gas}$, T$_{cr}$, P$_{cr}$, Acentric factor, γ, thermal conductivity and viscosity. These values for the different working fluids are taken from CoolProp thermodynamic library [53]. The Peng-Robinson model implemented in SU2 is based on the assumption that isentropic processes are completely polytropic and thus takes in a constant γ value. This constant value is calculated using $C_{p,ideal} = \frac{R_{gas}\gamma}{\gamma-1}$, where C$_{p,ideal}$ is the ideal specific heat at constant pressure and its value is taken from CoolProp as well.

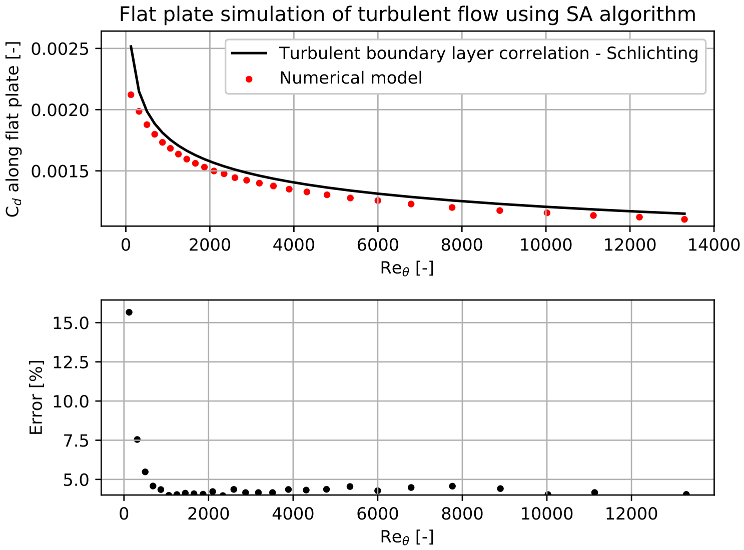


Figure 3.4: Performance of SA (Spalart–Allmaras) turbulence model implemented in SU2 on a flat plate simulation. The measurements of C$_d$ along the flat plat, obtained from the simulations, and the theoretical turbulent boundary layer correlation for C$_d$ devised by Schlichting is shown in the plot above. The % error between the theoretical correlation and the simulation results for each measurement along the flat plate is shown in the plot below.

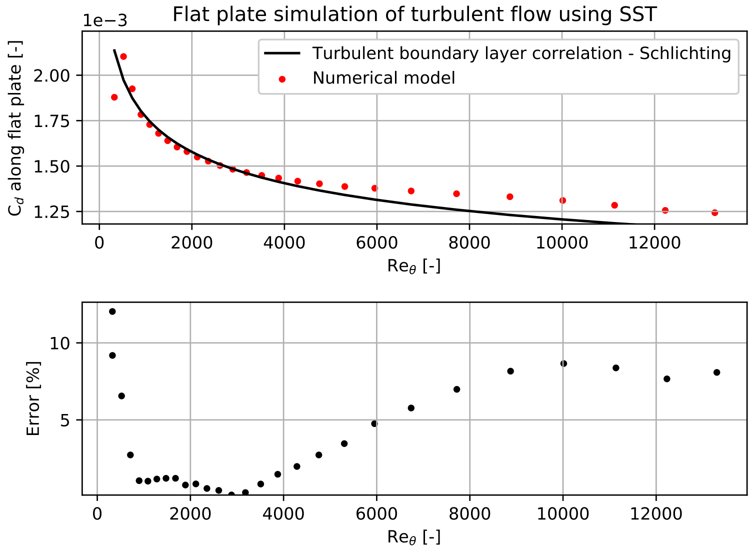


Figure 3.5: Performance of SST (Menter's Shear Stress Transport) turbulence model implemented in SU2 on a flat plate simulation. The measurements of C_d along the flat plat, obtained from the simulations, and the theoretical turbulent boundary layer correlation for C_d devised by Schlichting is shown in the plot above. The % error between the theoretical correlation and the simulation results for each measurement along the flat plate is shown in the plot below.

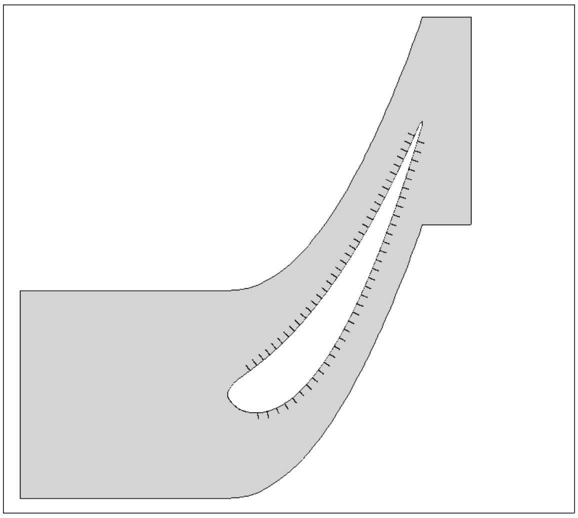


Figure 3.6: Discretisation of the suction and pressure side of the blade into segements demarcated by lines normal to the blade surface. There are 35 normals on each side of the blade placed from 10% to 95% of the suction/pressure side blade surface length.

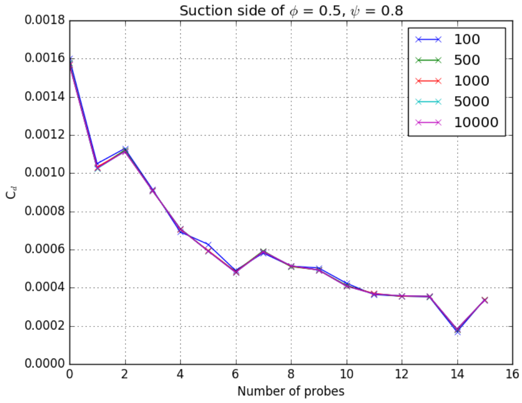


Figure 3.7: Comparison of the distribution of C_d, along the suction side of a blade, obtained by changing the number of measurement points along each normal. The blade geometry is characterised by Design 1 and ϕ, ψ = 0.5, 0.8 and simulated with Toluene and Z = 1 as fluid and inlet conditions.

Except for varying the test case specific parameters (inlet conditions and thermodynamic parameters of working fluid) all other settings are held constant for simulating all the test cases.

Output - Flow solutions

The results of the flow solutions are stored in two output files, namely *surface_flow.dat* and *flow.dat*. The prior contains measurements at the surface of the blade and the latter of the domain. The two files are interpreted using Tecplot 360 to calculate C_d along the suction, pressure and entire blade surface.

3.2.4. Computation of dissipation coefficient

To solve the integral for the total rate of entropy generated per unit surface area (shown in Equation (2.8)) first the region close to the blade surface is discretised. Second, the boundary layer height is determined. Eventually, values at the edge of the boundary layer are substituted into Equation (2.9) to determine the C_d distribution along the blade surface.

Flow solution interpretation using Tecplot

The flow solutions are post processed using Tecplot because the PyTecplot feature allows data to be accessed using Python script [54]. The use of customisable macros eases automation of calculations and extraction of data from the CFD flow solution files. The detailed description of the steps taken to estimate C_d are explained in the subsequent section.

Calculate C_d along blade surface

The steps taken to calculate the dissipation coefficient on both the suction and pressure side are as follows:

1. The blade surface coordinates are split into the pressure and suction side because the procedure to evaluate both sides is the same. The starting coordinate of both sides is at the origin, as shown in Figure 3.2, and the end is at the mid point of the trailing edge.

2. Each surface side is discretised into fixed segments starting from 10% to 95% of the blade length. Each segment is 2.5% of the blade length and is demarcated by normals to the blade surface. An example of this is shown in Figure 3.6. The height of each normal is uniform and adequate enough to capture the boundary layer thickness. Measurement points along each normal have been distributed exponentially, with higher concentration of measurement points close to the surface to capture the innermost layer of the boundary layer.

   The number of points per normal dictates the computational time and the resolution of the boundary layer. Figure 3.7 shows the results of varying different number of measurement points along the normal. Based on these results 500 measurement points per normal is selected.

3. Recall, the boundary layer thickness is defined as the distance normal from the wall to a height where the velocity profile has reached the asymptotic velocity. The first measurement point that either reaches the maximum of all the recordings or within 99% of the free stream velocity from the surface along each normal is defined as the edge of the boundary layer. $s_{\delta_e}, \rho_{\delta_e}, T_{\delta_e}$ and v_{δ_e} are extracted at this edge to solve Equation (2.8) and Equation (2.9).

4. The automation of the boundary edge detection and the numerical implementation of Equation (2.8) and Equation (2.9) leads to fluctuations in the distribution of C_d obtained from the 35 normals along both the blade surfaces. This statement is exemplified using Figure 3.8. This figure plots the actual and best fit C_d distribution per surface side of a test case. The distributions without the oscillations are best fit curves of polynomial order 3. From the best fit curve of the pressure surface a C_d value of 0.0015 can be approximated. In Figure 3.8 the best fit curve of the suction surface is calculated only till the 24^{th} normal. This is because of shockwaves impinging on the boundary layer on the suction surface in the flow passage.

5. Recall that one of the objectives of the present study is to analyse the influence of the volumetric flow on the dissipation coefficient. Consequently, several test cases involve

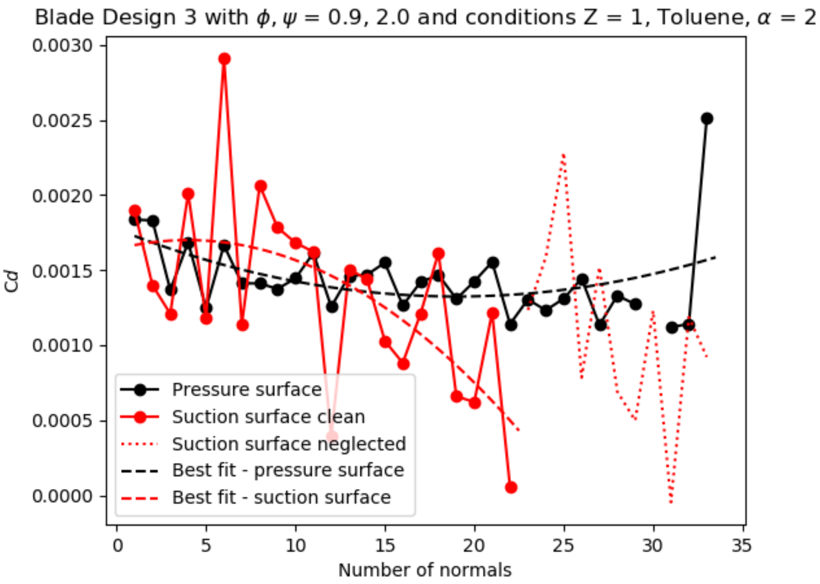
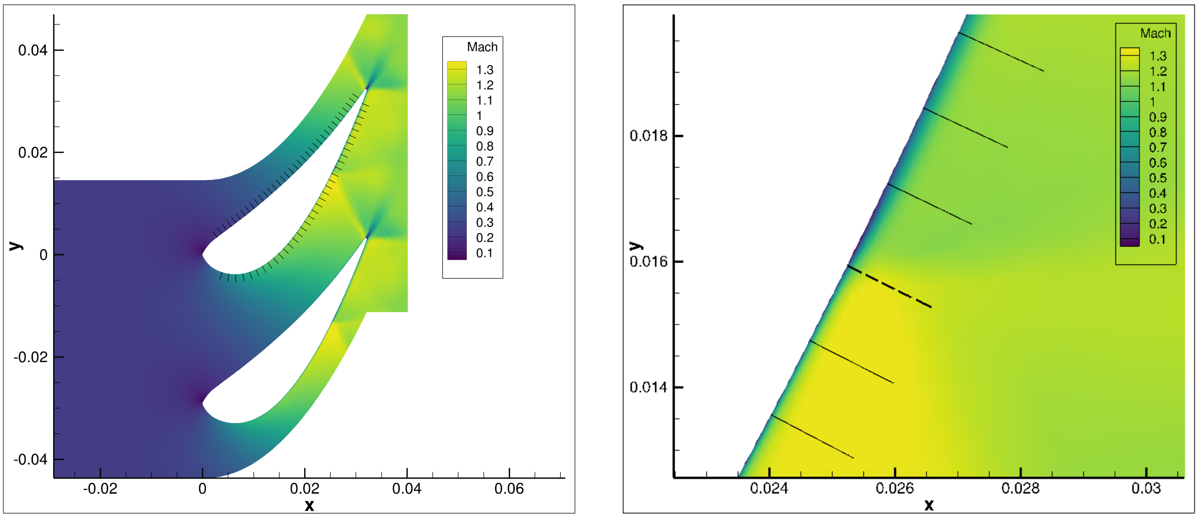


Figure 3.8: Fluctuation of C_d values along the blade surfaces due to numerical implementation of theoretical equations to determine C_d. A best fit polynomial curve of order 3 is used to generalise the fluctuations and determine C_d value per blade surface side. The result is for blade profile with $\phi = 0.9$, $\psi = 2.0$ and Design 3 corresponds to blade geometrical parameters $tc = 0.22$ and $sf = 1.75$.

simulating the stator blade with extreme total-to-static volumetric ratio. This leads to the generation of shock waves in the blade passage. This is exemplified with the simulation of a blade with Toluene as the working fluid, Z at inlet = 1.0 and α = 2.0. Figure 3.9a and Figure 3.9b show the Mach contour flow field of this simulation and the shock interaction with the boundary layer of the suction side respectively.

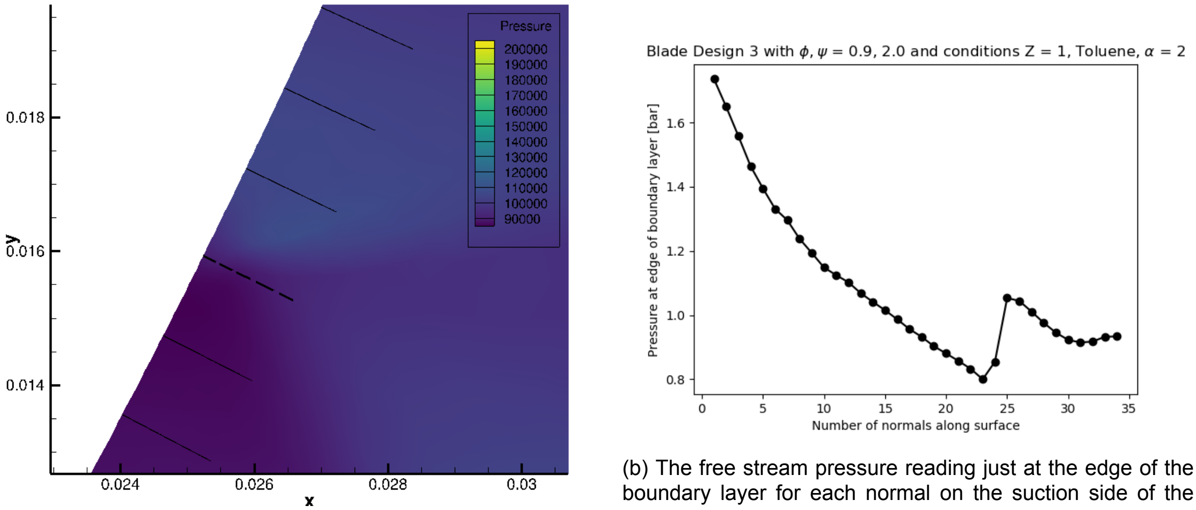


(a) Mach contour for the cascade.

(b) Boundary layer - shock interaction on a Mach contour.

Figure 3.9: Mach contours for blade with parameters ϕ = 0.9, ψ = 2.0 and Design 3, and simulation conditions Toluene, Z = 1.0 and α = 2.0 for the entire cascade in (a). (b) shows the local point of interaction between the shock and the boundary layer. The dotted normal is the first after which all normals are excluded from consideration as the focus of the work is solely on loses due to the boundary layer.

As the focus of the present work is solely on boundary layer losses all normals beyond the shock impingement are neglected from consideration. This is visually explained with the help of Figure 3.9b. The dotted normal in this figure is the normal beyond which the measurement of the other normals are neglected. To automate the procedure of finding shocks in the flow passage for any test case, the gradient of the pressure at the edge of the boundary layer is used. This is because pressure is inherently related to the Mach number. This is visible from Figure 3.10, where Figure 3.10a and Figure 3.10 are the pressure contour and the pressure readings at the edge of the boundary layer for each normal. The sudden change in the gradient of pressure recorded by normal 23 to 26 indicates the presence of a shock. It is because of this that the best fit of the suction surface is calculated only till the 24^{th} normal in Figure 3.8.



(a) Boundary layer - shock interaction on a pressure contour.

(b) The free stream pressure reading just at the edge of the boundary layer for each normal on the suction side of the blade.

Figure 3.10: Pressure as an indicator of the presence of a shock in the flow field. (a) shows the influence of boundary layer shock interaction on pressure contour for a blade with parameters ϕ = 0.9, ψ = 2.0. Design 3 bldae geometrical parameters $tc = 0.22$ and $sf = 1.75$, and simulation conditions Toluene, Z = 1.0 and α = 2.0. (b) shows the trend for the pressure at the edge of the boundary layer along the suction surface of the blade. The jump in pressure reading at immediately after normal 23 indicates the presence and influence of a shock on the boundary layer.

6. The dissipation coefficient for the entire blade, $C_{d,blade}$, is the sum of the weighted average of the C_d distribution for each surface side. This is because Denton states that the total entropy generated due to boundary layer losses is the entropy generated on both surfaces [2]. Furthermore, as C_d is not evaluated beyond the point of shock impingement, the C_d calculated in the present work is representative of the entropy generated solely due to the boundary layer.

Output - $C_{d,ss}$, $C_{d,ps}$, $C_{d,blade}$
The final output of the Data Generation block is dissipation coefficient for the suction side, pressure side and the entire blade. These values correspond to input parameters, namely duty coefficients (ϕ, ψ), blade geometry coefficients (σ), boundary conditions (Z, α) and working fluid. Together the inputs and $C_{d,blade}$ are appended to a database. The range of inputs used for the present analysis is the topic of the following section.

3.3. Selection of test cases

The following section discusses the criteria used to select the range of test cases for the present analysis. Each test case is a unique combination of a value of the input parameters listed in Table 3.1. First the range of duty coefficients and blade geometric parameters considered are discussed. Eventually, the range of boundary conditions, such as inlet reduced temperature, pressure, total-to-static volumetric flow ratio and working fluid, selected is discussed.

3.3.1. Selection of the range of duty coefficients

The flow, stage loading and degree of reaction coefficients are non-dimensional similarity parameters used to identify the architecture of the turbine and to preliminary evaluate its performance during the design phase. Recall that it is assumed that $r*$ = 0.5 for all the blades that are generated to populate the database. For ϕ and ψ, the Smith Chart is considered as the basis for the design space selection. Both coefficients are taken in steps of 0.1. The range for ϕ is 0.5 to 0.9 whereas that of ψ is 0.8 to 2.0. Figure 3.11 visualises this range on the Smith Chart. The selection of this range is attributed to the limit imposed by the blade design tool Meangen and the time constraint for the present work.

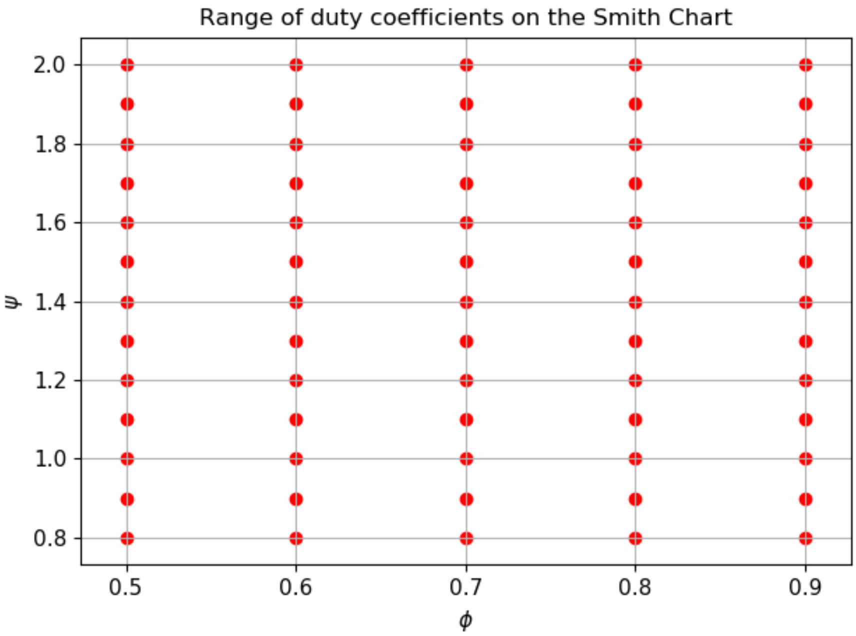


Figure 3.11: The range of the duty coefficients considered to design test cases and populate the database.

3.3.2. Selection of geometrical parameters and its range

Several geometric parameters that are used to identify and design a blade profile are listed in Section 3.2.1. Considering all of these parameters is out of scope for the present analysis because it significantly increases the number of test cases and subsequently the computational time. A similar problem is faced if optimisation algorithms are used to optimise the blade design for each test case. As the present work is first of its kind the idea of performing optimisation is dropped in exchange for a design of experiments study on a few selected geometrical parameters.

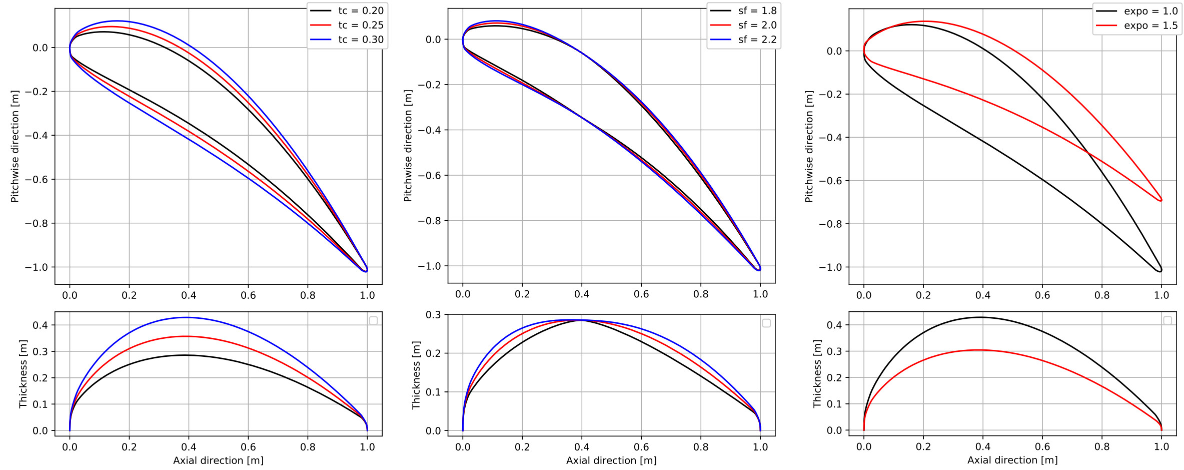
The selection of these parameters is limited by the capabilities of Meangen and Stagen. Table 3.4 lists the parameters that can be manipulated in the software. Out of all the parameters listed, the trailing (t_{te}/c_{ax}) and leading edge (t_{le}/c_{ax}) thickness to axial chord ratios are not considered because the trailing edge contains the wake and the leading edge has a boundary layer of an insignificant thickness. To ensure uniformity the position of the maximum thickness ($x_{t_{le}}/c_{ax}$) is kept constant across all blade designs. As a result, maximum thickness to chord ratio (t_{max}/c_{ax} or tc), shape factor (sf) and camberline shape ($expo$) are selected for the design of experiments study.

Table 3.4: The different controllable geometrical parameters available in Meangen.

Symbol	**Geometric parameter**
t_{le}/c_{ax}	Leading edge thickness to axial chord ratio
t_{te}/c_{ax}	Trailing edge thickness to axial chord ratio
t_{max}/c_{ax} or tc	Maximum thickness to chord ratio
$x_{t_{le}}/c_{ax}$	Fraction of axial chord at maximum thickness
sf	Shape factor or Blade thickness distribution
$expo$	Vary camber line shape

Maximum thickness to chord ratio, Blade thickness distribution and Camber line shape

Figure 3.12 shows how the parameters, namely tc, sf and $expo$, selected for a design of experiments study influence the blade geometry. As $x_{t_{le}}$ is constant, the position of the maximum thickness is fixed at 40% of the axial chord in Figure 3.12a and tc controls the thickness of the blade as a proportion to the axial chord. sf controls the thickness distribution. The baseline default value in Meangen for this parameter is 2.0. Figure 3.12b shows that any value lower than this implies a more triangular distribution, and higher values a more rectangular one. Lastly, $expo$ is responsible for the transformation of the axial position. it is used to vary the camber line shape. The default value is 1 and values lower than this means moving the blade loading downstream and vice versa. This is visible from Figure 3.12c.

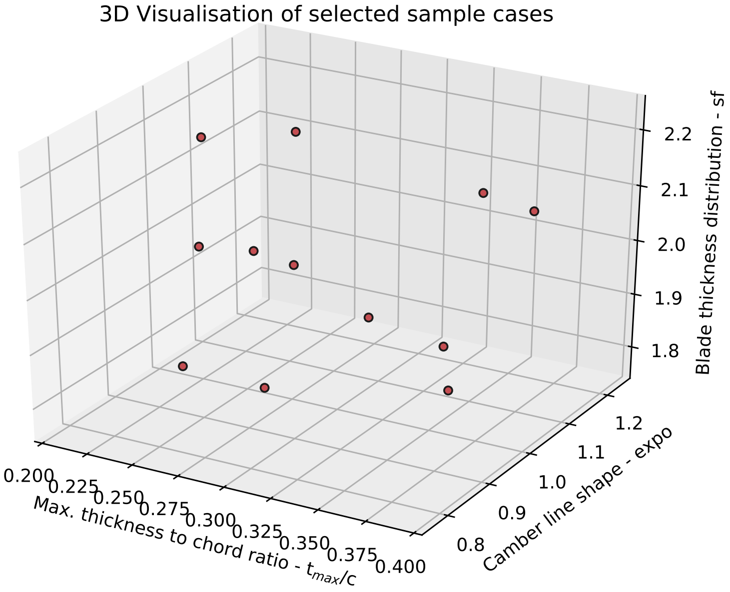


(a) x_{tc} or tc (b) Thickness distribution sf (c) $expo$

Figure 3.12: The influence of the geometrical parameters on geometry and the thickness of the blade.

Design of experiments study

The influence of the blade geometric parameters introduced in the previous subsection on the value of the dissipation coefficient $C_{d,blade}$ is studied. To achieve this, a design of experiments study on the blade geometrical parameters is performed because it helps identify correlations between different parameters quicker than experimenting with one parameter at a time [55]. In this approach, three parameters, namely tc, sf and $expo$, which are the input parameters for the blade generation software Meangen are considered. Among all the combinations of duty coefficients shown in Figure 3.11, only 4 combinations are considered. These are listed in Table 3.5. Two combinations of these, (0.5, 0.8) and (0.9,2.0), correspond to the extreme corners of the design space shown in Figure 3.11, whereas the other two, (0.6, 1.7) and (0.7,1.1), correspond to the mid point of the bottom right and top left quadrant of Figure 3.11. For each of these combinations, 12 distinct values of each parameter tc, sf and $expo$ is considered. These are visualised and tabulated in Figure 3.13. These values are selected using a Latin hypercube sampling algorithm [56]. This algorithm provides coverage of the entire design space by discretising the design space into rows and columns and ensuring there is only one sample per row and column. The result of selecting 12 unique combinations of geometric parameters for each of the 4 combinations of duty coefficients is a total of 12 x 4 = 48 unique blade designs. Each of the 48 blades is simulated in SU2 three times for each of the conditions reported in Table 3.5. These conditions differ in the value of inlet total temperature and pressure, α and the working fluid. T_r and P_r determine the value of the compressibility factor Z, while α determines the outlet static pressure boundary condition. The end result is a database of 48 x 3 = 144 instances.



$\frac{t_{max}}{c_{ax}}$	sf	$expo$
0.30	1.87	0.84
0.27	1.98	1.01
0.37	2.09	0.77
0.38	1.90	0.89
0.35	2.03	1.23
0.33	1.80	1.15
0.20	1.93	1.09
0.24	2.23	0.93
0.26	2.19	1.07
0.22	1.78	0.96
0.30	2.12	0.81
0.33	2.07	1.18

Figure 3.13: A Design of Experiments study of the geometrical parameters tc, sf and $expo$. The visualisation and tabulation of 12 samples selected using Latin hypercube sampling technique. This technique provides good coverage over the entire design space being analysed.

Table 3.5: Flow inlet conditions and combinations of duty coefficients for the design of experiments study of the geometrical parameters tc, sf.

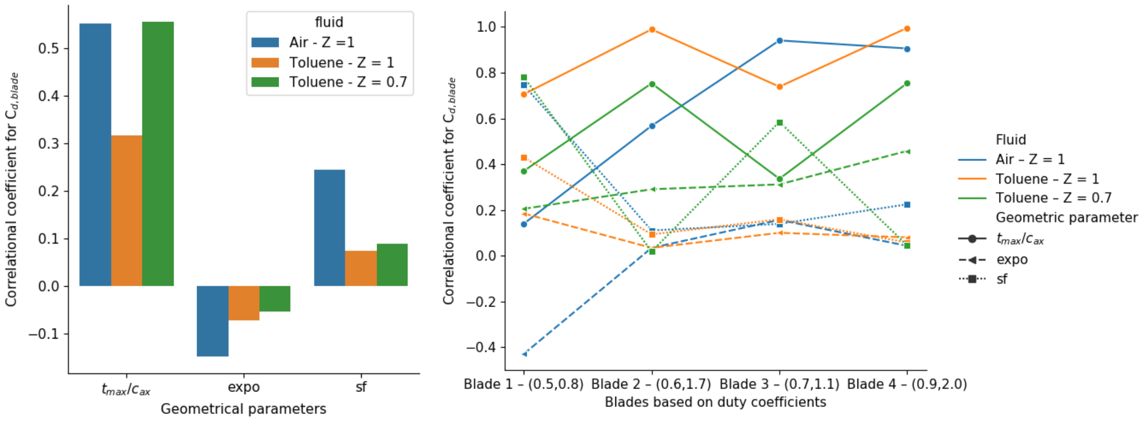
$fluid$	Z	α	(ϕ, ψ)
Toluene	1	2	(0.5,0.8), (0.6,1.7), (0.7,1.1), (0.9,2.0)
Toluene	0.7	2	(0.5,0.8), (0.6,1.7), (0.7,1.1), (0.9,2.0)
Air	1	2	(0.5,0.8), (0.6,1.7), (0.7,1.1), (0.9,2.0)

The dissipation coefficient is calculated for each of the 144 test cases using the Data Generation procedure described in Section 3.2. The influence of each parameter tc, sf and $expo$ on the dissipation coefficient $C_{d,blade}$ is quantified using the Pearson correlation coefficient.

The values of the Pearson coefficient range between +1 and -1, where the sign preceding the number indicates a positive or negative correlation and a value of 0 means no correlation.

Figure 3.14 shows the correlation coefficients when data is grouped on the basis of fluid and grouped on the basis of fluid and set of duty coefficients. According to Figure 3.14a there is a positive correlation between $C_{d,blade}$, tc and sf and a negative between $C_{d,blade}$ and $expo$. Additionally, the magnitude of the correlation coefficient of tc and $C_{d,blade}$ is the highest followed by that of sf and eventually $expo$. This holds for all fluids. Such trends imply that $expo$ is a less significant geometric parameter for C_d than the other two. Figure 3.14b shows the correlation per fluid and per blade design. The solid lines represent tc and depict the strength of tc with C_d. The second strongest correlation is shown by the finely dotted line represents sf, followed by $expo$. Exceptions to these trends are blades with duty coefficient combinations (0.6, 1.7) and (0.9, 2.0), working fluid Toluene and Z = 0.7. These exceptions have stronger correlations for $expo$ than sf because for these cases the duty coefficients influence $C_{d,blade}$ more than the geometric parameters.

From Figure 3.14, it follows that $expo$ has negligible influence on C_d compared to the other two. In light of these considerations, the influence of geometric parameter $expo$ on $C_{d,blade}$ is not considered for the present work. Consequently, it eliminates a weakly correlated parameter to $C_{d,blade}$.



(a) Correlation coefficients per fluid.

(b) Correlation coefficients per blade design.

Figure 3.14: Correlation coefficients, between C_d and geometrical parameters tc, sf and $expo$, obtained from Design of Experiments study. The data is based on CFD simulations performed on 48 unique blade designs (12 sets of geometrical variations and 4 sets of duty coefficients). The 48 blades are simulated with 3 unique inlet flow conditions that are a combination of the fluid or compressibility factor (Z). (a) Depicts the correlation coefficients obtained if the results are segregated on the basis of fluid and (b) on the basis of combinations of duty coefficients.

If the variation of a square design space of sf and tc with 6 unique values each is to be studied, the total number of test cases increases by a multiple of 6^2. This increases the computational power and time requirements. To prevent this, Latin hypercube sampling is used to determine six unique combinations of both the geometrical parameters, sf and tc. Each unique combination is assigned a design number. The results are summarised in Table 3.6

3.3.3. Final selection of duty and blade design coefficients

The range of duty coefficients shown in Figure 3.11 in combination with six sets of geometrical variables tabulated in Table 3.6 leads to 65 x 6 = 390 unique blade designs. Each unique design is represented by a single point and each design variation as per Table 3.6 is represented by a unique colour in Figure 3.15.

Set	tc	sf
Design 1	0.26	1.80
Design 2	0.28	1.90
Design 3	0.22	1.75
Design 4	0.24	1.85
Design 5	0.20	1.95
Design 6	0.30	2.00

Table 3.6: The values of tc and sf considered to setup the final test cases.

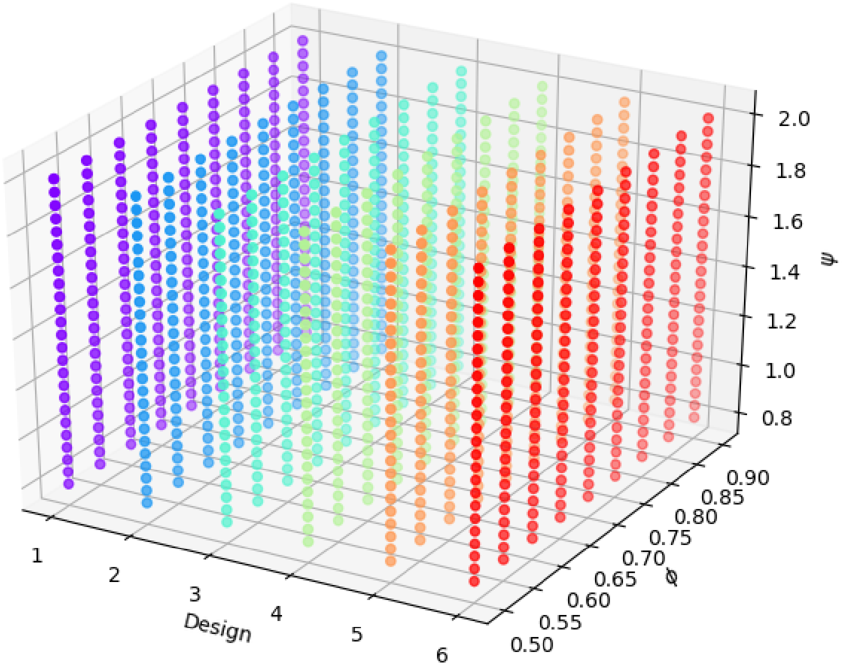


Figure 3.15: The range and design points of the duty coefficients and geometrical parameters considered to setup the final test cases. Each colour or 'Design' corresponds to a set of (tc and sf).

3.3.4. Selection of working fluid, volumetric flow ratio and Z

Recall the objective of the present work is to study the influence of the working fluid, inlet compressibility factor Z and total-to-static volumetric flow ratio on the dissipation coefficient $C_{d,blade}$. The range of selection of the latter three used for the present analysis is the topic of this section.

Working fluid

Different working fluids affect the flow field differently because of their thermodynamic properties. For instance, an organic fluid has a lower speed of sound than air so if the working fluid of a machine is changed from air to an organic fluid, there are bound to be more shocks in the flow field. Similarly, the boundary layer over a fixed geometry is different for two fluids because viscosity, thermal conductivity and diffusivity. In the light of these considerations, 4 working fluids, namely N_2, CO_2, Toluene and MM, are selected for the present study. These fluids are selected on the basis of their γ value. The latter two, Toluene and MM, are organic fluids, whereas the prior two are fluid similar to and lighter than air in terms of γ. Other properties of these fluids are tabulated in Table 3.7.

Table 3.7: Thermodynamic properties of the 4 selected fluids. The fluids are selected on the basis of the γ value

Fluid	γ [-]	m_{molar} [g/mol]	$\mathbf{R}_{gas}$ [J/kg/K]
N_2	1.40	28.01	296.80
CO_2	1.30	44.01	188.92
Toluene	1.05	92.14	90.24
MM	1.03	162.38	51.20

Three fluids, N_2, Toluene and MM are used to study the influence of NICFD on the flow field when γ_{pv} » γ and γ_{pv} = γ. Giuffre shows that the prior condition (γ_{pv} » γ) has a significant increase in the local flow Mach number. The results of this study are discussed in Section 4.6.

Volumetric flow ratio

As per isentropic expansion relations the volumetric flow is related to the pressure ratio via γ. Larger values of α correspond to larger pressure ratios and therefore more prominent compressibility effects. Three α values, 2, 3 and 4, are selected for the analysis. Each represents low, medium and extreme cases respectively. Usually, the shape of turbine expanders blades operating with $\alpha = 4$ is convergent - divergent. However as the present research is first of its kind only axial blades are considered.

Compressibility factor and inlet conditions

For this analysis the 2 Z conditions are considered, 1.0 and 0.7. Each represents an ideal and non-ideal case. According to the theorem of corresponding states, when different fluids are compared at the same reduced temperature and pressure the Z deviates for all fluids to approximately the same degree. One approach to select the inlet conditions is to select the same reduced temperature and pressure for all fluids. However, this approach is inapplicable practically due to the following reasons:

- The expansion ratio for different fluids is different for the same α because of γ_{pv}. For instance, N_2 will have a higher pressure ratio than MM. Given inlet conditions of Z = 0.7 lie in regions close to the saturation curve and critical point, an expansion of the two fluids at a fixed α might lead to phase change for one and not for the other.
- The extent of the thermodynamic library and formulation of equation of states used by SU2.

Consequently, a variety of inlet temperature and pressure is used for different fluids, while ensure the value Z is either 0.7 or 1. These varying inlet conditions are determined using a trial and error approach in SU2 for each fluid. Table 3.8 shows the finalised combinations of the $fluid, \alpha$ and Z.

3.3.5. Performance of the Data Generation Block

The Data Generation block shown in Figure 3.1 is used to generate multiple instances of data. Each instance correlates the inputs, namely ϕ, ψ, tc, sf, working fluid, Z and α, to the blade dissipation coefficient $C_{d,blade}$. 9360 unique test cases presented in Table 3.1 are simulated using the steps shown in the Data Generation block. 7485 out of these 9360 simulations run to completion successfully and produce $C_{d,blade}$ values. The remaining 1875 do not run successfully because the residuals of the CFD simulation diverge. This divergence is traced back to the quality of the mesh for several blade geometries; all of which turn the flow the most.

Table 3.8: 24 unique combinations of 4 fluids, 3 α and 2 Z values. The reduced inlet temperature and pressure dictate the value of Z.

Z = 1.0						**Z = 0.7**				
Fluid	α	**P**$_{r,in}$	**T**$_{r,in}$	**P**$_{in}$	**T**$_{in}$	α	**P**$_{r,in}$	**T**$_{r,in}$	**P**$_{in}$	**T**$_{in}$
N_2	2	0.15	3.00	509370	378.52	2	0.95	1.18	3226010	148.91
N_2	3	0.15	3.00	509370	378.52	3	0.95	1.30	3226010	164.05
N_2	4	0.15	3.00	509370	378.52	4	0.98	1.40	3327884	176.67
CO_2	2	0.50	2.50	3688650	760.37	2	0.95	1.18	7008435	358.87
CO_2	3	0.50	2.50	3688650	760.37	3	0.95	1.18	7008435	358.87
CO_2	4	0.50	2.50	3688650	760.37	4	0.98	1.40	7229754	425.78
MM	2	0.03	0.92	58170	477.07	2	0.95	1.08	1842050	560.25
MM	3	0.03	0.92	58170	477.07	3	0.95	1.08	1842050	560.25
MM	4	0.03	0.92	58170	477.07	4	0.95	1.10	1842050	570.60
Toluene	2	0.05	1.20	206300	709.43	2	0.90	1.08	3713400	639.09
Toluene	3	0.05	1.20	206300	709.43	3	0.90	1.08	3713400	639.09
Toluene	4	0.05	1.20	206300	709.43	4	0.90	1.08	3713400	639.09

Blade profiles within combinations of duty coefficients $\phi \leq 0.6$ and $\psi \geq 1.5$ and with $\phi = 0.5$ $\psi = 1.4$ do not have low unstructured meshes for the inputs reported in Table 3.3. Consequently, these 13 blade profiles for each combination of input parameters mentioned in Table 3.1 are neglected from the analysis.

Table 3.9 summarises the range and combination of parameters that are successfully executed using the methodology of the Data Generation Block. These combinations are used to study the influence of each parameter on C$_{d,blade}$ and to develop a machine learning model.

Table 3.9: Combination of parameter and their ranges successfully executed through Data Generation block to obtain populate a database for machine learning.

Parameter	**Range**
ϕ	0.5 - 0.9 (Δ 0.1) for $\psi \leq 1.3$, 0.6 - 0.9 (Δ 0.1) for $\psi = 1.4$, 0.7 - 0.9 (Δ 0.1) for $\psi \geq 1.5$
ψ	0.8 - 1.3 (Δ 0.1) for $\phi = 0.5$, 0.8 - 1.4 (Δ 0.1) for $\phi = 0.6$, 0.8 - 2.0 (Δ 0.1) for $\phi \geq 0.7$
tc, sf	(0.26,1.80),(0.28,1.90),(0.22,1.75), (0.24,1.85),(0.20,1.95),(0.30,2.00)
$fluid$	N_2, CO_2, Toluene, MM
Z	0.7, 1.0
α	2, 3, 4

3.4. Machine learning procedure

The objective of this section is to elucidate the steps shown in the Machine Learning block in Figure 3.1 to train, develop and validate a machine learnt model capable of predicting C$_{d,blade}$. The machine learnt or data-driven model is built on a database generated by executing each combination of parameters presented in Table 3.9 through the Data Generation Block introduced in Section 3.2.

3.4.1. Description of the database

Table 3.10 shows the total number of unique data measurements, maximum, minimum and standard deviation of the numerical features, namely $\phi, \psi, tc, sf, Z, \alpha$, and the label C$_{d,blade}$. The explanation for 7485 unique points in this database is provided in Section 3.3.5. The mean value of the label is comparable to that of Denton's 0.002 for turbomachines. However, it is important to note that this the mean from all the simulations and is not indicative of the behaviour of all fluids. The small standard deviation indicates that the spread of C$_{d,blade}$ values are close to the mean.

Table 3.10: The count, mean, standard deviation, minimum and maximum values of the features and label in the database generated.

	ϕ	ψ	tc	sf	Z	α	**C$_{d,blade}$**
Count	7485.0	7485.0	7485.0	7485.0	7485.0	7485.0	7485.0
Mean	0.738	1.319	0.250	1.875	0.850	3.000	2.011E-3
Standard	0.130	0.365	0.034	0.085	0.150	0.817	0.396E-3
Min	0.500	0.800	0.200	1.750	0.700	2.000	0.566E-3
Max	0.900	2.000	0.300	2.000	1.000	4.000	3.182E-3

3.4.2. Categorisation and Standardisation

Categorisation

Machine learning models mostly work with features that are numeric, however, in some cases the feature is categorical and not quantified. For the present study, ϕ, ψ, tc, sf, Z and α are numeric and the working fluid is categorical data. For the machine learning algorithm the working fluid needs to be translated into a numerical form. To do this the OneHot Encoder technique is used, which transforms the input using a one-hot (aka 'one-of-K' or 'dummy') encoding scheme [46]. The result of this is a sparse matrix with a binary column for each category. Table 3.11 shows an example of the output produced by the OneHot Encoder. For repeating data, such as MM in this table, the same binary position in the columns as the earlier instance of MM is used.

Table 3.11: An example of the scheme produced by One-hot encoder for categorical data, namely the working fluid.

Fluid	**OneHot encoding scheme**			
CO_2	1	0	0	0
MM	0	1	0	0
N_2	0	0	1	0
Toluene	0	0	0	1
MM	0	1	0	0

The numeric features are standardised using the feature of Sklearn. The numerical features are standardised to prevent biases and make the data have a normal distribution with the mean at 0 and unit variance. Standardisation eliminates the biases that may otherwise be present due to the magnitude of the absolute values of each parameter. If not standardised, the algorithm will prioritise and have a bias towards α, which has values ranging from 2 to 4 compared to tc which has values ranging from 0.2 and 0.3.

3.4.3. Splitting data into training and testing sets

Before using the data to start a training model analysis, the data is split into a training and testing set. 80% of the data is taken as the training and 20% as the testing. Rather than taking a random sample every time the code is executed, the random number generator's seed is fixed. This way it always produces the same shuffled indices and keeps the testing set hidden.

A random sample works best for cases when the data set is large. As there are 7485 entries in the database, a random sample is bound to pick data for training and testing sets that is biased. To prevent this stratified sampling is used. Stratified sampling is a method of sampling from a dataset which can be partitioned into sub groups. The database generated is grouped on the basis of working fluids. This way both the training and testing set contain samples of all four fluids equally. Table 3.12 shows the percentage distribution of the fluid data in the entire, training and testing set.

Table 3.12: Representation of the subgroups of data in the entire, training and testing datasets. The subgroups are on the basis of the working fluid.

Dataset	N_2	CO_2	Toluene	MM	Size
Entire	25	25	25	25	7485
Training	25	25	25	25	5988
Testing	25	25	25	25	1497

3.4.4. Training models

As the objective is to develop a model based on data that has features and labels, supervised machine learning is the most suitable choice. In Section 2.6, different training models are discussed, however before selecting them a performance measure is selected. The primary performance measure used for the present work is the R-square value, and the secondary is the Root Mean Square Error (RMSE).

Another method that is used for assessing the performance of models and for tuning of parameters for optimal performance is cross-validation. The k-fold cross-validation method is used for the present work. The training data is split into k random subsets, out of which the algorithm uses k - 1 to learn. The remaining data is used to test the model. This is repeated for all the possible k arrangements of the subsets. The performance measure i.e. R-square and RMSE is calculated for each iteration. The final estimate of the model performance is the average of all these iterations.

Models tried are Linear Regression, Polynomial Regression, Regularised models and Support Vector Regression. Table 3.13 summarises the average validation scores and training scores. The validation test scores are the average of 10 cross-validation sets. From this table it is observed that Lasso and SVR are the worst performing ones as the R-square parameter is nil. Lasso and Ridge has the same performance and Polynomial regression of order 2 gives the highest R-square and the lowest RMSE of all. Based on this it is decided that Polynomial regression is the best choice of model.

The optimal degree of order for the polynomial regression is found using hyperparameter tuning. For polynomial regression hyperparameter tuning is the degree of order. A range of values from 2-8 is tried. Table 3.14 shows the average validation test scores and training scores. The average validation test scores is based on 5 cross validation sets. The lowest R-square value for the validation test set is that of degree 4 and for the training test set it is degree 8. However, degree 8 also has the lowest validation test score. This means that the data for training is being overfitted. On the other hand the polynomial of degree 1 is underfitted as

Table 3.13: The performance scores for the validation and training data for different models. The validation score is calculated by taking the average of 10 k-cross validation subsets.

	Validation Test		**Training**	
	R2	**RMSE** $[10^{-4}]$	**R2**	**RMSE** $[10^{-4}]$
Linear	0.74	2.00	0.74	2.00
Ridge	0.74	2.00	0.74	2.00
Lasso	0.00	3.95	0.00	3.95
Polynomial (2)	0.94	1.00	0.94	0.99
SVR	0.00	3.96	0.00	3.96

both the validation and the training score is weak. In light of these considerations, polynomial model of order 4 is considered to be the best choice.

Table 3.14: Validation test and training scores of the polynomial regression algorithm.

Degree	**Validation test score**	**Training score**
1	0.743	0.744
2	0.935	0.936
3	0.949	0.953
4	0.956	0.963
5	0.955	0.969
6	0.950	0.974
7	0.940	0.979
8	0.900	0.983

The performance of the model is checked using Figure 3.16. This figure shows the model's performance on the training and validation set with respect to training set size. When the training set size consists of a few instances, the model is perfectly able to fit the data, hence the RMSE is nil. As more data is exposed the data becomes more noisy and the RMSE value increases before reaching a plateau. On the other hand, the performance of the validation set shows that the model is unable to generalise properly based on a small training set size. However, as this size is increased the validation error reaches a plateau and reaches the same value as that of the training set. Both the curves are converging to a single RMSE value, that means that both the models are neither over- nor underfitted.

3.4.5. Machine learnt model

The evaluation scores of the training model will evaluates how well the model is trained. The testing score evaluates the model's ability to predict new data. The scores of the 4^th degree polynomial regression model, which is the finalised machine learnt model, are presented in Chapter 5.

This machine learnt model is substituted in TurboSim, the physics-based loss model, to assess its influence on the profile loss and eventually the total-to-total stage efficiency. The results are presented in Chapter 5 as well.

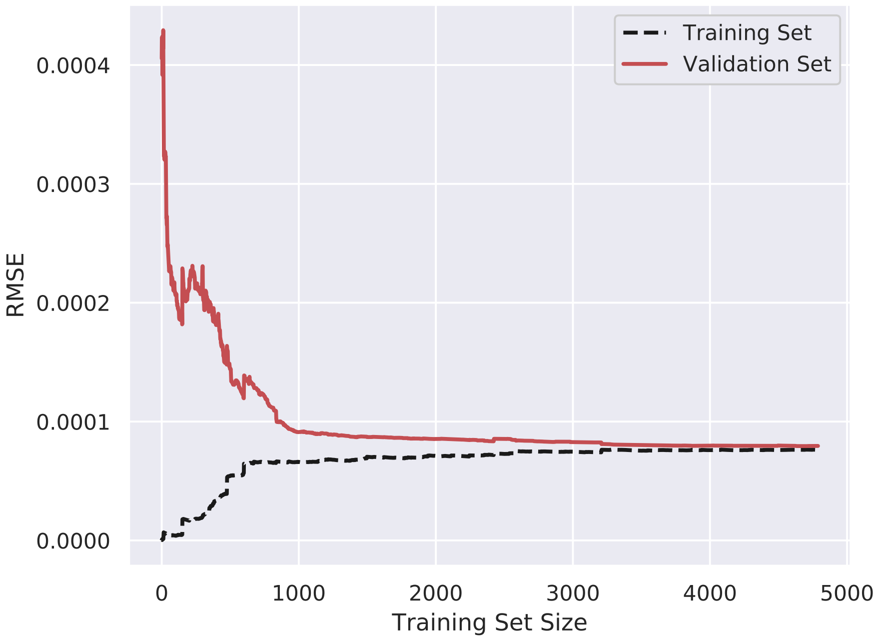


Figure 3.16: The learning curve of 4th order polynomial regression model. The performance of the model, in terms of RMSE, is checked on the training and the validation set as a function of the size of the set.

4

Results of parametric analysis

This chapter discusses the trends obtained from the 2D CFD RANS simulations of the different test cases introduced in Chapter 3. These trends are used to study the individual influence of the input parameters, namely ϕ, ψ, tc, sf, Z, α and working fluid, on the boundary layer dissipation coefficient $C_{d,blade}$. The effect of each input parameter on $C_{d,blade}$ is studied by holding the other input parameters constant. CO_2 is taken as the base reference fluid and is used for the analysis of all parameters except working fluid.

Section 4.1 - 4.4 presents the individual impact of the duty coefficients (ϕ, ψ), blade geometrical coefficients (tc, sf), compressibility (Z), volumetric flow ratio (α) and fluid on $C_{d,blade}$ respectively. The influence of γ_{pv} on the flow field is also studied; its results are shown in Section 4.6. Eventually, the influence of each parameter on $C_{d,blade}$ is quantified in Section 4.7.

4.1. Influence of duty coefficients

To illustrate the influence of only the duty coefficients on $C_{d,blade}$ all other parameters are kept constant. Figure 4.1 shows the trend of $C_{d,blade}$ with ϕ and ψ for the combination of parameters reported in Table 4.1.

Table 4.1: Values of parameters used for analysing the influence of duty coefficients on $C_{d,blade}$.

Parameter	$fluid$	ϕ	ψ	tc	sf	Z	α
Value	CO_2	-	-	0.3	2.0	1	4

The blank space on the upper left corner of Figure 4.1 corresponds to the blade geometries that are not simulated due to errors in meshing. This blank space does not affect the quality of the results and conclusions significantly because it corresponds to regions of high work and low flow coefficients. Combinations of high work and low flow coefficients produce blades with extreme flow turning angles. For example a blade capable of turning the flow by 116° is required for ϕ, ψ = 0.5, 2.0 respectively. Such blades are subjected to high blade loading and often cause the flow to separation on the suction surface. Consequently, these combinations of duty coefficients are often avoided in practice in industry and can be avoided for the present analysis as well.

Lighter shades of contours in Figure 4.1 signify higher $C_{d,blade}$ values and vice versa. This means that for the set of parameters tabulated in Table 4.1 $C_{d,blade}$ increases with increasing duty coefficients. Increasing ψ increases the blade loading and the flow turning angle for a fixed ϕ value. Such blade geometries are prone to thicker boundary layers and flow separation. Thus, $C_{d,blade}$ increases with increasing ψ. ϕ is a ratio of the meridional to tangential flow. For a fixed value of ψ, increasing ϕ does not influence the blade loading and decreases the flow turning angle. Blade geometries with high ϕ or lesser flow turning are more prone to

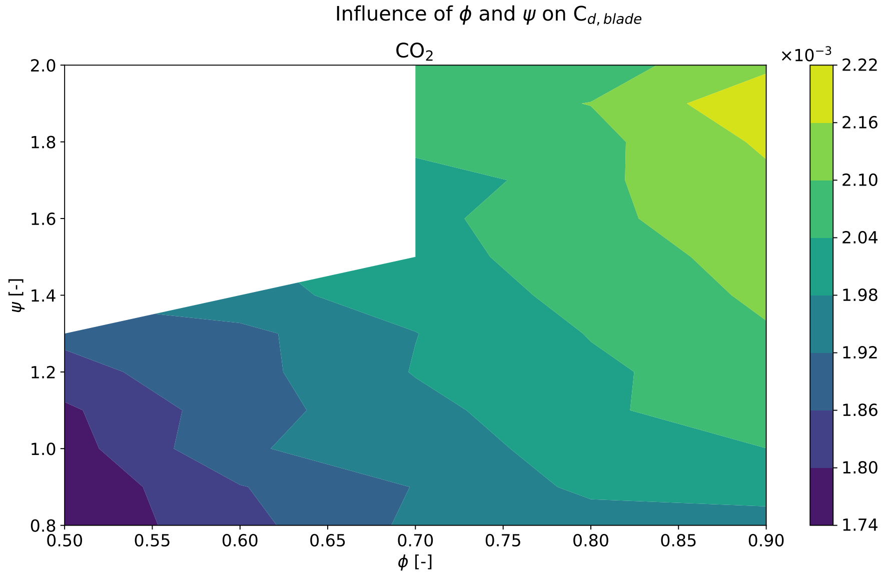


Figure 4.1: The influence duty coefficients (ϕ, ψ) on $C_{d,blade}$. The results are obtained by simulating blades with geometrical parameters of tc = 0.3, sf = 2.0 and operating conditions Z = 1.0, α = 4.

viscous losses than blade geometries with low ϕ. Blade geometries with low ϕ values have lower suction surface pressures. As a result the velocity over the suction surface increases thereby reducing the $C_{d,blade}$ according to Equation (2.9).

According to Figure 4.1 $C_{d,blade}$ is more sensitive to increase in ϕ than ψ. This is demonstrated by calculating the slope of $C_{d,blade}$ with respect to the same magnitude of step increase in ϕ and ψ individually. For ψ = 1.0, $\frac{\Delta C_{d,blade}}{\Delta\phi} = \frac{1.98-1.74}{(0.9-0.5)} \times 10^{-3} = 6 \times 10^{-4}$. For ϕ = 0.8, $\frac{\Delta C_{d,blade}}{\Delta\psi} = \frac{1.98-1.92}{(1.2-0.8)} \times 10^{-3} = 15 \times 10^{-5}$. It is important to note that these gradients correspond to combination of parameter values presented in Table 4.1.

4.2. Influence of geometrical parameters

For the set of duty coefficients presented in Figure 3.15, six unique designs are considered. These designs are different from one another in terms of the combinations of tc and sf and are tabulated in Table 3.6. To illustrate the influence of each unique combination of tc and sf on $C_{d,blade}$ all parameters are kept constant. Figure 4.2 shows the trend of $C_{d,blade}$ with unique designs on ϕ-ψ diagrams for the combination of parameters reported in Table 4.2.

Table 4.2: Values of parameters used for analysing the influence of geometrical parameters tc and sf on $C_{d,blade}$.

Parameter	$fluid$	ϕ	ψ	tc	sf	Z	α
Value	CO_2	-	-	-	-	1	4

Each of the 6 subplots shown in Figure 4.2 corresponds to a unique design. The combination of tc and sf the subplot represents is mentioned in its title.

The blank space on the upper left corner of each subplot in Figure 4.2 corresponds to the blade geometries that are not simulated due to errors in meshing. The detailed explanation for why this does not interfere in drawing conclusions is mentioned in Section 4.1.

A sudden change of contour levels at combination of duty coefficients (ϕ, ψ) = (0.6, 1.0), (0.7, 1.3), (0.8, 1.7) and (0.9, 2.0) in Design 1 is observed. The numerical implementation

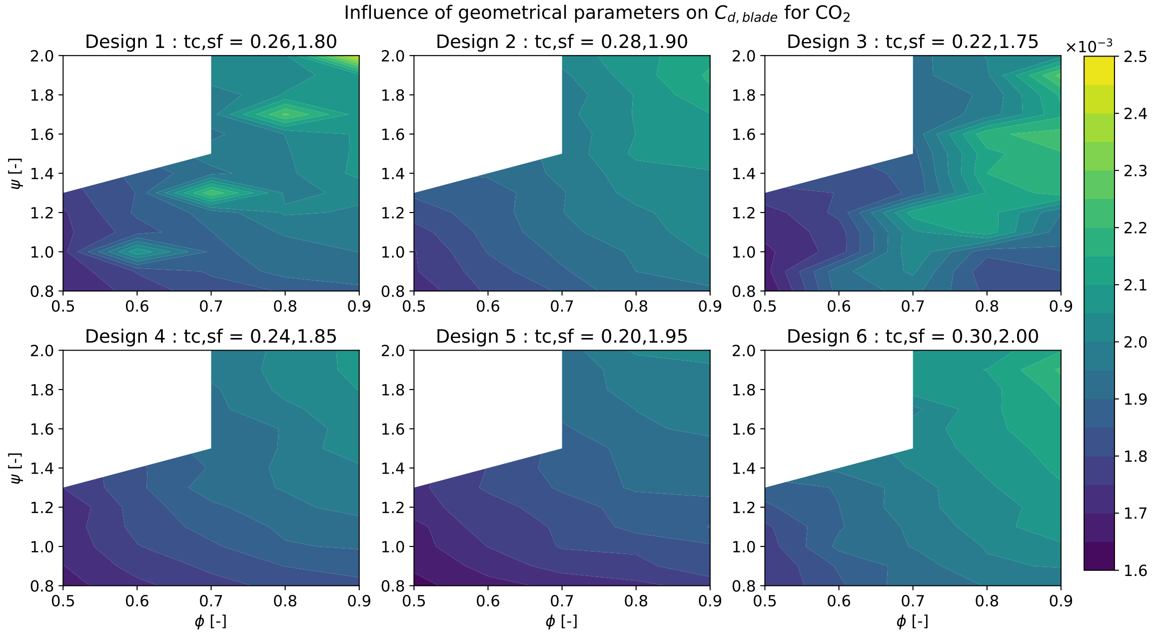


Figure 4.2: The influence geometrical parameters (tc, sf) on Cd. The results are obtained by simulating blades with operating conditions Z = 1.0, α = 4.

of Equation (2.8) and Equation (2.9) leads to fluctuations in the distribution of C$_{d,ss}$ obtained from the 17 normals along the blade suction surface shown in Figure 3.6. The reason for this sudden change in contour levels for the 4 combinations of duty coefficients mentioned above is attributed to the incorrect estimation of the dissipation coefficient on the suction surface C$_{d,ss}$ due to these fluctuations. The discrepancies in the subplot representing Design 3 in Table 4.2 are because of the same reason.

The influence of duty coefficients on C$_{d,blade}$ across all 6 designs is the same i.e. C$_{d,blade}$ increases with increasing duty coefficients. As the blade geometrical parameters tc and sf are selected in combinations their individual influence on C$_{d,blade}$ can not be observed explicitly.

From the discussion of Design of Experiments in Section 3.3.2 tc is strongly correlated to the C$_{d,blade}$ than sf. This is further substantiated by the subplot contours of Design 5 and 6 shown in Figure 4.2. Design 5 and 6 are used to demonstrate the influence of tc on C$_{d,blade}$ because both these designs represent the extremes of the range of blade thicknesses (tc) and have similar thickness distribution (sf). Design 5 consists of blade profiles with the lowest thickness to chord ratio and has tc = 0.20 and sf = 1.95. Design 6 consists of blade profiles with the highest thickness to chord ratio and has tc = 0.30 and sf = 2.0. Blades with geometry parameters of Design 6 have longer suction suction surfaces than that of Design 5. As a result the flow has to traverse over a longer distance on these blades. Increase in the blade surface length (l) directly increases the Reynolds number according to Equation (4.1).

$$Re = \frac{Vl}{\nu} \tag{4.1}$$

Depending on the value of Reynolds number, increasing Reynolds number may cause the flow over the blade surface to transition from laminar to turbulent. This would lead to an increase in entropy production due to the boundary layer and therefore increase C$_{d,blade}$. This is why the entire $\phi - \psi$ diagram for Design 5 has lower C$_{d,blade}$ values than that of Design 6.

To further strengthen this argument the entropy along the surface of two blade profiles is analysed. Both these profiles have the same duty coefficients (ϕ = 0.8 and ψ = 0.9) and

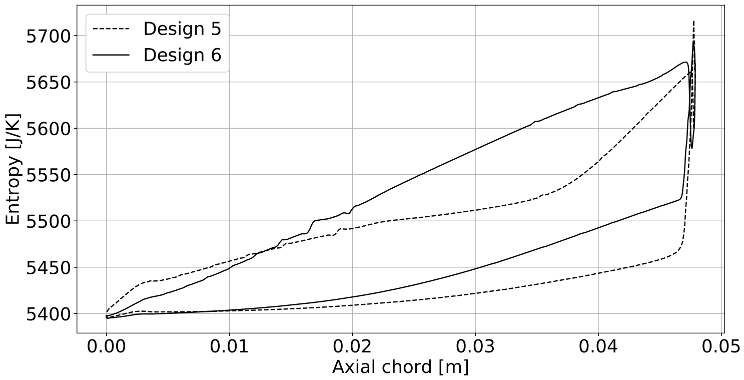


Figure 4.3: Entropy measured along the blade wall of blades with ϕ, ψ = 0.8,0.9 and geometrical parameters of Design 5 and 6. The blades are simulated with CO_2 and operating conditions of Z = 1 and α = 4.

simulation conditions reported in Table 4.2. However, they differ from one another in terms of the geometrical parameters. The first blade has tc and sf combination of Design 5 and the second has that of Design 6. The results are shown in Figure 4.3. The blade with features of Design 6 has higher entropy generation than that of the thinner blade with features of Design 5 after the axial chord distance of 0.0128m, which is also the location of maximum thickness for both blades.

From the discussion in Section 3.3.2 sf is weakly correlated to $C_{d,blade}$. Rather than comparing two contour plots shown in Figure 4.2, the entropy along the blade surface of two blade profiles with similar duty coefficients, ϕ = 0.7 and ψ = 1.4, is considered to study the influence of sf on $C_{d,blade}$. Both of these blade profiles are simulated with parameters mentioned in Table 4.2, however, differ from each other in terms of geometrical parameters. One blade profile has geometrical parameters of Design 3 and the other of Design 5. Both these Designs are similar in tc and represent the extremes of the range considered for the thickness distribution sf. Design 3 has blade profiles sf = 1.75 and tc = 0.22, whereas Design 5 consists of blade profiles with sf = 1.95 and tc = 0.20.

Figure 4.4 shows the entropy distribution along the surface of the two blades with different geometrical parameters. The entropy distribution of both blades is almost identical to each other. The entropy of the blade with geometrical parameters of Design 3 is higher than that of Design 5 between axial chord positions 0.010 and and 0.025. However, this increase is insignificant to draw conclusions from.

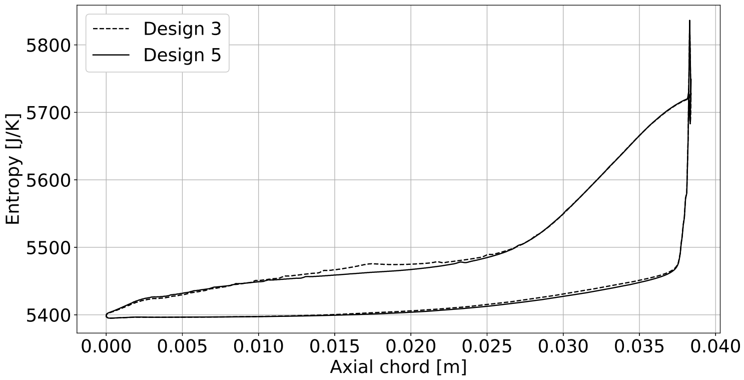


Figure 4.4: Entropy measured along the blade wall of blades with ϕ, ψ = 0.7,1.4 and geometrical parameters of Design 3 and 5. The blades are simulated with CO_2 and operating conditions of Z = 1 and α = 4.

4.3. Influence of thermodynamic inlet conditions

The compressibility factor Z is used to define the thermodynamic inlet conditions of the different test cases introduced in Chapter 3. For this analysis two input Z conditions, namely 1.0 and 0.7, are considered. Z = 1.0 represents ideal conditions and the other represents non ideal conditions. The reduced temperature and pressure (T_r and P_r) together dictate the value of Z. All combinations of inlet conditions considered for the present study are reported in Table 3.8.

To instantiate the influence of Z on $C_{d,blade}$ a set of stator blades corresponding to the range of ϕ and ψ reported in Table 3.9 are simulated with inlet conditions of Z = 1.0 and 0.7. Z = 1.0 is obtained by using inlet T_r = 2.5 and P_r = 0.5, whereas, Z = 0.7 is obtained by using inlet T_r = 1.4 and P_r = 0.98. All other parameters are kept constant; their values are listed in Table 4.3. Figure 4.5 shows the trend of $C_{d,blade}$ with Z on a $\phi - \psi$ diagram.

Table 4.3: Values of parameters used for analysing the influence of Z on $C_{d,blade}$.

Parameter	$fluid$	ϕ	ψ	tc	sf	Z	α
Value	CO_2	-	-	0.30	2.00	-	2

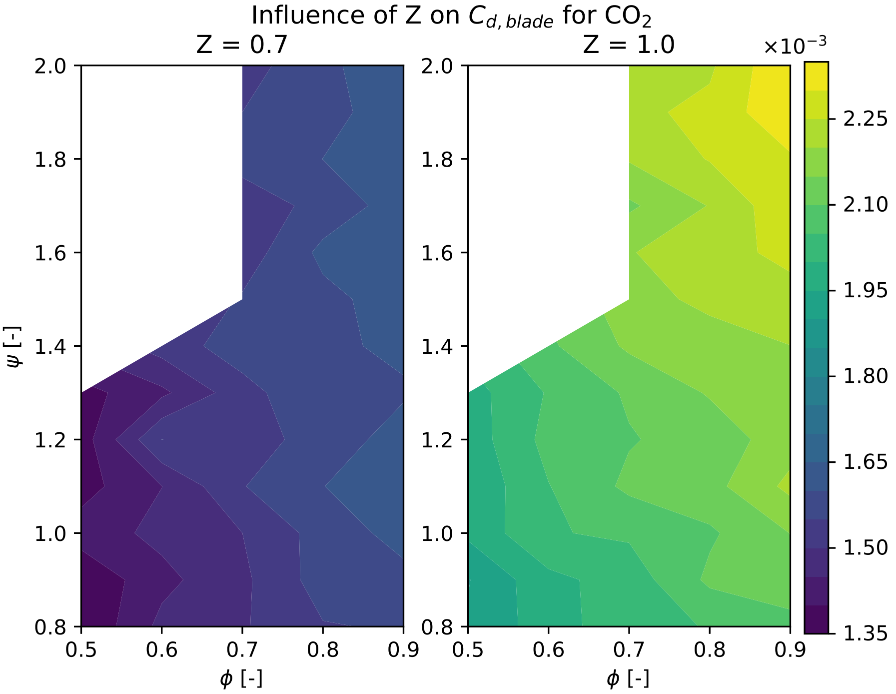


Figure 4.5: The influence of Z on Cd. The results are obtained by simulating blades with geometrical parameters of tc = 0.3, sf = 2.0 and fluid = CO_2, α = 2.

Just like the previous sections in this chapter, the blank space in the upper left corner of both subplots shown in Figure 4.5 corresponds to the blade geometries that were not simulated due to errors in meshing and does not influence the conclusions.

The contour subplot of Z = 0.7 in Figure 4.5 has lower $C_{d,blade}$ values than the subplot of Z = 1.0. Decreasing Z decreases $C_{d,blade}$. The inlet conditions for Z = 0.7 are close to the critical point. Isentropic expansions close to the critical point intrinsically have lower entropy than expansions away from the critical point. This behaviour is testified using Figure 4.6. This T-s diagram shows the isentropic expansion of CO_2 from the inlet conditions, namely Z = 1.0

and Z = 0.7, for α = 2.0. According to Equation (2.8) lower entropy leads to lower entropy generation from the boundary layer and therefore lower $C_{d,blade}$ values.

Figure 4.6 also shows that the expansion process of fluids with inlet conditions close to the critical point experiences a stronger gradient of Z and γ_{pv} compared to inlet conditions away from the critical point. Change in γ_{pv} along the expansion of Z = 0.7 implies different pressure ratios for the same α according to isentropic relations shown in Equation (2.2). These local changes along the expansion highlights the importance of determining the local thermodynamic conditions, such as Mach, when away from Z = 1. Giuffre and Pini showed that the local Mach flow number changes significantly when $\gamma_{pv} > \gamma$ [14]. The influence of γ_{pv} on local Mach is explored in Section 4.6.

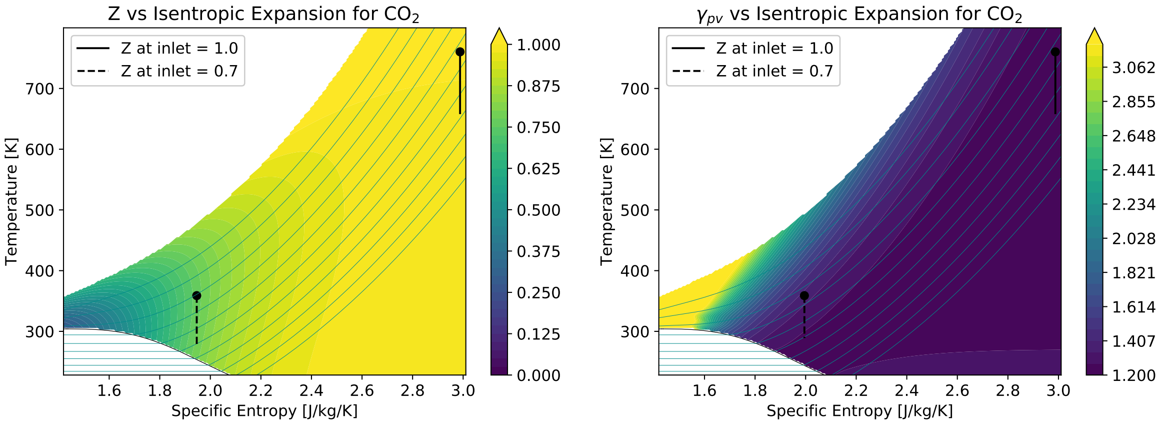


Figure 4.6: Isentropic expansion of CO_2 at inlet conditions Z = 1.0 and Z = 0.7 on contour plots of Z and γ_{pv}. Z = 1.0 corresponds to T_r = 2.5 and P_r = 0.5. Z = 0.7 corresponds to T_r = 1.4 and P_r = 0.98.

4.4. Influence of volumetric flow ratio

To study the influence of only the total-to-static volumetric flow ratio α on $C_{d,blade}$ a set of stator blades based on the range of duty coefficients listed in $Table$ 3.9 are simulated with 3 values of α, namely 2, 3 and 4. All other simulation parameters are kept constant; their values are reported in Table 4.4. Figure 4.7 shows the trend of $C_{d,blade}$ with α on three subplot $\phi - \psi$ diagrams; each corresponding to a value of α.

Table 4.4: Values of parameters used for analysing the influence of Z on $C_{d,blade}$.

Parameter	$fluid$	ϕ	ψ	tc	sf	Z	α
Value	CO_2	-	-	0.30	2.00	1.0	-

The blank corners on the upper left side of the subplots in Figure 4.7 represents the region of blade geometries that were not simulated due to errors in meshing. The absence of this region does not influence the conclusions drawn from this figure because of reasons stated in Section 4.1. Based on the three contour plots shown in Section 4.4 it is visible that the $C_{d,blade}$ decreases with increase in α. The subplot representing α = 2 has the lightest shades of the contour levels compared to the subplot representing α = 4.

α determines the outlet static boundary conditions. α is related to the total-to-static pressure ratio β via the isentropic relation shown in Equation (4.2). In this equation k assumes the value of γ_{pv}.

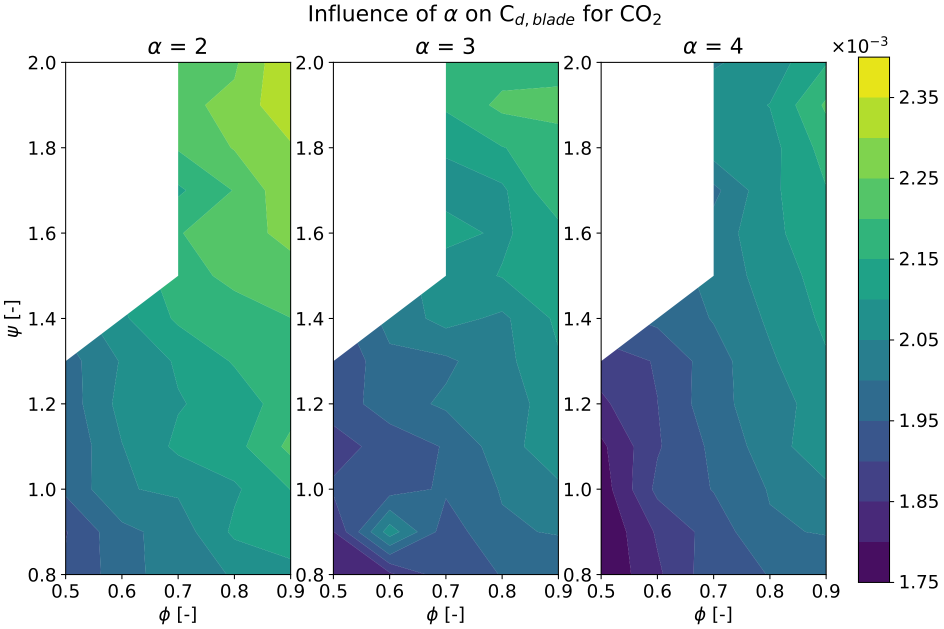


Figure 4.7: The influence of α on Cd. The results are obtained by simulating blades with geometrical parameters of Design 6 and operating conditions fluid = CO_2, Z = 1.0.

$$\beta = \alpha^k \tag{4.2}$$

According to Equation (4.2), an increase in α leads to an increase in β for the same k. An increase in β implies the static pressure at the outlet of the stator decreases, thereby increasing the dynamic pressure and the local flow velocity. The compressibility effects on the fluid become prominent as the flow of the fluid increases. When the local flow moves faster than the speed of sound i.e. when the Mach values along the blade passage increase and the flow becomes more susceptible to shock losses. The focus of the present study is on losses from boundary layers, so the interference of shocks or separation is neglected from the calculation of $C_{d,blade}$ as discussed in Section 3.2.4. As a result, the losses from only boundary layer is higher for α = 2 than α = 4 in Figure 4.7.

Another reason for decrease in $C_{d,blade}$ for increase in α can be explained using Equation (2.9). The ratio of $v_{\delta_e}^3$ to S_a decreases as v_{δ_e} increases. Consequently, leading to lower $C_{d,blade}$ values.

4.5. Influence of the working fluid

Four working fluids, namely N_2, CO_2, Toluene and MM, are selected for the present analysis. These four fluids represent increasing molecular mass and decreasing γ. To illustrate the influence of only the fluid on the dimensionless blade boundary layer dissipation coefficient all other parameters are kept constant. Figure 4.8 presents the trend of $C_{d,blade}$ with ϕ, ψ for the four fluids and combinations of parameters reported in Table 4.5.

Table 4.5: Values of parameters used for analysing the influence of the fluid on $C_{d,blade}$.

Parameter	$fluid$	ϕ	ψ	tc	sf	Z	α
Value	-	-	-	0.30	2.00	1	4

The presence of a blank region on the top left corner of all four $\phi - \psi$ diagrams is explained in Section 4.1. Figure 4.8 shows that C$_{d,blade}$ increases with increase in molecular complexity. As all the results displayed in Figure 4.8 are simulated using ideal inlet conditions (Z = 1) k = γ for Equation (4.2). For the same α the β decreases for increasing molecular complexity because γ decreases with increasing molecular complexity as reported in Table 3.7. An increasing β implies an increase in local flow velocity. As the focus of the present study is to assess losses from boundary layers, so the interference of shocks or separation is neglected from the calculation of C$_{d,blade}$ as discussed in Section 3.2.4. As a result, the losses from only boundary layer is higher for lower values of γ i.e for MM than N_2.

The decrease in C$_{d,blade}$ for increase in α can be explained using Equation (2.9). The entropy generation decreases with respect to the increase in the bulk kinetic energy of the flow thereby leading to lower C$_{d,blade}$ values.

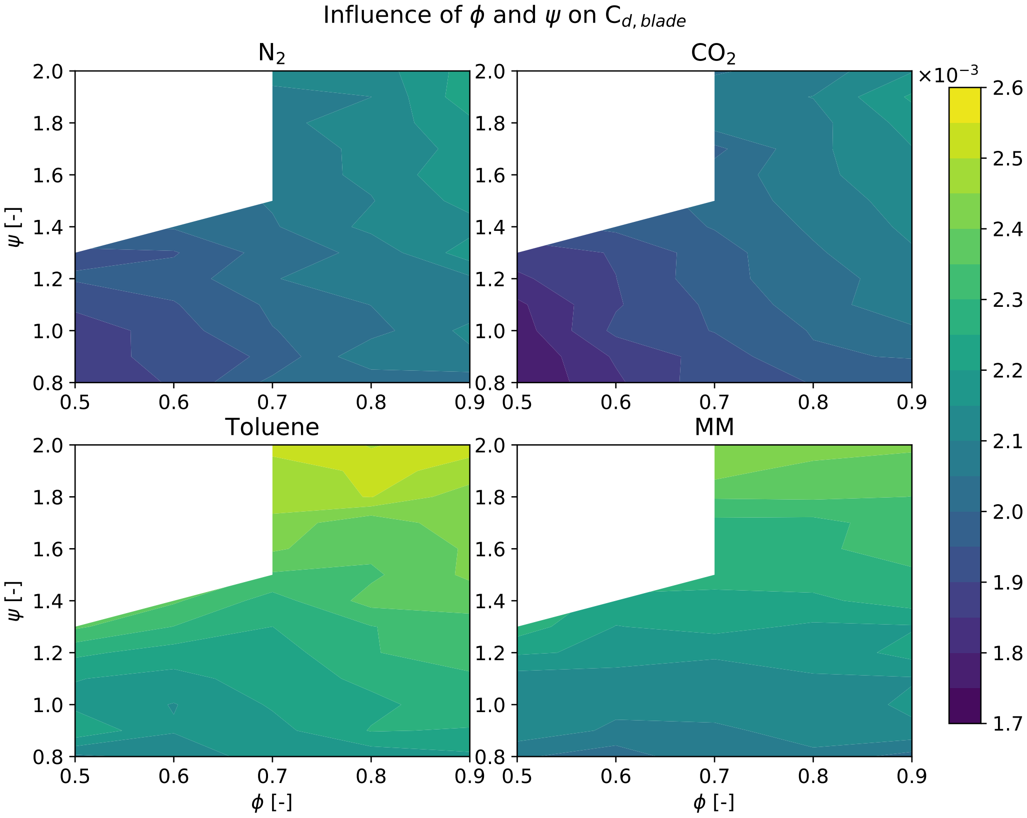


Figure 4.8: The influence fluid on Cd. The results are obtained by simulating blades with geometrical parameters of tc = 0.3, sf = 2.0 and operating conditions Z = 1.0, α = 4.

4.6. Influence of polytropic exponent on flow field

Section 4.5 and 4.3 analyse the influence of the fluid and inlet boundary conditions characterised by Z respectively. Z is taken as a measure of non-ideal effects of the fluid at the inlet and is determined using the inlet T_r and P_r. The polytropic exponent γ_{pv} is also a measure of non-ideal effects. The discussion from Section 2.2.1 shows that γ_{pv} differs from the isentropic expansion factor γ in regions close to the critical point, whereas regions away from the critical point has γ_{pv} values similar to γ. Giuffre and Pini compare stage performance for cases where γ_{pv} differs significantly from γ. They find that for the same working fluid the local Mach flow number changes significantly when $\gamma_{pv} > \gamma$ [14]. In light of these findings, the objective of this section is to study the influence of a fixed γ_{pv} on the flow field of different working fluids.

The study is performed on two working fluids, namely N_2 and MM for the cases reported in Table 4.6. T_r and P_r are chosen to ensure constant γ_{pv} = 1.42 at the inlet. Figure 4.9 shows the isentropic expansion of N_2 and MM on the contours of γ_{pv} and Z. This figure shows that the expansion of N_2 and MM for the same inlet γ_{pv} corresponds to Z = 0.98 and Z = 0.57 respectively. Additionally, expansion of MM experiences a variation of Z and γ_{pv} compared to N_2. With these considerations the blade with geometric parameters mentioned in Table 4.6 is simulated with the two fluids.

Table 4.6: Blade geometric parameters, inlet conditions used to study the influence of a fixed inlet polytropic exponent on the flow field of two working fluids.

$fluid$	ϕ	ψ	tc	sf	α	$T_{r,inlet}$	$P_{r,inlet}$	γ_{inlet}	$\gamma_{pv,inlet}$
N_2	0.70	1.30	0.20	1.95	2.00	1.90	0.19	1.42	1.42
MM	0.70	1.30	0.20	1.95	2.00	1.15	1.96	1.27	1.42

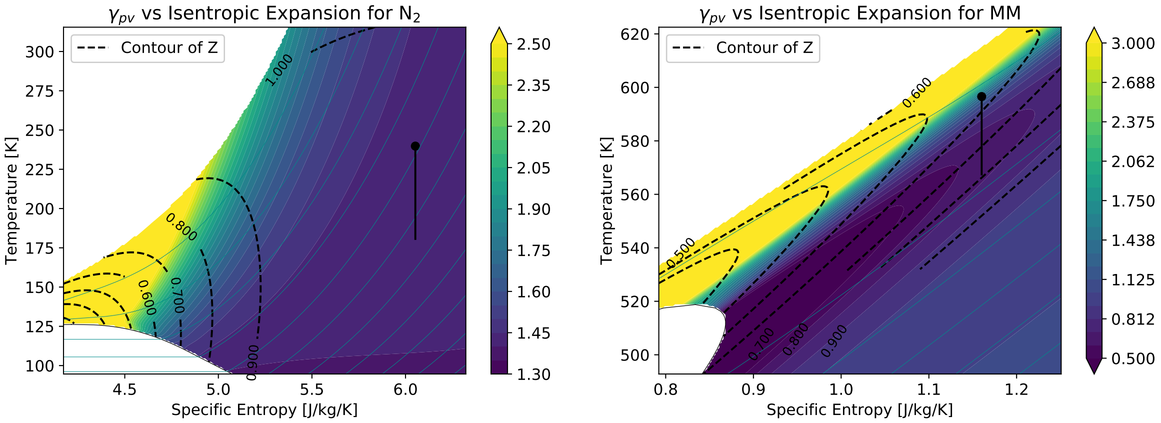


Figure 4.9: Isentropic expansion of N_2 and MM at constant inlet γ_{pv} = 1.42 on γ_{pv} and Z contours.

Figure 4.10 shows the flow field obtained from the simulation of the blades with the two fluids. The local Mach number in the blade passage of MM is visually higher than that of N_2. To quantify the values of Mach along the blade passage normals along the axial direction of the blade passage are drawn in Tecplot as shown in Figure 4.11. Values of γ_{pv} and Mach values are extracted using these normals and are averaged on the basis of the suction and pressure side. Figure 4.12 and 4.13 show the progression of average γ_{pv} and average Mach respectively.

The average γ_{pv} decreases in Figure 4.12 along the blade passage for the suction and pressure side of the blade for both N_2 and MM. γ_{pv} decreases more for MM than N_2 because the inlet conditions are close to the critical point in comparison to N_2. Therefore a larger variation of γ_{pv} is observed along the expansion of MM. Figure 4.9 corroborates these observations. However, γ_{pv} in Figure 4.12 is not as low as Figure 4.9 because the flow solver SU2 used to simulate the blade passages assumes that isentropic processes are polytropic. As a result the value of γ_{pv} is taken as a constant and is not updated along the expansion process.

Figure 4.13 shows the average Mach along the domain. The Mach for both MM and N_2 are similar till axial direction of 0.02m. Beyond this axial position the fluid accelerates more for MM than N_2. This means that local Mach increases for cases where $\gamma_{pv} > \gamma$ more than for cases where $\gamma_{pv} = \gamma$.

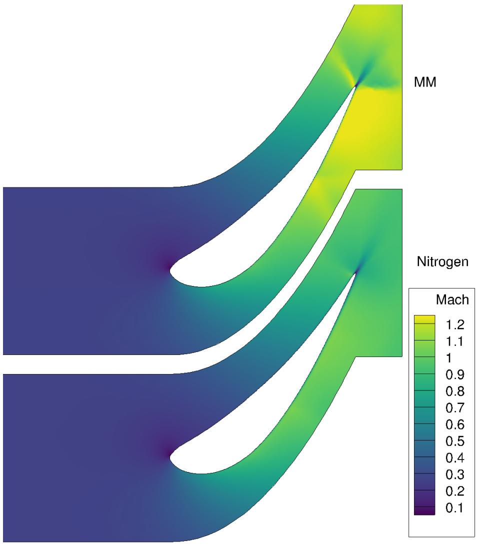


Figure 4.10: Flow field of blade with geometrical parameters ϕ = 0.7, ψ = 1.3, tc = 0.2, sf = 1.95 for working fluid N_2 and MM for inlet γ_{pv} = 1.42.

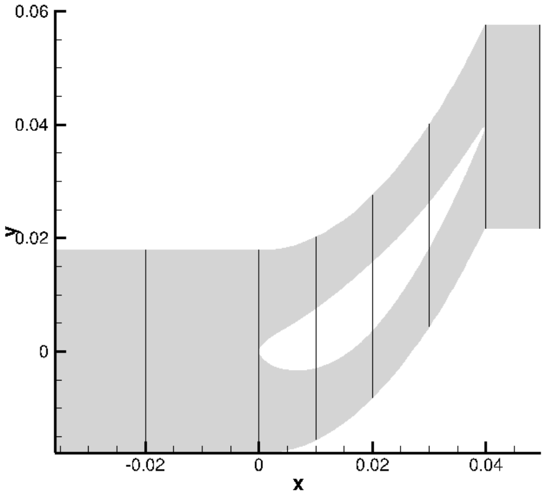


Figure 4.11: Normals along the domain are used to extract the Mach, entropy and γ_{pv}. Values passing over the blade are not considered.

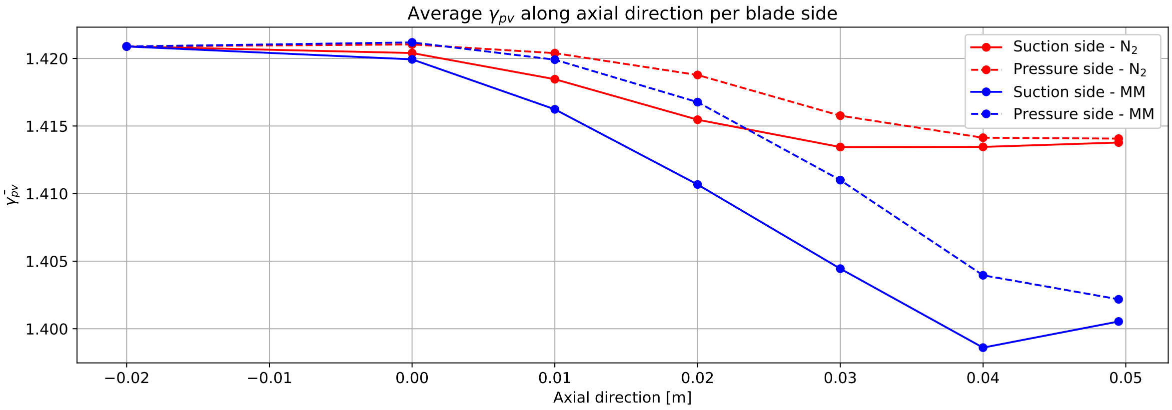


Figure 4.12: The progression of average γ_{pv} per surface fluid and per surface side along the axial direction of the blade passage.

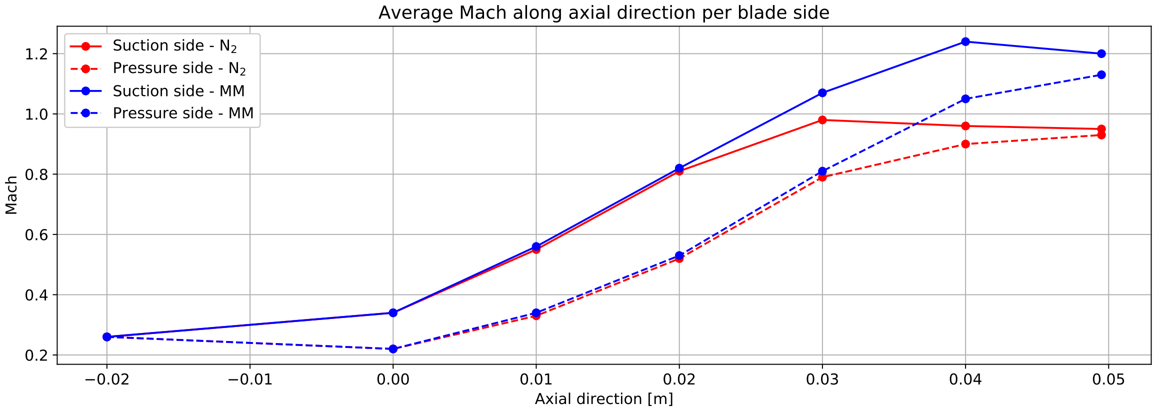


Figure 4.13: The progression of average Mach per surface fluid and per surface side along the axial direction of the blade passage.

4.7. Correlation coefficients

Section 4.1 - 4.4 present the individual influence of parameters, namely $\phi, \psi, tc, sf, \alpha, Z$ and the working fluid, on the dissipation coefficient C$_{d,blade}$. The analysis on the influence of one parameter is performed by holding all other parameters constant. Consequently, it is important to note that the results obtained for each analysis corresponds to a specific set of input parameters and therefore drawing generic conclusions based on these results would be misleading. For instance consider the trend of ϕ, ψ with C$_{d,blade}$ presented in Section 4.1. This trend shows that C$_{d,blade}$ is more sensitive to changes in ϕ than ψ. However, this conclusion holds only for the set of parameters reported in Table 4.1 and cannot be extrapolated to other working fluids, tc, sf, Z or α or generalised. To show the influence of only ϕ, ψ on C$_{d,blade}$ for all combinations of input parameters means generating 4 (fluids) x 6 (combinations of tc, sf) x 2 (Z) x 3 (α) = 144 $\phi - \psi$ diagrams and is not a plausible approach to illustrate the association of each parameter with C$_{d,blade}$.

To tackle this challenge the database containing C$_{d,blade}$ of the 7485 test cases introduced in Chapter 3 is used. Each test case is a combination of $\phi, \psi, tc, sf, Z, \alpha$ and working fluid and their range of values are reported in Table 3.9. Data of this entire database is used to present the relationship between parameters and between each parameter and C$_{d,blade}$ on a pair plot. A pair plot is a grid of scatter plots that shows the bivariate relationship between all pairs of variables in a multivariate database [57].

Figure 4.14 shows a pair plot for the database discussed in Section 3.4. The grid consists of two types of plots, namely density plot and scatter plot. The density plot along the diagonal shows the distribution of the parameter mentioned along the x-axis. The scatter plots under the diagonal of density plots depict the relationship between two parameters. For instance, the bottom-most left scatter plot shows ϕ with C$_{d,blade}$ for all data in the database. The last row of scatter plots is important because it helps visualise and quantify the strength the relation of each parameter with C$_{d,blade}$. The Pearson correlation coefficient introduced in Section 3.3.2 signifies the direction and strength of each relation.

A correlation coefficient matrix is used to tabulate the Pearson correlation coefficients between each variable. Figure 4.15 shows the correlation matrix corresponding to the pair plot shown in Figure 4.14. Each cell in this matrix corresponds to the correlation coefficient between two variables. The diagonal of this matrix is 1 because it shows the correlation of each parameter with itself.

The following observations are derived from both Figure 4.14 and Figure 4.15:

- The correlation coefficients of $tc - sf$, $\phi - \psi$ and $Z - \alpha$ are 0.37, 0.32 and 0.068. This parameter dependency is simply due to the selection of values for these parameters. For instance, the scatter plot in Figure 4.14 shows a positive linear relation between tc and sf, however, this is because the combination of values tabulated in Table 3.6. Selection of different combinations of $tc - sf$ would show a different correlation coefficient.

- C$_{d,blade}$ increases with increase in duty coefficients ϕ and ψ. This is visible from the positive gradient in the scatter plots of ϕ vs C$_{d,blade}$ and ψ vs C$_{d,blade}$ in Figure 4.14. Additionally, the dissipation coefficient is more sensitive to changes in ψ than ϕ. This trend is testified by the stronger correlation coefficient for the prior than the latter in Figure 4.15. The correlation coefficient of ψ is 0.22 and ϕ is 0.19.

- C$_{d,blade}$ increases with increase in tc and is almost insignificant to changes in sf. The correlation coefficient of the two are 0.12 and -0.008 respectively. As this conclusion is similar to the one of the Design of Experiments discussion in Section 3.2.1 and results presented in Section 4.2 the explanation provided in these sections can be used to explain this behaviour.

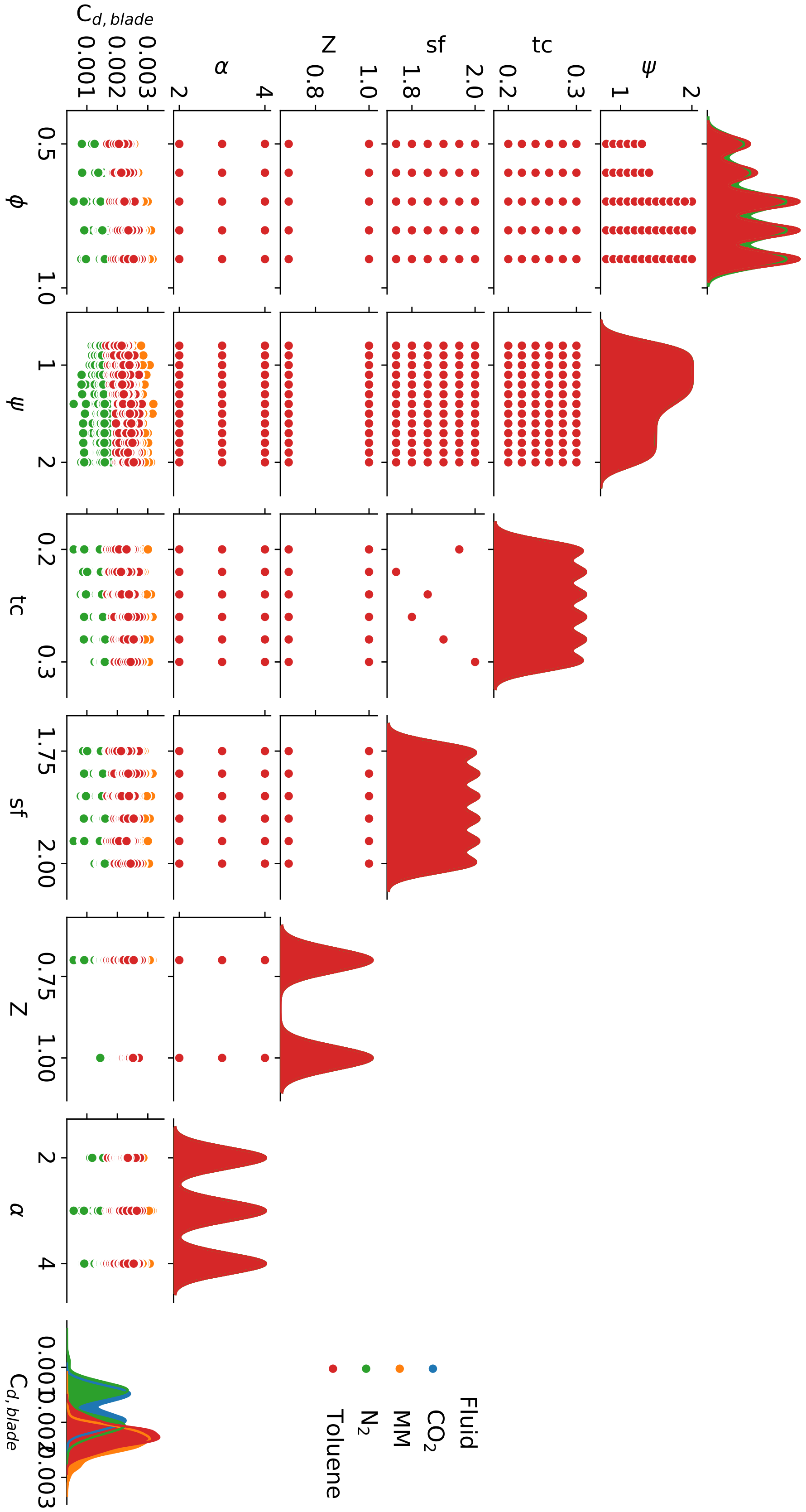
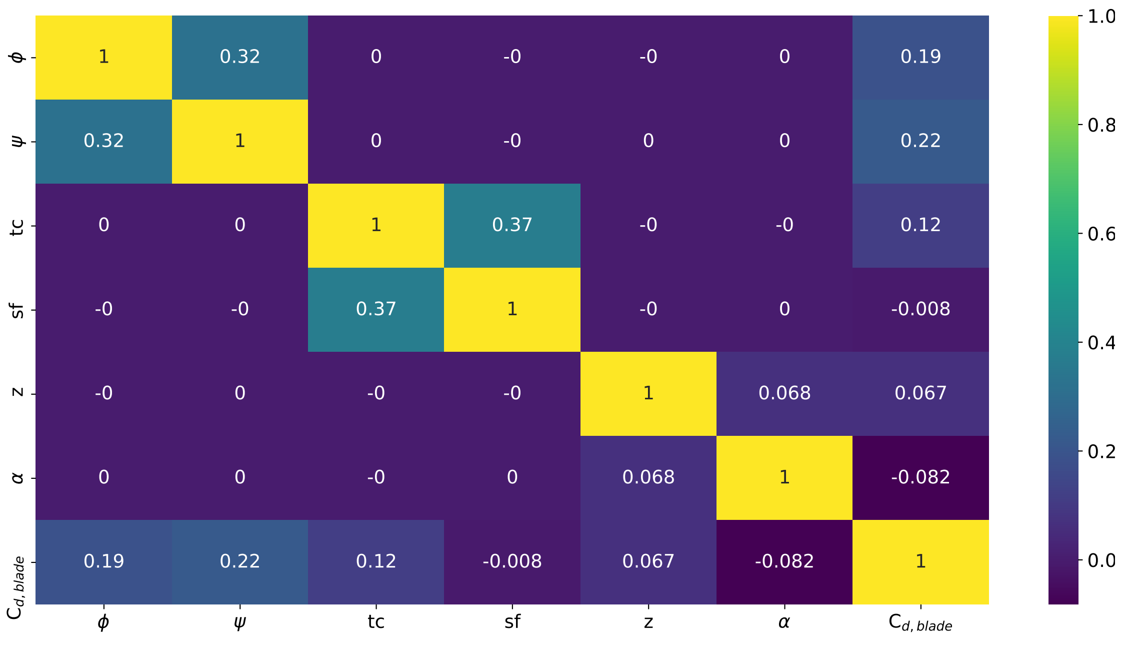


Figure 4.14: Pair plot for the database containing data on $C_{d,blade}$ for 7485 unique test cases. The data corresponding to each working fluid, namely N_2, CO_2, Toluene and MM, is highlighted using different colours.



	ϕ	ψ	tc	sf	z	α	$C_{d,blade}$
ϕ	1	0.32	0	-0	-0	0	0.19
ψ	0.32	1	0	-0	0	0	0.22
tc	0	0	1	0.37	-0	-0	0.12
sf	-0	-0	0.37	1	-0	0	-0.008
z	-0	0	-0	-0	1	0.068	0.067
α	0	0	-0	0	0.068	1	-0.082
$C_{d,blade}$	0.19	0.22	0.12	-0.008	0.067	-0.082	1

Figure 4.15: Correlation coefficients for the parameters and $C_{d,blade}$ and between parameters. The sign of the coefficients indicates positive or negative correlation and the magnitude determines the strength of the correlation.

- According to the results in Section 4.3, increase in Z leads to a significant increase in $C_{d,blade}$ for CO_2 and other parameters mentioned in Table 4.2. When all test cases are considered Z is weakly correlated to $C_{d,blade}$ by a correlation coefficient of 0.067 determined in Section 4.7.
- Results in Section 4.4 demonstrated that $C_{d,blade}$ increases with decreasing α. The negative correlation coefficient value of -0.082 shown in Figure 4.15 indicates that this is true for all cases considered in database.
- As the working fluid is not expressed as a number there is no means to measure the correlation coefficient. However conclusions can be drawn from the distribution of $C_{d,blade}$ per fluid shown in the distribution plot at the bottom-most right corner of Figure 4.14 and from Figure 5.2. The value of $C_{d,blade}$ is highest for MM followed by Toluene and CO_2 and lowest for N_2. The average $C_{d,blade}$ for all test cases simulated with N_2 is 0.0017, whereas that for MM is 0.0023.

The quantification of the influence of parameters on $C_{d,blade}$ show that the duty coefficients influence $C_{d,blade}$ the most. This is then followed by the maximum thickness to chord ratio, α and Z respectively. The geometrical parameter sf has negligible influence on $C_{d,blade}$. The correlation coefficient of the working fluid with $C_{d,blade}$ is not calculated, however, the effect is visible from the peak positions of the bell curve in Figure 4.14. The mean value of the dissipation coefficient for MM increases by 35% as compared to N_2.

5

Results of Machine learnt model

This chapter discusses the quality of the machine learnt model and the influence of the machine learnt model on the profile losses and total-to-total stage efficiency estimated by the physics loss model introduced in Section 2.3. In Section 5.1 the evaluation of the training model and validation scores of a 4^{th} degree polynomial regression model introduced in Section 3.4 is discussed. The following section discusses the implementation of the machine learnt model in the physics based model. The subsequent sections, Section 5.3 and 5.4, present the influence of the machine learnt model on the profile losses and stage efficiency estimated by the physics loss model, respectively.

5.1. Evaluation of the polynomial regression machine learnt model

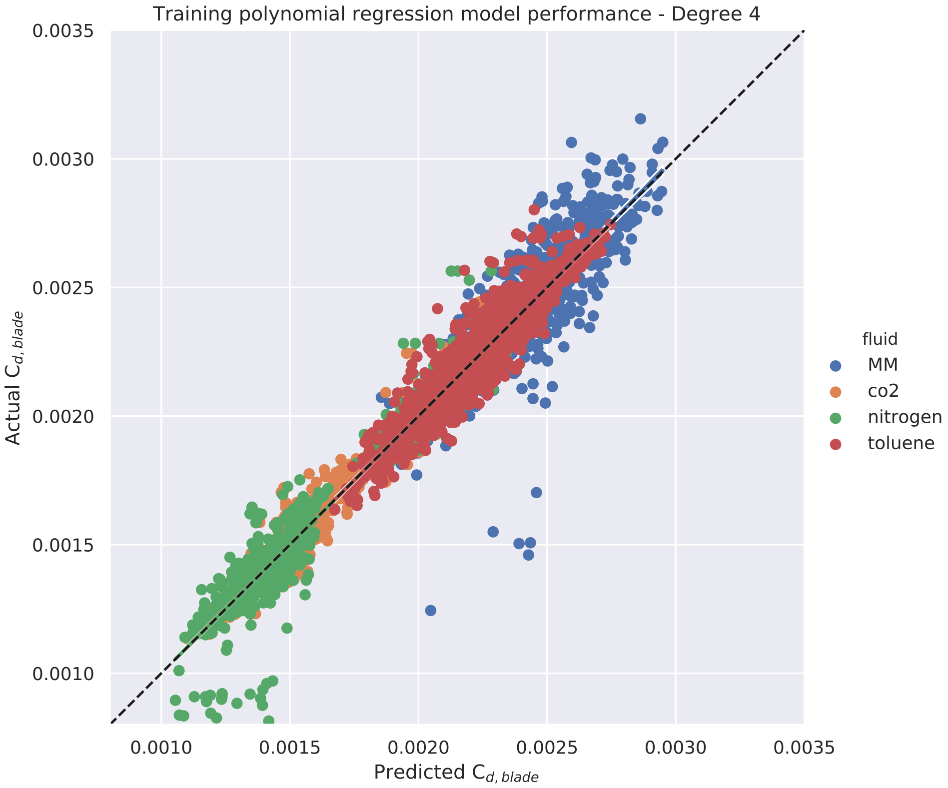


Figure 5.1: Comparison of $C_{d,blade}$ determined from CFD simulations considered actual data, and $C_{d,blade}$ predicted from a 4^{th} degree polynomial regression machine learnt model. The diagonal represents the line of perfect prediction.

The explanation for using a polynomial regression model for training the data and developing the model is explained in Section 3.4. The quality of training the prediction of C$_{d,blade}$ by the polynomial regression model compared to the actual C$_{d,blade}$, determined from the 2D RANS simulation, is visualised in Figure 5.1. The predicted values are represented on the x-axis and the actual on the y-axis. The diagonal dotted line is the line of perfect prediction. Points along this line signify a perfect match between the actual and predicted values.

Figure 5.1 shows that most of the data lies in close proximity to the line of perfect prediction. The outliers shown in Figure 5.1 for MM and N_2 are attributed to simulations with inlet condition Z = 0.7. For several blade profiles, the simulations for both these fluids leads to a lower order of convergence. The Root Mean Squared Error (RMSE), which is a measure of the dispersion of data around the line of perfect prediction, for the trained model is 7.63E-05. Overall the R-square, which measures the proximity of data to the fitted regression line, for the trained model is 0.963.

The trained model is then exposed to the testing data, which is data that the model has never studied before. The predictability of new data by the model ultimately defines its performance. Figure 5.2 compares the predicted C$_{d,blade}$ with the actual C$_{d,blade}$ values from the testing dataset. As the size of the testing dataset is smaller than the training fewer points are observed in Figure 5.2. Consequently, the number of outliers is lesser than that of the training set. Hence, the RMSE of this model for the testing data is 8.120E-05 and is lower than that from the training dataset. The R-square of the model with the testing dataset is 0.959 and is lower than the one observed from the training dataset because the model is exposed to lesser data and is unable to generalise as well as it did with the training set. In light of these results, it is concluded that the model is suitable for predicting C$_{d,blade}$.

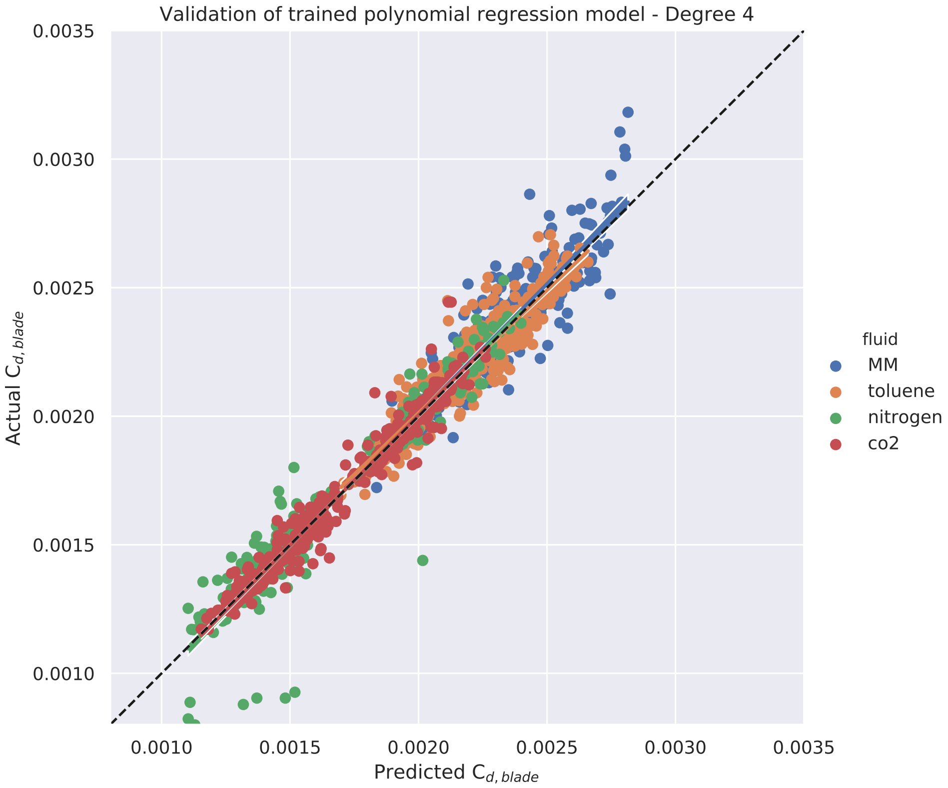


Figure 5.2: Comparison of actual and predicted machine learnt C$_{d,blade}$ values. The predictions are based on values that are taken from the test set, which is data that the model has never been subjected to.

5.1.1. Distribution of dissipation coefficient with Reynolds number

From the results of the machine learnt model presented in Figure 5.2, it is observed that the cluster of C$_{d,blade}$ for different fluids is distributed differently. For instance MM has a higher range of C$_{d,blade}$ values than that of CO$_2$. This is also observed in Figure 4.14. This section attempts to provide an explanation for this behaviour.

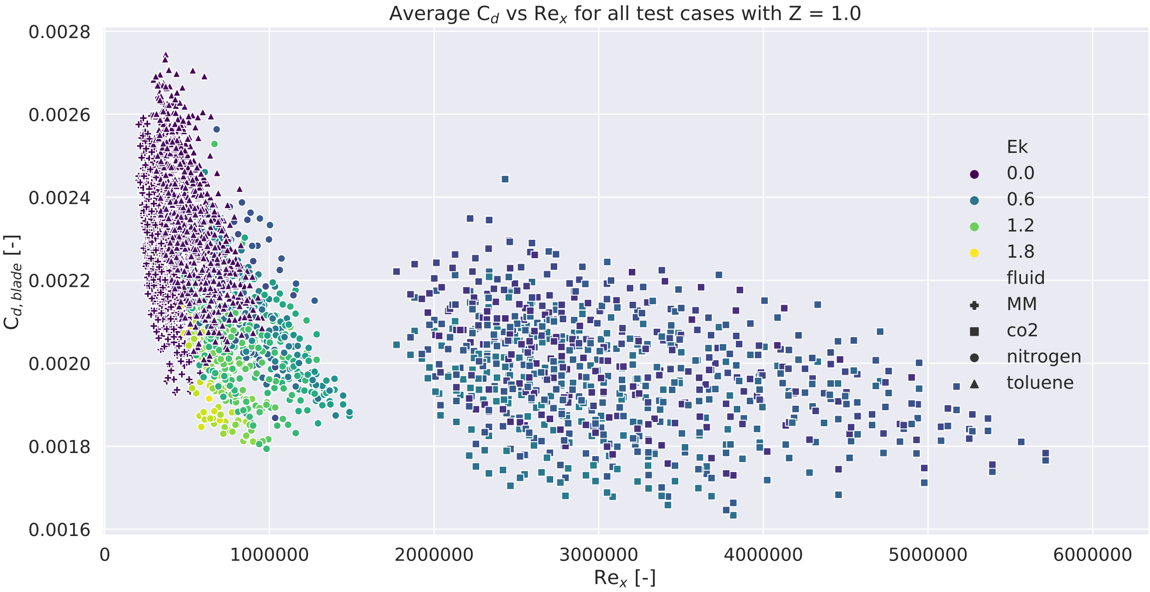


Figure 5.3: Distribution of average C$_{d,blade}$ with average Re$_x$ for all the test cases considered for Z = 0.7. The corresponding working fluid and Eckert number (Ek) for these test cases are highlighted as well.

To ease the comparison of C$_{d,blade}$ between different fluids, C$_{d,blade}$ is plotted against the corresponding freestream Reynolds number Re$_x$. Figure 5.3 shows this distribution for all test cases simulated with Z = 1. Each point on this plot represents a single test case. The position of each point on the x-axis (Re$_x$) is directly the consequence of the inlet conditions selected for each test case. The generic trend for all fluids shows that increase in Re$_x$ leads to a decrease in C$_{d,blade}$. As the Reynolds number is the ratio of inertial to viscous forces, higher Re$_x$ values imply a decrease in the viscous effects and therefore a decrease in C$_{d,blade}$. Unlike a correlation where a fixed Re$_x$ provides a unique C$_{d,blade}$ a range of C$_{d,blade}$ for a fixed Re$_x$ and fluid is observed in Figure 5.3. This is attributed to other parameters such as the duty coefficients (ϕ, ψ), geometrical parameters (tc, sf) and volumetric flow ratio (α) considered for the present study.

There is also a range of C$_{d,blade}$ for a fixed Re$_x$ but *varying* fluids in Figure 5.3. For instance at Re$_x$ = 620000 C$_{d.blade}$ ranges from 0.0018 to 0.0024. In this range the test cases simulated with Toluene are clustered at the top and those simulated with N$_2$ are clustered at the bottom. The cause for this behaviour can be explained using the Eckert number and the C$_d$ - Re$_x$ relation.

The Eckert number (Ek = $\frac{v^2}{C_pT}$) is the ratio of the kinetic energy to the enthalpy of the thermal boundary layer. Pini et al. [58] express C$_d$ using Falkner-Skan transformation as:

$$C_d = \frac{1}{\sqrt{Re_x)}} \int_0^{\eta_e} CR \frac{T_{\delta_e}}{T} f''^2 d\eta \tag{5.1}$$

where Re$_x$ is the freestream Reynolds number, T$_{\delta_e}$/T the ratio of freestream temperature and temperature used to measure viscous dissipation in the boundary layer, CR is the

Chapman-Rubesin parameter, η is a similarity variable and f" = $\frac{1}{v_{\delta_e}}\frac{dV_x}{d\eta}$.

Complex molecules have a higher heat capacity (Cp) than simple molecules and thus a lower Ek. This translates in a less steeper gradient of temperature in proximity of the wall, hence, the kinematic and thermal boundary layers can be assumed to be decoupled and consequently complex molecules can be assumed to have isothermal boundary layers. This means $\frac{T_{\delta_e}}{T}$ in Equation (5.1) is ≈ 1. Findings of Pini et al. show that for complex fluids CR does not exhibit appreciable gradients. Hence the integral to determine C$_d$ in Equation (5.1) is irrelevant and C$_d$ is inversely related to $\sqrt{Re_x}$ for complex fluids. On the contrary, simpler molecules have higher Ek and steeper temperature gradients so $\frac{T_{\delta_e}}{T} < 1$ that translates in lower C$_d$ as compared to complex fluids for the the same Re$_x$. This corroborates the results shown in Figure 5.3.

Figure 5.4 shows the distribution of C$_{d,blade}$ against the corresponding freestream Reynolds number for all test cases simulated with Z = 0.7. Similar to Figure 5.3 the position of each point on Re$_x$ is determined by the inlet conditions. As the Re$_x$ for all fluids for Z = 0.7 is higher than the case for Z = 1, as per the definition of Reynold's number the C$_{d,blade}$ for Z = 0.7 should be lower. This holds for N_2 and CO_2 but not for MM and Toluene. This discrepancy is attributed to the influence of Z and the molecular complexity of the working fluid.

The range of Ek for the case of Z = 0.7, shown in Figure 5.4, is lower than that of Z = 1. Results of Pini et al. find that the reduction in CR close to the wall is stronger for CO_2 than MM as critical conditions are approached [58]. Thus, in the case of Z = 0.7 C$_{d,blade}$ decreases more for simpler fluids CO_2 and N_2 than the case of Z = 1, whereas is almost no change is observed in C$_{d,blade}$ for MM and Toluene.

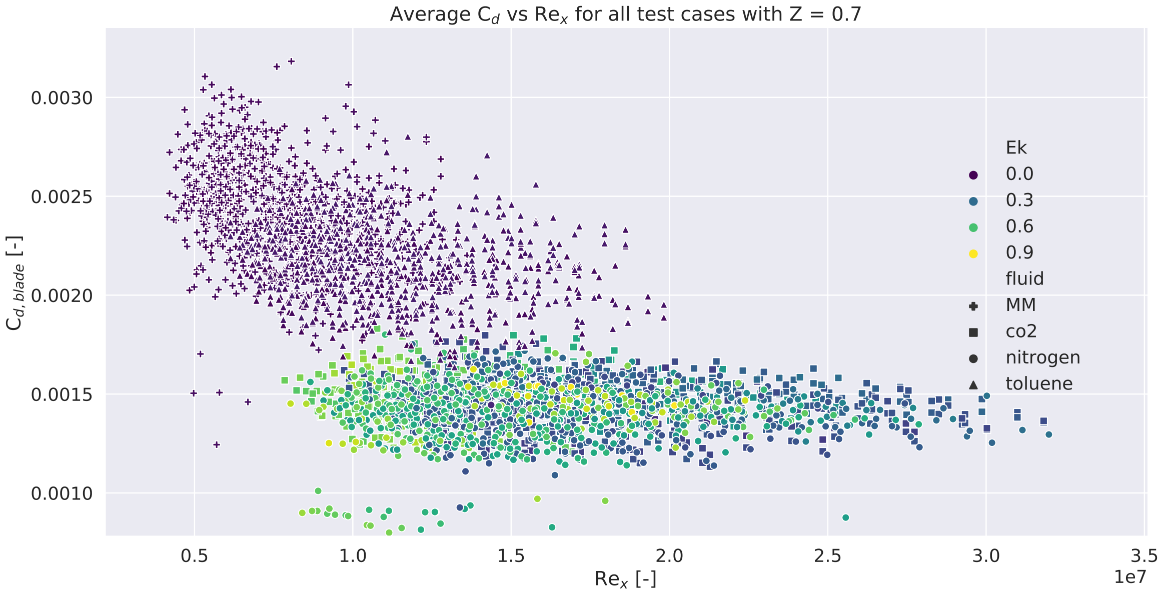


Figure 5.4: Distribution of average C$_{d,blade}$ with average Re$_x$ for all the test cases considered for Z = 0.7. The corresponding working fluid and Eckert number (Ek) for these test cases are highlighted as well.

For the sake of completeness the relation of C$_{d,blade}$ and Re$_\theta$ is illustrated in Figure 5.5. Re$_\theta$ is calculated by taking the average of the pressure and suction surface of the blade. Compared to Figure 2.5 and 2.6, the Re$_\theta$ of Figure 5.5 is an order higher.

Figure 5.5: Distribution of average $C_{d,blade}$ with average Re_θ for all the test cases. The corresponding working fluid and Eckert number (Ek) for these test cases are highlighted as well. $0.0056Re_\theta^{-1/6}$ line is the turbulent correlation developed by Schlichting.

5.2. Integration of the machine learnt model in TurboSim

The machine learnt model described in Section 5.1 considers $\phi, \psi, tc, sf, Z, \alpha$ and working fluid as inputs, however, the physics-based loss model TurboSim considers all inputs except the blade geometrical parameters (tc and sf). This calls the need for a new machine learnt model that is not a function of the blade geometrical parameters. Subsequently, tc and sf are removed from the database and therefore the new database contains features ϕ, ψ, Z, α and working fluid and label $C_{d,blade}$. This new database is used to train and validate a new model using the same methodology discussed in Section 3.4.

The hyperparmeter tuning finds that polynomial regression of degree 4 performs the best for the new model too. The evaluation scores of training and validating the machine learnt model *without* tc and sf are summarised and compared with the machine learnt model *with* tc and sf in Table 5.1.

Table 5.1: Evaluation scores RMSE and R-square for a 4^{th} order polynomial regression machine learnt model with and without blade geometrical parameters tc and sf.

	Training data		**Testing data**	
Machine learnt model	RMSE [10^{-4}]	R-square	RMSE [10^{-4}]	R-square
with tc and sf	0.763	0.963	0.812	0.959
without tc and sf	1.028	0.932	1.042	0.932

As tc is correlated to $C_{d,blade}$ as per the discussion in Section 4.7 the performance score of the machine learnt model without tc is lower than that of the one with this parameter. The evaluation scores of the model without tc and sf obtained is considered to be suitable for integrating the model into TurboSim. For the sake of brevity the polynomial equation of the machine learnt model without geometric factors is presented in Appendix A.

5.3. Influence of the machine learnt model on profile losses

There are two variants of TurboSim. The first is with the machine learnt model and the second without the machine learnt model. For conciseness from hereon TurboSim with and without the machine learnt model are abbreviated as TSML and TSC respectively. Both of these models output the profile loss percent (η_{bbl}), which is defined as the difference between 100% efficiency and the efficiency of the blade performance in terms of profile loss.

$$\eta_{bbl} = 100\% - \eta \tag{5.2}$$

To estimate the profile loss percent both TSC and TSML require α, Z, fluid and $C_{d,blade}$ as inputs. TSC assumes $C_{d,blade}$ = 0.002 and is independent of α, Z and fluid. Whereas for TSML $C_{d,blade}$ is determined using the machine learnt model and varies with change in α, Z and fluid. The objective of this section is to study the change in profile loss percent (η_{bbl}) estimated by TSML and TSC for varying input parameters.

The setup of this study is as follows. Both, TSML and TSC are executed for a fixed set of working fluid, Z, α and a fixed range of duty coefficients. The difference of the two estimations of profile losses is calculated using Equation (5.3).

$$\Delta\eta_{bbl} = \eta_{bbl,TSC} - \eta_{bbl,TSML} \tag{5.3}$$

where $\eta_{bbl,TSC}$ is the profile loss percent determined using TSC and $\eta_{bbl,TSML}$ is the profile loss percent determined using TSML. For a fixed set of inputs the change in profile loss percent $\Delta\eta_{bbl}$ is a measure of over or underestimation of profile loss by TSC with respect to TSML.

Since a range of duty coefficients are considered $\Delta\eta_{bbl}$ is plotted as a contour on a $\phi - \psi$ diagram. It is important to note that the upper left hand corner of the plot is blank for all the contour plots shown in the following subsections. This is because the range of duty coefficients considered for this study is the same as the one mentioned in Table 3.9. The discussion in Section 3.3.5 describes why some blade geometries are not simulated due to errors in meshing.

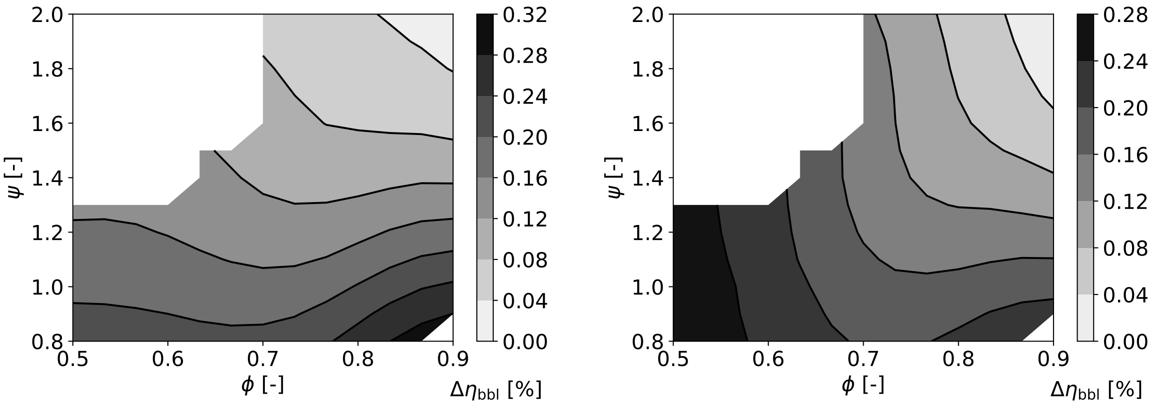
5.3.1. Influence of fluid on the profile losses

To illustrate the influence of varying the fluid on the profile loss percentage estimated by TSC and TSML all other input parameters are kept constant; their values along with loss model and optimisation setting used in TurboSim are tabulated in Table 5.2. The fluid is varied by considering N_2, CO_2, Toluene and MM as inputs. Figure 5.6 shows the trend of $\Delta\eta_{bbl}$ with ϕ and ψ for four different fluids.

Table 5.2: Values of parameters used for generating the variation of $\Delta\eta_{bbl}$ with $\phi - \psi$ for four fluids with inlet Z = 1. Other inputs required by TurboSim are provided as well.

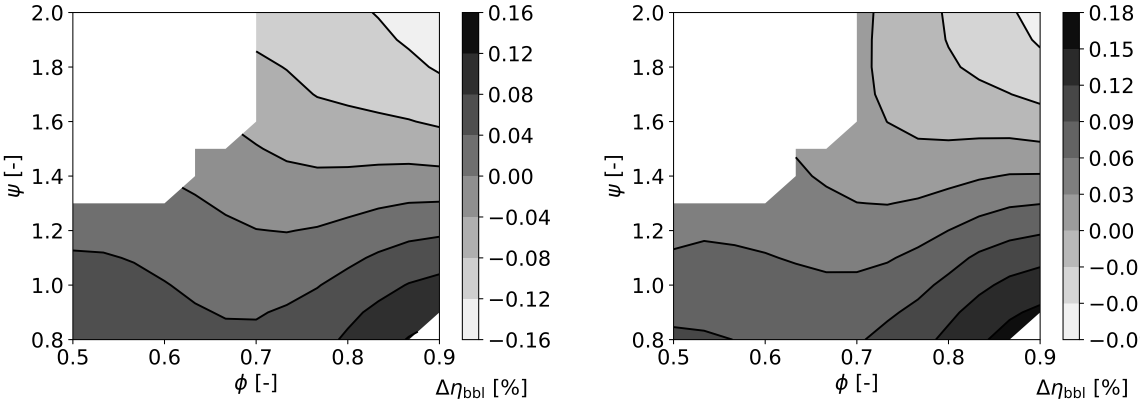
$fluid$	**(T_r, P_r, Z)$_{inlet}$**	α	**Loss Model**	**Optimisation**
N_2	(3.00,0.15,1.0)	2	Physical - BBL	No
CO_2	(2.50,0.50,1.0)	2	Physical - BBL	No
Toluene	(1.20,0.05,1.0)	2	Physical - BBL	No
MM	(0.92,0.03,1.0)	2	Physical - BBL	No

In Figure 5.6, the positive values of $\Delta\eta_{bbl}$ indicates that TSC overestimates the value of the profile loss or η_{bbl}. Negative values indicate underestimation. The magnitude of over or underestimation of profile losses varies with varying fluids. The largest absolute value $\Delta\eta_{bbl}$ observed for each fluid is 0.32%, 0.28%, 0.18% and 0.16% for N_2, CO_2, MM and Toluene



(a) N_2

(b) CO_2



(c) Toluene

(d) MM

Figure 5.6: Contour maps of the difference in profile loss percentage points $\Delta\eta_{bbl}$ obtained from TSC and TSML. All contour plots are generated for the same α = 2 and Z = 1 but represent different working fluids.

respectively. In Figure 5.6a and 5.6b TSC overestimates $C_{d,blade}$ than TSML for N_2 and CO_2 for the fixed set of parameters reported in Table 5.2. This behaviour is justified using the conclusions drawn from distribution plot of $C_{d,blade}$ in Figure 4.14. The mean of the distribution of $C_{d,blade}$ for N_2 was the lowest followed by that of CO_2. The mean $C_{d,blade}$ for all the observations of these fluids is lower than 0.002. This also explains why the largest value for $\Delta\eta_{bbl}$ is observed for N_2.

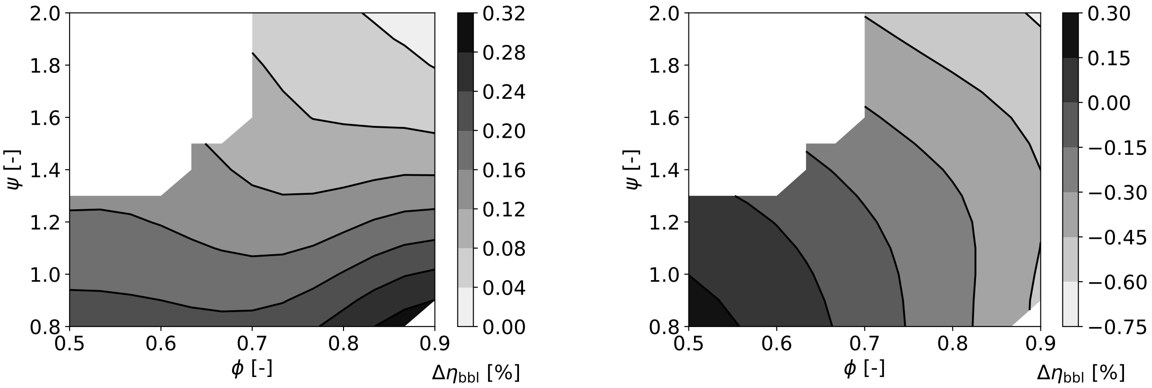
5.3.2. Influence of Z on profile losses

The Z is determined using a combination of T_r and P_r. Two values of Z are considered, namely Z = 1 and Z = 0.7. The influence of varying the Z on the profile loss percentage estimated by TSC and TSML is studied for two fluids differing in molecular complexity, namely N_2 and MM. All other input parameters are kept constant; their values are tabulated in Table 5.3. Figure 5.7 shows the trend of $\Delta\eta_{bbl}$ with ϕ and ψ for the two different fluids.

The TSC underestimates the profile loss percentage for more combinations of duty coefficients for Z = 0.7 than Z = 1. For Z = 1 the profile loss percentage and therefore $C_{d,blade}$ for all combinations of N_2 are overestimated by TSC. On the other hand TSC underestimates profile losses for all combinations of duty coefficients simulated using MM and Z = 0.7 (Figure 5.7d). The largest absolute value of $\Delta\eta_{bbl}$ observed for this trend is 3.50% percentage points. This

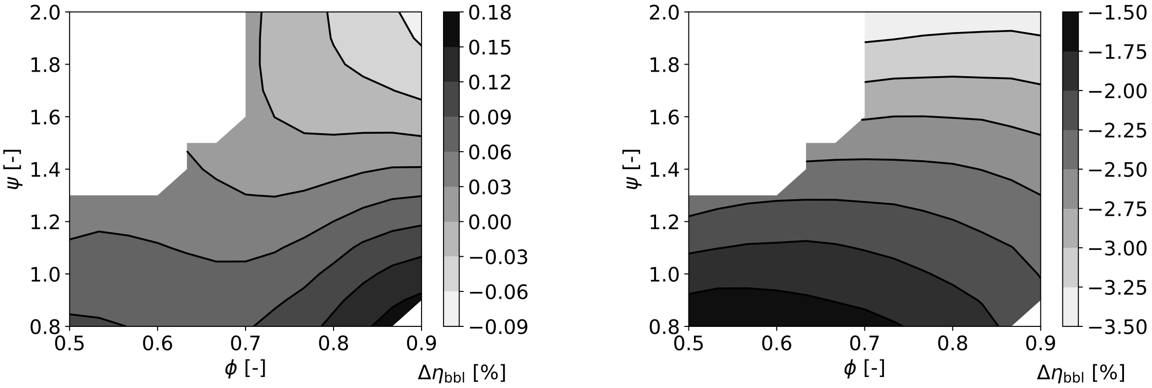
Table 5.3: Values of parameters used for generating the variation of $\Delta\eta_{bbl}$ with $\phi - \psi$ for four fluids with inlet Z = 0.7. Other inputs required by TurboSim are provided as well.

$fluid$	**(T$_r$, P$_r$, Z)$_{inlet}$**	α	**Loss Model**	**Optimisation**
N_2	(3.00,0.15,1.0), (1.18, 0.95, 0.7)	2	Physical - BBL	No
CO_2	(2.50,0.50,1.0), (1.18, 0.95, 0.7)	2	Physical - BBL	No
Toluene	(1.20,0.05,1.0), (1.08, 0.90, 0.7)	2	Physical - BBL	No
MM	(0.92,0.03,1.0), (1.08, 0.95, 0.7)	2	Physical - BBL	No



(a) N_2, Z = 1

(b) N_2, Z = 0.7



(c) MM, Z = 1

(d) MM, Z = 0.7

Figure 5.7: Contour maps of the difference in profile loss percentage points $\Delta\eta_{bbl}$ obtained from TSC and TSML. All contour plots are generated for the same α = 2 and $fluid$ =MM and N_2. Each column represents different Z.

shows that the use of a constant value for C$_{d,blade}$ can severely underestimate the losses being generated in blade passages of blade profiles with low duty coefficients and operating with the conditions reported in Table 5.3. Another plausible explanation for this behaviour is the distribution of C$_{d,blade}$ for MM higher than 0.002 for Z = 0.7 in Figure 5.5. However, this is the case ofr Z = 1.0 as well yet there are not such significant estimations of the profile loss percent.

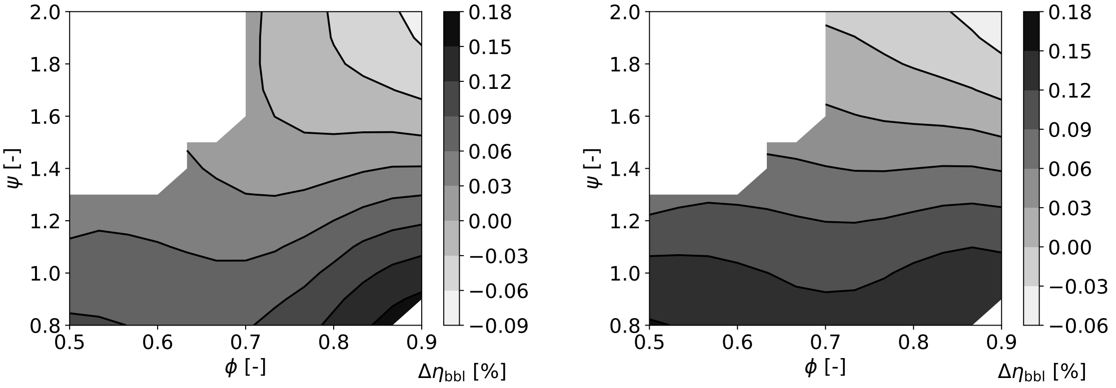
The almost uniform influence of Z on the gradient of $\Delta\eta_{bbl}$ vs ϕ, ψ is visible from the two contour plots for each fluid in Figure 5.7. For both fluids the trends are similar.

5.3.3. Influence of volumetric flow ratio on profile losses

To demonstrate the influence of varying the α on the profile loss percentage estimated by TSC and TSML all other input parameters are kept constant; their values are tabulated in Table 5.4. Figure 5.8a and 5.8b show the trend of $\Delta\eta_{bbl}$ with ϕ and ψ for α = 2 and 3, respectively.

Table 5.4: Values of parameters used for generating the variation of $\Delta\eta_{bbl}$ with $\phi - \psi$ for α = 2 and 3. The values of inlet Z = 1.0 and working fluid = MM is fixed. Other inputs required by TurboSim are provided as well.

$fluid$	**(T$_r$, P$_r$, Z)$_{inlet}$**	α	**Loss Model**	**Optimisation**
MM	(0.92,0.03,1.0)	2	Physical - BBL	No
MM	(0.92,0.03,1.0)	3	Physical - BBL	No



(a) α = 2

(b) α = 3

Figure 5.8: Contour maps of the difference in profile loss percentage points $\Delta\eta_{bbl}$ obtained from TSC and TSML. All contour plots are generated for the same Z = 1 and $fluid$ = MM but represent different α.

The range of $\Delta\eta_{bbl}$ observed for the range of duty coefficients and parameters reported in Table 5.4 is similar for both α = 2 and 3. $\Delta\eta_{bbl}$ for high ϕ ψ is where the difference in the two contours Figure 5.8. The maximum $\Delta\eta_{bbl}$ is 0.18% for both trends.

Results of this subsection show that the use of a machine learning model to predict C$_{d,blade}$ can help correct the estimate of profile loss percentage by as high as 3.50% points. Given that this value is observed for MM operating at inlet Z = 0.7 substantiates the importance of using a machine learnt model to estimate losses of axial turbomachines operating with organic fluids, such as MM.

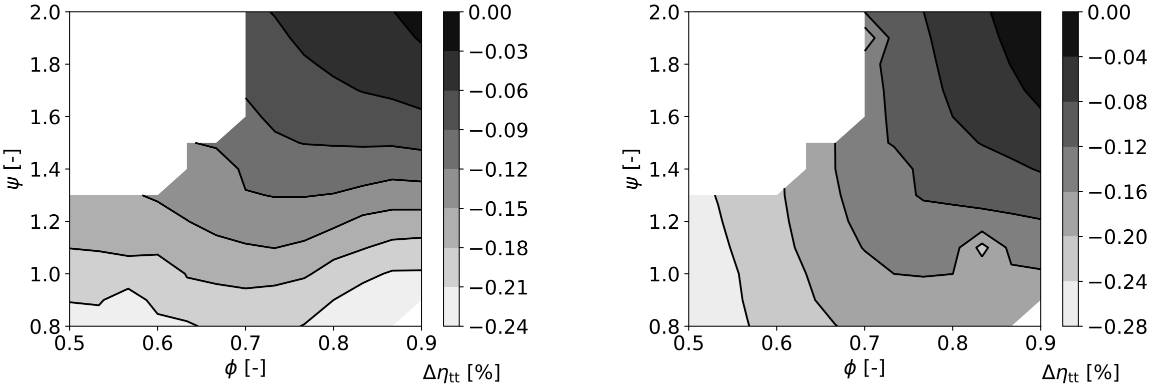
5.4. Influence of the machine learnt model on stage efficiency

The results of previous section highlights the importance of using a machine learnt model to predict the dissipation coefficient, which in turn helps estimate the profile loss percent better. The objective of this section is to present the impact of the machine learning model developed on the global model of TurboSim. The impact of the machine learnt model on the total-to-total stage efficiency η_{tt} for the same cases as those presented in Section 5.3 is discussed in this section.

The influence of varying the inputs fluid, α and Z on the η_{tt} is assessed. The setup of this study is same as that described in Section 5.3. Equation (5.4) is used to determine the difference between the $\eta_{tt,TSC}$ obtained from the TSC model and $\eta_{tt,TSML}$ obtained from the TSML model. The contour of $\Delta\eta_{tt}$ is made on a $\phi - \psi$ diagram to present the influence of TSC and TSML with varying input parameters.

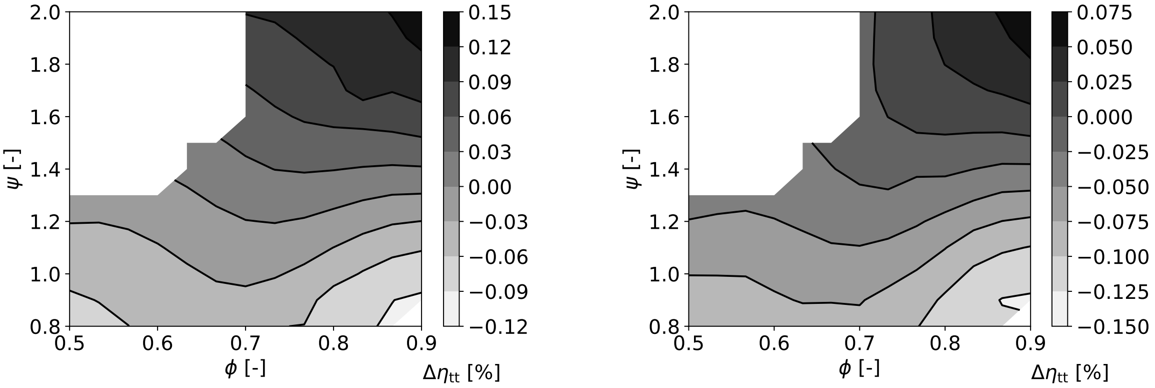
$$\Delta\eta_{tt} = \eta_{tt,TSC} - \eta_{tt,TSML} \tag{5.4}$$

5.4.1. Influence of fluid on stage efficiency



(a) N_2

(b) CO_2



(c) Toluene

(d) MM

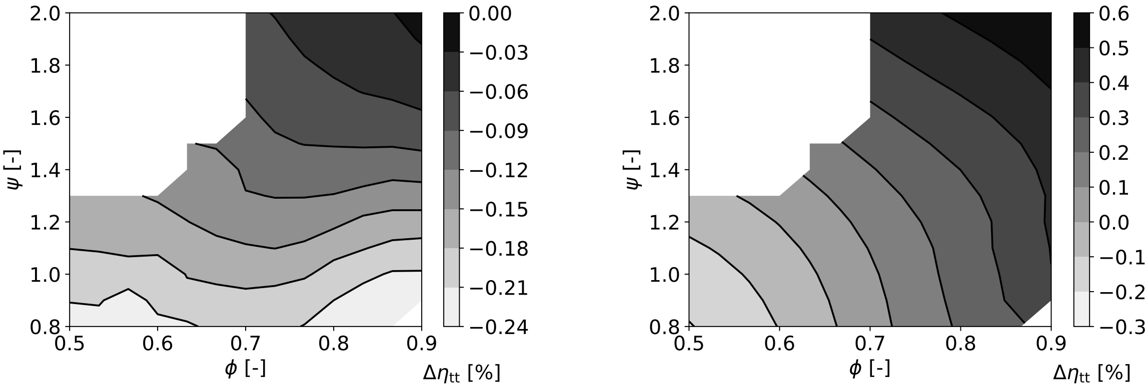
Figure 5.9: Contour maps of the difference in total-to-total stage efficiency $\Delta\eta_{tt}$ obtained from TSC and TSML. All contour plots are generated for the same α = 1 and Z = 1 but represent different fluids.

Figure 5.9 presents the influence of fluid on the stage efficiency for the set of parameters tabulated in Table 5.2. Negative values of $\Delta\eta_{tt}$ imply the TSC under estimates the efficiency and vice versa. All efficiencies for all combinations of duty coefficients are underestimated for N_2 and CO_2 by the constant TurboSim model. These results are in line with those found in Figure 5.6a. The regions of the highest under estimation of η_{tt} corresponds to the regions of highest over estimations of η_{bbl}. Consequently, the contour map is similar visually. The same is the case for the other fluids. Note, that the difference in profile loss $\Delta\eta_{bbl}$ is not the same as the difference in the total-to-total efficiency $\Delta\eta_{tt}$.

5.4.2. Influence of Z on stage efficiency

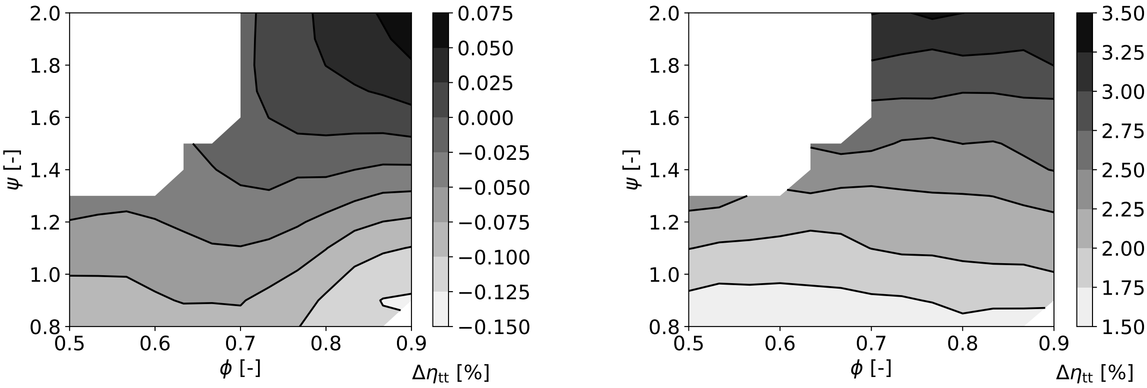
Figure 5.10 illustrates the influence of $\Delta\eta_{tt}$ with $\phi - \psi$ for different Z and fluids. The parameters used to generate these results are summarised in Table 5.3.

According to Figure 5.10, η_{tt} is overestimated for most combination of duty coefficients simulated with fluids N_2 or MM under Z inlet = 0.7. Whereas, its the opposite for results



(a) N_2, Z = 1

(b) N_2, Z = 0.7



(c) MM, Z = 1

(d) MM, Z = 0.7

Figure 5.10: Contour maps of the difference in total-to-total efficiency $\Delta\eta_{tt}$ obtained from TSC and TSM. All contour plots are generated from the same α = 2 and $fluid$ = MM and N_2. Each column of plots represents different Z.

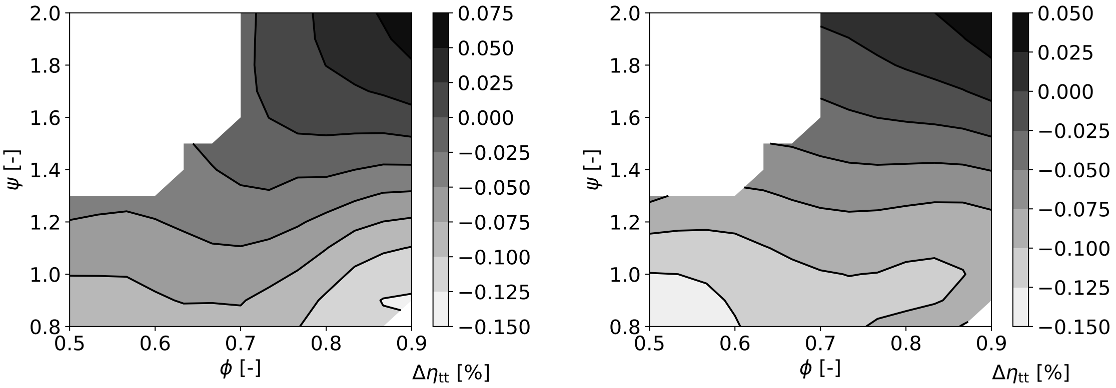
simulated with inlet Z = 1.0. These results too are in line with those found in the discussion of Figure 5.7. The values do not directly correlate but are within the same range. While the highest $\Delta\eta_{tt}$ of 3.50% points is for regions of high duty coefficients, it is still a significant quantity to over estimate. The highest value of under estimation is -0.3% points for blade profiles with low duty coefficients and operating with N_2.

5.4.3. Influence of volumetric flow ratio on stage efficiency.

Simulations are performed on a range of duty coefficients with input parameters reported in Table 5.4. α is varied from 2 to 3 to observe its influence on $\Delta\eta_{tt}$. This is done for MM. The results are plotted in Figure 5.10.

The change in α has no significant impact on the magnitude of $\Delta\eta_{tt}$ for the two cases. The TSC underestimates for regions with increasing ψ fom 0.8 to 1.6 and then over estimates for higher values of ψ greater than 1.6. The observations shown in Figure 5.11 correspond to the trends of the profile losses shown in Figure 5.8.

The results presented and discussed in Section 5.3 and 5.4 show that there is a difference in profile loss and η_{tt} when considering $C_{d,blade}$ = 0.002 or $C_{d,blade}$ as a function of α, Z and working fluid. The terms $\Delta\eta_{bll}$ and $\Delta\eta_{tt}$, introduced in these sections, are a measure of over or under estimation of the profile loss percentage and η_{tt} when determined using TSC as compared to TSML respectively.



(a) α = 2 (b) α = 3

Figure 5.11: Contour maps of difference in total-to-total efficiency $\Delta\eta_{tt}$ obtained from TSC and TSML. All contour plots are generated for the same Z = 1 and fluid = MM but represent different α.

Table 5.5 summarises the minimum and maximum $\Delta\eta_{bll}$ and $\Delta\eta_{tt}$ from the $\phi - \psi$ diagrams generated for each set of simulation executed in Section 5.3 and 5.4. Figure 5.12 visualises these values and depicts the influence of $\Delta\eta_{bbl}$ on $\Delta\eta_{tt}$. Together, the table and the figure show that the data driven loss model allows for an improvement in the estimation of the blade boundary layer losses or the profile losses which results in a total improvement of the stage efficiency. The extent to which the data-driven model allows for an improvement varies depending on the set of inputs. Based on the inputs tested the absolute value of improvement can be as small as 0.05% to as high as 3.50%. As the 3.50% improvement corresponds to MM operating close to critical conditions, which is similar to conditions of ORCs, the present study shows the importance of using a varying value of $C_{d,blade}$ in estimating losses than using a constant $C_{d,blade}$ = 0.002.

Table 5.5: Summary of the minimum and maximum values of $\Delta\eta_{bll}$ and $\Delta\eta_{tt}$ for the given inlet conditions. Negative values show underestimation by the constant $C_{d,blade}$ TurboSim model (TSC) as compared to the TurboSim model with the machine learnt model (TSML).

Input			$\Delta\eta_{bll}$		$\Delta\eta_{tt}$		Input conditions			$\Delta\eta_{bll}$		$\Delta\eta_{tt}$	
$fluid$	α	Z	Min	Max	Min	Max	$fluid$	α	Z	Min	Max	Min	Max
N_2	2	1.0	0.00	0.32	-0.24	0.00	N_2	2	0.7	-0.75	0.30	-0.30	0.60
CO_2	2	1.0	0.00	0.28	-0.28	0.00	MM	2	0.7	-3.50	-1.50	1.50	3.50
Tol	2	1.0	-0.16	0.16	-0.12	0.15	MM	2	1.0	-0.09	0.18	-0.15	0.07
MM	2	1.0	-0.09	0.18	-0.15	0.07	MM	3	0.7	-0.06	0.18	-0.15	0.05

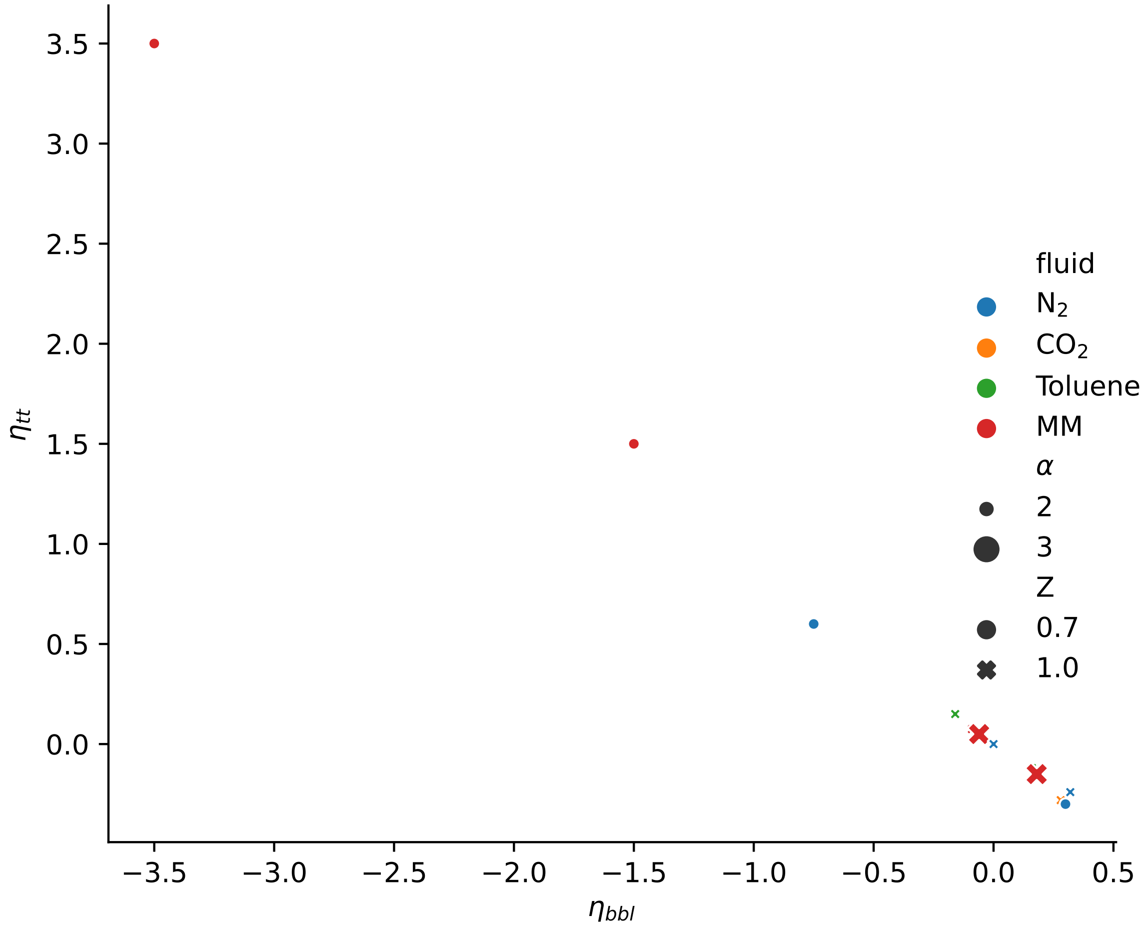


Figure 5.12: Visualisation of change in total-to-total efficiency percentage corresponding to change in profile loss percentage for the given inlet conditions. Negative values show underestimation by the constant C$_{d,blade}$ TurboSim model (TSC) as compared to the TurboSim model with the machine learnt model (TSML).

6

Conclusions and Recommendations

This chapter concludes the steps taken i) to generate a database of $C_{d,blade}$ for a combination of input parameters ii) to develop a machine learnt model using the generated database and iii) to study the over/underestimation of profile losses by the constant $C_{d,blade}$ model currently implemented in TurboSim. The key findings followed by the answers to the research questions obtained are presented in this chapter. Lastly, recommendations for future improvements are also listed.

6.1. Conclusion

It is realised that there remains a knowledge gap in determining $C_{d,blade}$ to predict boundary layer losses for axial turbines operating with organic fluids and the parameters that influence $C_{d,blade}$. Knowing Re_θ at a preliminary design phase is a challenge. So there is also a need for a model or function that allows a designer to estimate $C_{d,blade}$ using parameters, such as duty coefficients, volumetric expansion ratio, fluid, operating thermodynamic conditions and blade geometrical parameters, that are known at the preliminary design phase. In light of these considerations, the objective of the present work is to update the boundary layer loss estimation calculation in the physics-based model by developing a new data-driven loss model based on results from CFD simulations for the conceptual design phase for unconventional turbomachines.

A framework with two blocks is developed with the objective of generating a database containing measurements of $C_{d,blade}$ from CFD RANS simulations and using this database to develop a machine learnt model to estimate $C_{d,blade}$. The first block is used to generate 312 2D geometries of axial stators based on duty coefficients, ϕ, ψ and geometrical parameters, tc and sf. Meangen and Stagen of Multall suite are used for this task. The geometries are simulated using the open-source SU2 suite with varying Z, α and working fluid. In combination with the geometries this leads to 7485 unique test cases or instances for the database. Tecplot 360 is used to extract thermodynamic properties close to the edge of the boundary layer to estimate $C_{d,blade}$ for each test case and the value found is appended to the database.

A parametric analysis is performed using the results of the CFD simulations. The findings are as follows:

- The duty coefficients dictate the velocity triangle and the outlet flow angle of the stator, which in turn influences the pitch of the passage and therefore the flow through it. The dissipation coefficient increases with increasing duty coefficients. The increase is more dependent on ψ than ϕ.

- Out of the two geometrical parameters considered, the maximum thickness to chord ratio tc and the thickness distribution sf, the prior influences $C_{d,blade}$ more than the latter.

Higher tc increases the blade surface length the flow has to traverse, which increases the viscous dissipation in the boundary layer. The extent of the influence of sf on $C_{d,blade}$ is unknown as no significant variation in the entropy distribution over two blades with different sf and same simulation conditions was observed.

- The fluid affects the $C_{d,blade}$ value for a blade operating with same Z, α, duty coefficients and geometrical parameters. For test cases simulated with similar freestream Reynolds number, $C_{d,blade}$ of MM and Toluene is higher than that of CO_2 and N_2. Simpler molecules have higher Eckert numbers which means that the thermo-physical property gradients are more pronounced that that of complex molecules. This translates in lower $C_{d,blade}$ for simpler molecules and vice versa.
- Decreasing Z decreases $C_{d,blade}$ for CO_2 and N_2. The reason for this is attributed to the inlet thermodynamic conditions, especially the freestream entropy, when expanding a flow from close to critical conditions compared to ideal conditions. Expansion close to the critical points leads to a variation in γ_{pv} and thus an increase in Mach and presence of shockwaves in the flow. As the method used to estimate $C_{d,blade}$ does neglects the influence of boundary layer shockwave interaction and separation the $C_{d,blade}$ values observed are lower.

  Decreasing Z insignificantly decreases $C_{d,blade}$ for MM and Toluene. This discrepancy between complex and simpler fluids is due the property gradients within the boundary layer. Simpler fluids operating close to the critical point have stronger property gradients and therefore lower dissipation coefficient as compared to operating close to ideal conditions. Complex fluids have insignificant thermal gradients due to high heat capacity, hence insignificant changes in $C_{d,blade}$ for changes in operating conditions.
- A study is performed on the local Mach flow for inlet γ_{pv} lower than γ. Two cases are considered, N_2 ($\gamma_{pv} = \gamma$), MM ($\gamma_{pv} < \gamma$) . This study finds that the average Mach number along the flow passage increases for MM than for N_2 and the average γ_{pv} decreases for MM more than N_2. This is because the same γ_{pv} for two different fluids corresponds to Z = 1 for N_2 and Z = 0.57 for MM. Hence MM observes a more prominent variation in local flow conditions than N_2.
- Increasing α increases the local flow Mach number in the flow, which increases the presence of shockwaves and therefore decreases the $C_{d,blade}$ value. The decrease is due to the incapability for the $C_{d,blade}$ calculation method to account for shock interactions and separation.
- The Pearson correlation coefficient is used to quantify the influence of each input parameter with $C_{d,blade}$. The decrease order in which the parameters are found to be correlated with $C_{d,blade}$ is $\psi > \phi > tc > \alpha > Z > sf$.

The database, to which the results from the CFD are appended, contains $C_{d,blade}$, $fluid$, Z, α, ϕ, ψ, tc and sf. Multiple regression algorithms, such as Support Vector Machines, Polynomial Regression and Decision Trees, under the branch of supervised machine learning are trained. Polynomial Regression of 4th order degree is found to have the lowest RMSE and highest R-square scores for training and validation. The test scores of this model are at power with that of training and validation. However, as the physics based model does not contain tc and sf as inputs, another model without these parameters is trained. The RMSE and R-square for this model is 1.028 and 0.93, respectively. This model too has test scores at power with the training and validation.

To study the influence of the machine learnt model on profile and total-to-total stage efficiency estimated by TurboSim, the results of the original implementation of TurboSim i.e with constant C$_{d,blade}$ = 0.002 (TSC) is compared with the TurboSim with Machine Learning (TSML). The main findings are as follows:

- For the combinations of all inputs under which TSC and TSML are simulated, it is found that TSC over and understimates profile loss percent η_{bbl} within 0.32% and -0.75% respectively for all fluids and inlet thermodynamic conditions except MM close to critical conditions.
- For the combinations of all inputs under which TSC and TSML are simulated, it is found that TSC over and understimates total-to-total stage efficiency η_{tt} within 0.60% and -0.30% respectively for all fluids and inlet thermodynamic conditions except MM close to critical conditions.
- The exceptional case of MM operating close to critical conditions shows an underestimation of profile loss percent by 3.50%. This might be because the range of C$_{d,blade}$ values observed for all test cases simulated under similar conditions is mostly higher than 0.002. However, this range of C$_{d,blade}$ for MM at Z = 0.7 is similar to that of Z = 1, yet the latter case does not portray such significant profile loss percents.

  Therefore, it is concluded that the plausible cause for the 3.5% is incorrect simulations of different blade profiles at Z = 0.7. This highlights the importance of ensuring the quality of data in the database for machine learning.
- While there is no validation data to check the losses estimated by the machine learnt model, its implementation demonstrates that its application can lead to more accurate estimations of losses for different combinations of design parameters at the preliminary design phase.

6.2. Research answers

Based on the findings the answers to the research questions are as follows:

- **What is the effect of flow compressibility, working fluid and blade geometrical parameters on C$_d$ for boundary layers of axial stators?**

  The present work assesses the influence of seven parameters, namely $\phi, \psi, tc, sf, \alpha, Z$ and $fluid$ on the dimensionless dissipation coefficient.

  α is the volumetric flow ratio across the stator blade and influences the compressibility of the flow. An increase or decrease in this parameter influences the pressure ratio across the blade. Three values of α, 2,3 and 4 are considered for the analysis. According to the results obtained from the CFD simulations, increasing α leads to decreasing C$_{d,blade}$ values because of increase in shocks in the flow passage. The focus of the present work is to assess losses from boundary layers free of shocks or separation, so any such influence is neglected in the calculations, which therefore causes the value of C$_{d,blade}$ to decrease. Out of the three values, depending on the working fluid being used it is recommended that α = 4 should be avoided for axial stator blades as this introduces shock losses and calls for the use of convergent-divergent blades.

  The working fluid plays a role in the magnitude of C$_{d,blade}$. Four working fluids, namely N$_2$, CO$_2$, Toluene and MM are considered. Irrespective of the values of tc, sf, ϕ, ψ, Z or α N$_2$ and CO$_2$ has lower dissipation coefficient than MM and Toluene. Such behaviour is attributed to molecular complexity which subsequently influences the Eckert number and the inlet thermodynamic conditions which dictates Z and the inlet Reynolds number Re$_{x}$. The influence of Eckert number is more prominent on C$_{d,blade}$ for simpler molecules whereas the inverse square root of Re$_{x}$ influences C$_{d,blade}$ for complex molecules.

tc and sf are the maximum thickness to chord ratio and blade thickness distribution respectively. These two parameters are selected to define the blade geometry because of the capability of the 1D meanline design tool Meangen. Rather than changing the values of these parameters individually, the influence is studied by taking these two in sets. This reduces the dimensionality of the problem while providing some insight into its influence. Both increase the magnitude of C$_{d,blade}$, however, the results show that the blade thickness changes the entropy generation along the blade surface more than the thickness distribution of the blade. Therefore, out of the two the maximum thickness to chord ratio must be considered in estimating C$_{d}$.

The order in which these parameters influence the C$_{d,blade}$ is the working fluid, followed by α and tc.

- **Which 2D blade profiles define the design space being studied?**

  The 2D profiles of the stator blades are defined on the basis of duty coefficients ϕ and ψ. A, commonly used, constant degree of reaction value of 0.5 is assumed for all the combinations of $\phi - \psi$ considered for this analysis. The range of the two parameters is set based on the limit imposed by Meangen and second by the number of cases. The initial design space considered was from 0.8 to 4.0 in steps of 0.05 for ψ and 0.5 to 1.2 in steps of 0.1 for ϕ. The limit imposed by Meangen led to the decrease in this design space from 4.0 to 2.0 for ψ and from 1.2 to 0.9 for ϕ. This led to 225 unique designs for the blades.

  The simulation of 225 blades with combination of different parameters increases the size of the database being generated but at the cost of computational time. As the present work is limited in time this design space is kept the same in size but made less dense by increasing the steps of ψ to 0.1. So 65 unique designs instead of 225 cover the same design space.

- **What strategy will be used to develop the trained data-drivel loss model from 2D-CFD RANS simulations?**

  The database created contains features, the parameters, and a label, the C$_{d,blade}$. Before the models are trained the database is prepared to prevent biased or skewing of results. The approach used involves the following steps. First the data is split into a training and testing set using a ratio of 8:2. A pipeline consisting 2 steps, converting categorical data to numerical and standardisation of data, is created. The working fluid, which is the categorical data in the training and testing sets is converted to numeric values using the One-Hot Encoder. This approach generates a sparse matrix with a binary column for each category. The remaining features are standardised by subtracting the mean of the feature from the sample and dividing the difference by the standard deviation of that feature.

  This pipeline is used to convert the database to a standardised one. Five different models, namely Linear, Ridge, Lasso, second order Polynomial and Support Vector Machines, are trained using this database. Each model is trained using the k-cross validation technique. Out of all the models Polynomial regression shows the lowest Root Mean Square Error and highest R-square values. The scores of degrees of up to 8 are explored. From all, the 4th order polynomial has similar validation and training scores, implying that the model is neither over-fitting nor under-fitting the data. The RMSE and R-square scores of the final model is 7.628E-05 and 0.963 respectively.

Additional results, such as the learning curve plots (Figure 3.16), provide insight into the smallest size of database required to develop a Polynomial model with high R-square and low RMSE scores. The RMSE scores from the training and validation set are similar when the training set size is of 4000 instances. If the same ratio of 8:2 is used, then the minimum number of instances in the database should be 5000.

6.3. Recommendations

Given the use of multiple topics utilised for the present research, there are insights that can be used to improve the devised framework or for future research.

- Improving the simulation setup: The present work sets a basic foundation and demonstrates that results from CFD simulations can be used to create a database, which can further be used to develop a machine learnt model capable of predicting losses. The current setup has a few drawbacks that must be improved in order to add more instances to the database. The first is the automation of the meshing. The current automation procedure lacks an automated check on each grid. It was due to this that there was a region with no results in the selected design space. The second problem is in the determination of the number of cores required per simulations. Several simulations had to be rerun on different number of cores till the maximum for each simulation setup could be found. Consequently, manual intervention was required to ensure the convergence of several simulations. One option is to run all simulations on 2 cores, however that is at the cost of time. Lastly, the termination criteria of the simulation should be changed from number of iterations to reduction in order of convergence. This approach guarantees enhanced results but at the cost of time.
- Using a better tool for Design of Experiments: The present work lacks in an explicit relationship between the geometrical parameters and $C_{d,blade}$. A tool like Dakota, that is dedicated to performing design of experiments study, should be used. The features of this tool would allow one to add and explore more dimensions (parameters) effectively.
- Thorough analysis of results from a physical perspective: The results obtained demonstrate similar behaviour to what is expected in theory, however, this is not substantiated using a thorough analysis of the CFD simulations from a physical perspective. The studying of the progression of boundary layers, and thermal gradients is recommended.
- Validation study: This present work assumes adiabatic walls and uses equations of states which are models on their own. The trends obtained by implementing the machine learnt model in TurboSim must be validated with results from experimental results. This would highlight the accuracy in estimating over or under prediction of by the machine learnt model and the machine learnt model itself.
- Extending the study to incorporate the infuence of shockwaves and separation on the dissipation coefficient: The present model neglects the measurements gathered by the normals post shock impingement location. Therefore, it is recommended a better method to estimate $C_{d,blade}$ for such flows is used.
- Extending the study to estimate the blade boundary layer dissipation coefficient for the rotor : The boundary layer of the rotor is influenced by the wake of the stator, therefore it

is recommended to study the C$_{d,blade}$ for rotors and eventually extend the machine learnt model to predict C$_{d,blade}$ for both the stator and rotor at a preliminary design phase.

- Extending the study to estimate other loss coefficients: The present work demonstrates that a machine learnt model based on CFD RANS simulations can be built to estimate the boundary layer dissipation coefficient. Even though the present study lacks validation data, nevertheless, it demonstrates that coefficient values may deviate from what is commonly used in literature. A similar approach can be used to estimate the base pressure coefficient, which relies on experimental data. This coefficient too can be determined from 2D RANS simulations.
- Discover new relations: Instead of using supervised machine learning, unsupervised machine learning can be used either on a similar or a more detailed database to explore new relations. The challenge lies in connecting the obtained statistical relation to laws of physics.
- Increasing the size of the database: A larger database would allow one to implement state of the art Artificial Neural Networks or genetic programming algorithms.

Bibliography

[1] F. Guterl, nov 2018.

[2] J. D. Denton, “Loss mechanisms in turbomachines,” in *ASME 1993 International Gas Turbine and Aeroengine Congress and Exposition*, pp. V002T14A001–V002T14A001, American Society of Mechanical Engineers, 1993.

[3] U. E. I. Administration, “International energy outlook 2019 with projections to 2050,” tech. rep., U.S Energy Information Administration, sep 2019.

[4] I. E. Agency, “World energy balances 2020 edition : Database documentation.” Online, 2020.

[5] Ren21, “Renewables 2019 - global status report,” Tech. Rep. ISBN 978-3-9818911-7-1, Ren21: Renewables Now, 2019.

[6] S. Quoilin, M. Van Den Broek, S. Declaye, P. Dewallef, and V. Lemort, “Techno-economic survey of organic rankine cycle (orc) systems,” *Renewable and Sustainable Energy Reviews*, vol. 22, pp. 168–186, 2013.

[7] H. Refrigerants, “Turboden,” 2020.

[8] U. Drescher and D. Brüggemann, “Fluid selection for the organic rankine cycle (orc) in biomass power and heat plants,” *Applied thermal engineering*, vol. 27, no. 1, pp. 223–228, 2007.

[9] E. Macchi, “Theoretical basis of the organic rankine cycle,” in *Organic Rankine Cycle (ORC) Power Systems*, pp. 3–24, Elsevier, 2017.

[10] M. H. I. Group, “Turboden,” 2019.

[11] O. Balje and D. Japikse, “Turbomachines—a guide to design selection and theory,” 1981.

[12] S. L. Dixon and C. Hall, *Fluid mechanics and thermodynamics of turbomachinery*. Butterworth-Heinemann, 2013.

[13] A. Gioffre, “Design guidelines for axial turbines operating with non-ideal compressible flows.” Master’s thesis : Polytechnico di Milano, 2019.

[14] M. P. Andrea Giuffre, “Design guidelines for axial turbines operating with non-ideal compressible flows,” *ASME 2020 Turbo Expo - Virtual Conference Session Gallery*, vol. 46-08 Turbine Design Methods and Applications, sep 2020.

[15] H. Schlichting and K. Gersten, *Boundary-layer theory*. Springer, 2016.

[16] M. Pini and C. De Servi, “Entropy generation in laminar boundary layers of non-ideal fluid flows,” in *International Seminar on Non-Ideal Compressible-Fluid Dynamics for Propulsion & Power*, pp. 104–117, Springer, 2018.

[17] F. Pizzi, "Development and assessment of physical based profile loss model for orc turbine based on boundary layer analysis." Master's thesis : Polytechnico di Milano, 2018.

[18] B. G. Miller, "Clean coal technologies for advanced power generation," *Clean Coal Engineering Technology*, pp. 251–300, 2017.

[19] U. E. I. Administration, "Power plant operations report," Tech. Rep. EIA-920, U.S. Energy Information Administration, 2019.

[20] I. Johnson, W. T. Choate, and A. Davidson, "Waste heat recovery. technology and opportunities in us industry," tech. rep., BCS, Inc., Laurel, MD (United States), 2008.

[21] W.-h. Yang, X. Tao, J. Li-yue, G. Yue-jiao, C. Guang, and L. Wei, "Waste heat recovery and power generation in cement works," in *2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS)*, vol. 2, pp. 223–226, IEEE, 2009.

[22] S. Brückner, S. Liu, L. Miró, M. Radspieler, L. F. Cabeza, and E. Lävemann, "Industrial waste heat recovery technologies: An economic analysis of heat transformation technologies," *Applied Energy*, vol. 151, pp. 157–167, 2015.

[23] F. Starr, "High temperature materials issues in the design and operation of coal-fired steam turbines and plant," in *Structural Alloys for Power Plants*, pp. 36–68, Elsevier, 2014.

[24] I. H. Bell, J. Wronski, S. Quoilin, and V. Lemort, "Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library coolprop," *Industrial & engineering chemistry research*, vol. 53, no. 6, pp. 2498–2508, 2014.

[25] C. M. Invernizzi, "Closed power cycles," *Lecture Notes in Energy*, vol. 11, 2013.

[26] T. Yamamoto, T. Furuhata, N. Arai, and K. Mori, "Design and testing of the organic rankine cycle," *Energy*, vol. 26, no. 3, pp. 239–251, 2001.

[27] P. Colonna, E. Casati, C. Trapp, T. Mathijssen, J. Larjola, T. Turunen-Saaresti, and A. Uusitalo, "Organic rankine cycle power systems: from the concept to current technology, applications, and an outlook to the future," *Journal of Engineering for Gas Turbines and Power*, vol. 137, no. 10, 2015.

[28] P. A. Thompson, "A fundamental derivative in gas dynamics," *The Physics of Fluids*, vol. 14, no. 9, pp. 1843–1849, 1971.

[29] P. Colonna, S. Rebay, J. Harinck, and A. Guardone, "Real-gas effects in orc turbine flow simulations: influence of thermodynamic models on flow fields and performance parameters," in *ECCOMAS CFD 2006 Conference, Egmond aan Zee, NL*, 2006.

[30] E. S. Menon, *Pipeline planning and construction field manual*. Gulf Professional Publishing, 1978.

[31] D. Kouremenos and K. Antonopoulos, "Isentropic exponents of real gases and application for the air at temperatures from 150 k to 450 k," *Acta mechanica*, vol. 65, no. 1-4, pp. 81–99, 1987.

[32] P. Nederstigt, "Real gas thermodynamics: and the isentropic behavior of substances." Master's thesis : Delft University of Technology, 2017.

[33] P. Colonna, N. Nannan, A. Guardone, and T. Van der Stelt, “On the computation of the fundamental derivative of gas dynamics using equations of state,” *Fluid phase equilibria*, vol. 286, no. 1, pp. 43–54, 2009.

[34] A. Kluwick, “Shock discontinuities: from classical to non-classical shocks,” *Acta Mechanica*, vol. 229, no. 2, pp. 515–533, 2018.

[35] W. C. Reynolds and P. Colonna, *Thermodynamics*. Cambridge University Press, 2018.

[36] E. M. Greitzer, C. S. Tan, and M. B. Graf, *Internal flow: concepts and applications*, vol. 3. Cambridge University Press, 2007.

[37] T. Cebeci and L. W. Carr, “A computer program for calculating laminar and turbulent boundary layers for two-dimensional time-dependent flows,” Technical Report NASA-TM-78470, A-7340, NASA; United States, NASA Ames Research Center; Moffett Field, CA, United States, mar 1978.

[38] W. Dawes, “A comparison of zero and one equation turbulence modelling for turbomachinery calculations,” in *ASME 1990 International Gas Turbine and Aeroengine Congress and Exposition*, pp. V001T01A093–V001T01A093, American Society of Mechanical Engineers, 1990.

[39] E. Truckenbrodt, “A method of quadrature for the calculation of laminar and turbulent boundary layers in plane and rotational symmetric flow.,” Tech. Rep. NACA-TM-1379, National Advisory Committee for Aeronautics, National Advisory Committee for Aeronautics; Washington, DC, United States, may 1955.

[40] Q. D. Roberts and J. D. Denton, “Loss production in the wake of a simulated subsonic turbine blade,” in *ASME 1996 International Gas Turbine and Aeroengine Congress and Exhibition*, pp. V001T01A111–V001T01A111, American Society of Mechanical Engineers, 1996.

[41] D. Mee, N. Baines, M. Oldfield, and T. Dickens, “An examination of the contributions to loss on a transonic turbine blade in cascade,” in *ASME 1990 International Gas Turbine and Aeroengine Congress and Exposition*, pp. V005T16A012–V005T16A012, American Society of Mechanical Engineers, 1990.

[42] D. M. Pini, “Lecture 3: Similitude.” Lecture, feb 2017.

[43] T. M. Mitchell *et al.*, “Machine learning. 1997,” *Burr Ridge, IL: McGraw Hill*, vol. 45, no. 37, pp. 870–877, 1997.

[44] M. Copeland, “What’s the difference between artificial intelligence, machine learning and deep learning?.” NVIDIA website, jul 2016. `https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/.`

[45] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O’Reilly Media, 2019.

[46] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[47] COGNUB, “Cognitive computing and machine learning.” Online, 2019. http://www.cognub.com/index.php/cognitive-platform/.

[48] K. W. Lee and H. Choo, “Predicting user attitudes toward smartphone ads using support vector machine,” *International Journal of Mobile Communications*, vol. 14, no. 3, pp. 226–243, 2016.

[49] J. D. Denton, “Multall—an open source, computational fluid dynamics based, turbomachinery design system,” *Journal of Turbomachinery*, vol. 139, no. 12, 2017.

[50] A. Ghidoni, E. Pelizzari, S. Rebay, and V. Selmin, “3d anisotropic unstructured grid generation,” *International journal for numerical methods in fluids*, vol. 51, no. 9-10, pp. 1097–1115, 2006.

[51] T. D. Economon, F. Palacios, S. R. Copeland, T. W. Lukaczyk, and J. J. Alonso, “Su2: An open-source suite for multiphysics simulation and design,” *Aiaa Journal*, vol. 54, no. 3, pp. 828–846, 2016.

[52] A. Patel, R. Diez, R. Pecnik, *et al.*, “Turbulence modelling for flows with strong variations in thermo-physical properties,” *International Journal of Heat and Fluid Flow*, vol. 73, pp. 114–123, 2018.

[53] I. H. Bell, J. Wronski, S. Quoilin, and V. Lemort, “Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library coolprop,” *Industrial & Engineering Chemistry Research*, vol. 53, no. 6, pp. 2498–2508, 2014.

[54] Tecplot, “Tecplot 360.” Online, 2020. https://www.tecplot.com/products/tecplot-360/.

[55] K. M. Bower, “What is design of experiments?.” ASQ Excellence Through Quality, 2020.

[56] A. Lee, “pydoe: The experimental design package for python.” Online, 2013.

[57] J. W. Emerson, W. A. Green, B. Schloerke, J. Crowley, D. Cook, H. Hofmann, and H. Wickham, “The generalized pairs plot,” *Journal of Computational and Graphical Statistics*, vol. 22, no. 1, pp. 79–91, 2013.

[58] M. Pini and C. D. Servi, “Entropy generation in laminar and turbulent adiabatic boundary layers of non-idea fluid flows.” Delft University of Technology.

A

Regression equation for the machine learnt model

A polynomial regression model of degree 4 provides accurate approximation while preventing over-fitting of the data. The scikit-learn module of python is used to perform the polynomial fitting [46]. The polynomial equation that represents the machine learnt model consists of 5 inputs, $\phi, \psi, Z, \alpha, fluid$. As fluid is a classification term and is quantified using a binary matrix, the total number of inputs for the machine learnt model increases to 4 (ϕ, ψ, Z, α) + 4 $(fluids)$ = 8.

As the best fit is a 4th degree polynomial the resulting equation consists of 495 unique terms and corresponding coefficients. Out of these 495 terms 295 terms have zero as the coefficient. These 295 terms along with their coefficients are listed in Table A.1. The equation is formulated by multiplying the coefficient with the term and summing each combination. Therefore the equation is expressed as:

$$C_{d,blade} = (1\times 2.03E{-}3)+(1.36E{-}5\times\phi)+(2.56E{-}5\psi)+\ldots+(1.39E{-}5Z\alpha^2x4)+(7.27E{-}6Z\alpha x1^2) \tag{A.1}$$

where x1, x2, x3 and x4 correspond to MM, CO_2, N_2 and Toluene respectively.

Table A.1: Terms and coefficients of resulting 4th degree polynomial equation. x1, x2, x3 and x4 correspond to MM, CO$_2$, N_2 and Toluene respectively.

Terms	Coefficients	Terms	Coefficients	Terms	Coefficients
1	2.03E-03	$\phi\,\psi\,x1$	-6.43E-06	$Z\,\alpha\,x3$	-2.88E-06
ϕ	1.36E-05	$\phi\,\psi\,x2$	6.74E-06	$Z\,\alpha\,x4$	-7.34E-06
ψ	2.56E-05	$\phi\,\psi\,x3$	1.04E-05	$Z\,x1^2$	-5.87E-05
Z	3.99E-05	$\phi\,\psi\,x4$	-5.94E-06	$Z\,x2^2$	4.35E-05
α	-5.45E-06	$\phi\,Z^2$	1.36E-05	$Z\,x3^2$	5.88E-05
$x1$	6.06E-05	$\phi\,Z\,\alpha$	3.45E-06	$Z\,x4^2$	-3.80E-06
$x2$	-5.26E-05	$\phi\,Z\,x1$	-8.30E-06	α^3	-8.15E-06
$x3$	-4.72E-05	$\phi\,Z\,x2$	8.44E-06	$\alpha^2\,x1$	-2.32E-05
$x4$	3.92E-05	$\phi\,Z\,x3$	8.12E-07	$\alpha^2\,x2$	1.79E-05
ϕ^2	-4.99E-06	$\phi\,Z\,x4$	-6.93E-06	$\alpha^2\,x3$	1.12E-05
$\phi\,\psi$	4.73E-06	$\phi\,\alpha^2$	9.08E-06	$\alpha^2\,x4$	-5.53E-06
$\phi\,Z$	-5.99E-06	$\phi\,\alpha\,x1$	-1.07E-06	$\alpha\,x1^2$	-5.74E-06
$\phi\,\alpha$	1.77E-06	$\phi\,\alpha\,x2$	5.74E-07	$\alpha\,x2^2$	1.30E-06
$\phi\,x1$	-5.66E-06	$\phi\,\alpha\,x3$	7.82E-06	$\alpha\,x3^2$	5.50E-07
$\phi\,x2$	8.86E-06	$\phi\,\alpha\,x4$	-5.55E-06	$\alpha\,x4^2$	-1.57E-06
$\phi\,x3$	6.87E-06	$\phi\,x1^2$	-5.66E-06	$x1^3$	6.06E-05
$\phi\,x4$	3.54E-06	$\phi\,x2^2$	8.86E-06	$x2^3$	-5.26E-05
ψ^2	-5.26E-06	$\phi\,x3^2$	6.87E-06	$x3^3$	-4.72E-05
$\psi\,Z$	2.57E-06	$\phi\,x4^2$	3.54E-06	$x4^3$	3.92E-05
$\psi\,\alpha$	-9.78E-07	ψ^3	1.83E-06	ϕ^4	5.58E-06
$\psi\,x1$	2.54E-05	$\psi^2\,Z$	1.55E-06	$\phi^3\,\psi$	-6.01E-06
$\psi\,x2$	-6.97E-06	$\psi^2\,\alpha$	-5.47E-07	$\phi^3\,Z$	-3.37E-06
$\psi\,x3$	-4.46E-06	$\psi^2\,x1$	7.29E-07	$\phi^3\,\alpha$	-5.17E-07
$\psi\,x4$	1.17E-05	$\psi^2\,x2$	-2.24E-06	$\phi^3\,x1$	1.33E-06
Z^2	-2.66E-08	$\psi^2\,x3$	-4.77E-06	$\phi^3\,x2$	6.39E-07
$Z\,\alpha$	-4.79E-06	$\psi^2\,x4$	1.02E-06	$\phi^3\,x3$	6.20E-07
$Z\,x1$	-5.87E-05	$\psi\,Z^2$	2.56E-05	$\phi^3\,x4$	-9.96E-08
$Z\,x2$	4.35E-05	$\psi\,Z\,\alpha$	-7.50E-06	$\phi^2\,\psi^2$	2.15E-06
$Z\,x3$	5.88E-05	$\psi\,Z\,x1$	-7.78E-06	$\phi^2\,\psi\,Z$	5.24E-06
$Z\,x4$	-3.80E-06	$\psi\,Z\,x2$	4.75E-07	$\phi^2\,\psi\,\alpha$	2.99E-07
α^2	3.95E-07	$\psi\,Z\,x3$	7.57E-06	$\phi^2\,\psi\,x1$	-3.45E-06
$\alpha\,x1$	-5.74E-06	$\psi\,Z\,x4$	2.31E-06	$\phi^2\,\psi\,x2$	3.05E-06
$\alpha\,x2$	1.30E-06	$\psi\,\alpha^2$	-1.24E-05	$\phi^2\,\psi\,x3$	-4.16E-07
$\alpha\,x3$	5.50E-07	$\psi\,\alpha\,x1$	4.26E-06	$\phi^2\,\psi\,x4$	7.68E-06
$\alpha\,x4$	-1.57E-06	$\psi\,\alpha\,x2$	-7.68E-07	$\phi^2\,Z^2$	-4.99E-06
$x1^2$	6.06E-05	$\psi\,\alpha\,x3$	-9.84E-06	$\phi^2\,Z\,\alpha$	-5.89E-07
$x2^2$	-5.26E-05	$\psi\,\alpha\,x4$	5.36E-06	$\phi^2\,Z\,x1$	-7.81E-06
$x3^2$	-4.72E-05	$\psi\,x1^2$	2.54E-05	$\phi^2\,Z\,x2$	3.13E-06
$x4^2$	3.92E-05	$\psi\,x2^2$	-6.97E-06	$\phi^2\,Z\,x3$	6.08E-06
ϕ^3	2.49E-06	$\psi\,x3^2$	-4.46E-06	$\phi^2\,Z\,x4$	3.05E-06
$\phi^2\,\psi$	6.86E-06	$\psi\,x4^2$	1.17E-05	$\phi^2\,\alpha^2$	-5.63E-06
$\phi^2\,Z$	4.45E-06	Z^3	3.99E-05	$\phi^2\,\alpha\,x1$	5.22E-06
$\phi^2\,\alpha$	-2.74E-07	$Z^2\,\alpha$	-5.45E-06	$\phi^2\,\alpha\,x2$	-1.32E-06
$\phi^2\,x1$	4.79E-06	$Z^2\,x1$	6.07E-05	$\phi^2\,\alpha\,x3$	-6.65E-06
$\phi^2\,x2$	-6.62E-06	$Z^2\,x2$	-5.26E-05	$\phi^2\,\alpha\,x4$	2.48E-06
$\phi^2\,x3$	-7.02E-06	$Z^2\,x3$	-4.73E-05	$\phi^2\,x1^2$	4.79E-06
$\phi^2\,x4$	3.86E-06	$Z^2\,x4$	3.92E-05	$\phi^2\,x2^2$	-6.62E-06
$\phi\,\psi^2$	-4.10E-06	$Z\,\alpha^2$	2.13E-05	$\phi^2\,x3^2$	-7.02E-06
$\phi\,\psi\,Z$	-3.06E-06	$Z\,\alpha\,x1$	7.27E-06	$\phi^2\,x4^2$	3.86E-06
$\phi\,\psi\,\alpha$	-1.16E-07	$Z\,\alpha\,x2$	-1.84E-06	$\phi\,\psi^3$	-2.96E-06

Terms	Coefficients	Terms	Coefficients	Terms	Coefficients
$\phi\ \psi^2\ Z$	-8.79E-06	$\psi^3\ Z$	4.09E-06	$Z^3\ \alpha$	-4.79E-06
$\phi\ \psi^2\ \alpha$	2.11E-06	$\psi^3\ \alpha$	-4.68E-07	$Z^3\ x1$	-5.87E-05
$\phi\ \psi^2\ x1$	1.03E-05	$\psi^3\ x1$	-5.97E-06	$Z^3\ x2$	4.36E-05
$\phi\ \psi^2\ x2$	-4.70E-06	$\psi^3\ x2$	6.54E-06	$Z^3\ x3$	5.89E-05
$\phi\ \psi^2\ x3$	-3.81E-06	$\psi^3\ x3$	7.41E-07	$Z^3\ x4$	-3.83E-06
$\phi\ \psi^2\ x4$	-5.90E-06	$\psi^3\ x4$	5.13E-07	$Z^2\ \alpha^2$	3.80E-07
$\phi\ \psi\ Z^2$	4.73E-06	$\psi^2\ Z^2$	-5.26E-06	$Z^2\ \alpha\ x1$	-5.74E-06
$\phi\ \psi\ Z\ \alpha$	-5.76E-07	$\psi^2\ Z\ \alpha$	5.72E-07	$Z^2\ \alpha\ x2$	1.30E-06
$\phi\ \psi\ Z\ x1$	9.48E-06	$\psi^2\ Z\ x1$	-7.57E-07	$Z^2\ \alpha\ x3$	5.52E-07
$\phi\ \psi\ Z\ x2$	-2.18E-06	$\psi^2\ Z\ x2$	6.09E-07	$Z^2\ \alpha\ x4$	-1.56E-06
$\phi\ \psi\ Z\ x3$	-9.52E-06	$\psi^2\ Z\ x3$	5.82E-06	$Z^2\ x1^2$	6.07E-05
$\phi\ \psi\ Z\ x4$	-8.42E-07	$\psi^2\ Z\ x4$	-4.12E-06	$Z^2\ x2^2$	-5.26E-05
$\phi\ \psi\ \alpha^2$	6.36E-06	$\psi^2\ \alpha^2$	-2.33E-06	$Z^2\ x3^2$	-4.73E-05
$\phi\ \psi\ \alpha\ x1$	-1.52E-05	$\psi^2\ \alpha\ x1$	5.74E-06	$Z^2\ x4^2$	3.92E-05
$\phi\ \psi\ \alpha\ x2$	3.79E-06	$\psi^2\ \alpha\ x2$	-7.97E-08	$Z\ \alpha^3$	-7.15E-06
$\phi\ \psi\ \alpha\ x3$	1.78E-05	$\psi^2\ \alpha\ x3$	-9.99E-06	$Z\ \alpha^2\ x1$	4.80E-05
$\phi\ \psi\ \alpha\ x4$	-6.50E-06	$\psi^2\ \alpha\ x4$	3.79E-06	$Z\ \alpha^2\ x2$	-3.26E-05
$\phi\ \psi\ x1^2$	-6.43E-06	$\psi^2\ x1^2$	7.29E-07	$Z\ \alpha^2\ x3$	-8.08E-06
$\phi\ \psi\ x2^2$	6.74E-06	$\psi^2\ x2^2$	-2.24E-06	$Z\ \alpha^2\ x4$	1.39E-05
$\phi\ \psi\ x3^2$	1.04E-05	$\psi^2\ x3^2$	-4.77E-06	$Z\ \alpha\ x1^2$	7.27E-06
$\phi\ \psi\ x4^2$	-5.94E-06	$\psi^2\ x4^2$	1.02E-06		
$\phi\ Z^3$	-5.99E-06	$\psi\ Z^3$	2.56E-06		
$\phi\ Z^2\ \alpha$	1.77E-06	$\psi\ Z^2\ \alpha$	-9.73E-07		
$\phi\ Z^2\ x1$	-5.65E-06	$\psi\ Z^2\ x1$	2.54E-05		
$\phi\ Z^2\ x2$	8.85E-06	$\psi\ Z^2\ x2$	-6.97E-06		
$\phi\ Z^2\ x3$	6.87E-06	$\psi\ Z^2\ x3$	-4.47E-06		
$\phi\ Z^2\ x4$	3.54E-06	$\psi\ Z^2\ x4$	1.17E-05		
$\phi\ Z\ \alpha^2$	-2.39E-06	$\psi\ Z\ \alpha^2$	1.04E-05		
$\phi\ Z\ \alpha\ x1$	-1.22E-06	$\psi\ Z\ \alpha\ x1$	-6.06E-06		
$\phi\ Z\ \alpha\ x2$	-2.16E-06	$\psi\ Z\ \alpha\ x2$	3.05E-06		
$\phi\ Z\ \alpha\ x3$	-2.35E-06	$\psi\ Z\ \alpha\ x3$	7.08E-06		
$\phi\ Z\ \alpha\ x4$	9.18E-06	$\psi\ Z\ \alpha\ x4$	-1.16E-05		
$\phi\ Z\ x1^2$	-8.30E-06	$\psi\ Z\ x1^2$	-7.78E-06		
$\phi\ Z\ x2^2$	8.44E-06	$\psi\ Z\ x2^2$	4.75E-07		
$\phi\ Z\ x3^2$	8.12E-07	$\psi\ Z\ x3^2$	7.57E-06		
$\phi\ Z\ x4^2$	-6.93E-06	$\psi\ Z\ x4^2$	2.31E-06		
$\phi\ \alpha^3$	2.66E-06	$\psi\ \alpha^3$	-1.65E-06		
$\phi\ \alpha^2\ x1$	1.19E-05	$\psi\ \alpha^2\ x1$	-2.31E-05		
$\phi\ \alpha^2\ x2$	-1.40E-06	$\psi\ \alpha^2\ x2$	9.01E-06		
$\phi\ \alpha^2\ x3$	-5.76E-07	$\psi\ \alpha^2\ x3$	5.17E-06		
$\phi\ \alpha^2\ x4$	-8.30E-07	$\psi\ \alpha^2\ x4$	-3.44E-06		
$\phi\ \alpha\ x1^2$	-1.07E-06	$\psi\ \alpha\ x1^2$	4.26E-06		
$\phi\ \alpha\ x2^2$	5.74E-07	$\psi\ \alpha\ x2^2$	-7.68E-07		
$\phi\ \alpha\ x3^2$	7.82E-06	$\psi\ \alpha\ x3^2$	-9.84E-06		
$\phi\ \alpha\ x4^2$	-5.55E-06	$\psi\ \alpha\ x4^2$	5.36E-06		
$\phi\ x1^3$	-5.66E-06	$\psi\ x1^3$	2.54E-05		
$\phi\ x2^3$	8.86E-06	$\psi\ x2^3$	-6.97E-06		
$\phi\ x3^3$	6.87E-06	$\psi\ x3^3$	-4.46E-06		
$\phi\ x4^3$	3.54E-06	$\psi\ x4^3$	1.17E-05		
ψ^4	1.84E-07	Z^4	-5.33E-08		