UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Ingegneria Industriale DII

Corso di Laurea Magistrale in Ingegneria Aerospaziale

Surrogate Modelling for Short Intake Design

Relatore: Chiar.mo Prof. Ernesto Benini

Correlatore: Prof. David MacManus

Federico Cipriani

Matr. 1084256

Anno Accademico 2016/2017

ABSTRACT

The importance of decreasing the fuel consumption of aircrafts from generation to generation, along with the need to reduce emissions and noise, is leading to engines with higher By-Pass Ratio (BPR). This generally leads to an increase of weight and drag, which can outweigh the earned benefits. One approach which allows to reduce these effects implies the reduction of the length of the engine intake. Nevertheless, this can reduce the internal diffusion capability and lead to flow non-uniformities at the engine fan face, leading to reduced performance of the whole engine.

The aim of the current research project consists in the construction of a reliable surrogate model which reflects the behaviour of the intake bottom line when exposed to incidence conditions. This model represents an approximation of the real behaviour, but it provides a tool for the quick evaluation of bottom line designs. The initial objective of the project consisted in the determination of the bounds of the design space, which has to be populated for modelling purposes. Preliminary to any operation, a thorough analysis of the obtained design space has to be carried out for two main reasons: identify region of poor interest which have been included in the design space; identify geometries characterised by undesired features.

The surrogate modelling technique chosen is the Kriging method. Its application to the design space shows reasonable results, which, at this stage, cannot be considered acceptable for the final intended use. The best result achieved required an increase of dimensionality, moving from three to four variables, but leading to promising results demonstrated by, as indicator of the overall quality of the model, a Root Mean Square Error equal to 0.036, considering values of $DC_{60} \leq 0.1$ as aerodynamically acceptable values. The method has shown good potentialities, but more complex mathematical solutions have to be introduced to improve the model.

Keywords: 3D CFD, distortion, Kriging, metamodel, reduced order model

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Prof David MacManus, for giving me one of the greatest occasions of my academic carrier, in which I had the opportunity to grow both from a personal and professional point of view.

I would like to thank each member of the team at the Cranfield University, in particular Robert Christie, Alex Heidebrecht and Ioannis Goulos. Thanks for your patience, for your constant help and for the many things you taught me and which shaped my engineering approach to problems and to their resolution.

Thanks to all the people I met during this year at Cranfield, with whom I shared one of the most amazing time frames of this life. Without you the campus would not be the same.

Last but absolutely not least, thanks to the people who allow me to chase my dreams and who trust me in all the choices I make. You constantly give me the solid support I need in every tough decision I take. Thanks to my parents, and to my brother and sister, even if a thanks would never be enough.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

NOMENCLATURE

Roman symbols

$d_k^{(ij)}$	Cross-distance between samples i and j along
DC_{60}	Distortion coefficient
f_{IL}	Factor which control the radius of curvature at the highlight
$f_{\kappa_{FAN}}$	Ratio between radius of curvature at the fan face and fan diameter
k	Kinetic energy residual
K	Coefficient dependent from engine type and flight speed
L/D	Intake non-dimensional length (length intake/fan diameter)
L_{LIP}	Length of the intake lip
M	Mach number
P_f	Total pressure at the fan face
p_∞	Free-stream static pressure
P_∞	Free-stream total pressure
r	Pearson 'r' coefficient
r	Radius of curvature
r^2	Coefficient of determination
r_{IL}	Radius of curvature at the highlight
r_{TH}	Radius of curvature at the throat
R	Radial position of a point on the intake profile
R_{HI}	Radius of the intake cross-section at the highlight
R_{TH}	Radius of the intake cross-section at the throat
V	Vandermonde matrix
x	Location of a sample in the design space
y	Observation at location x
$\hat{y}$	Observation predicted at location x

Abbreviations

ACARE	Advisory Council for Aeronautical Research in Europe
AR	Aspect Ratio
BLUE	Best Linear Unbiased Estimator
BPR	By-Pass Ratio
CFD	Computational Fluid Dynamics
COBYLA	Constrained Optimisation BY Linear Approximation
CR	Contraction Ratio
CST	Class Shape Transformation
DoE	Design of Experiment
FF	Full Factorial
FPR	Fan Pressure Ratio
GA	Genetic Algorithms
GPML	Gaussian Processes for Machine Learning
HPC	High Computational Power
iCST	Intuitive Class Shape Transformation
IPR	Intake Pressure Recovery
LHS	Latin Hypercube Sampling
LOO	Leave-One-Out
MDO	Multi-Disciplinary Optimisation
MFCR	Mass Flow Capture Ratio
MLE	Maximum Likelihood Estimation
MSE	Mean Squared Error
OLS	Ordinary Least Square
PC	Parallel Coordinates
RANS	Reynolds-Averaged Navier-Stokes
RMSE	Root Mean Squared Error
ROCOR	Rate Of Change Of Curvature
RoI	Range of Interest ($DC_{60} < 0.1$)

RSM	Response Surface Methodology
SFC	Specific Fuel Consumption
SST	Shear Stress Transport
UDACP	User-Defined Auto-Correlation Parameters

Greek symbols

Δ	Variation
κ	Curvature
η	Intake diffusion efficiency
ω	Specific turbulent dissipation rate
θ	Auto-correlation parameter
θ_0	Initial auto-correlation parameter for MLE
θ_L	Lower limit of the auto-correlation parameter (thetaL)
θ_U	Upper limit of the auto-correlation parameter (thetaU)
Γ	Tikhonov matrix

Superscripts

ij	Pair of sample *i* and sample *j*
n	Number of samples in the design space
p	Number of features or dimensions

Subscripts

DIFF	Intake diffuser
FAN	Fan face
IL	Initial lip of the intake
LIP	Intake lip
max	Maximum diameter of the engine nacelle
nd	Non-dimensional value

TE	Trailing Edge
TH	Intake Throat
∞	Free-stream condition

INTRODUCTION

1.1 Context of the project and motivation

The main objective that is driving the innovation in the aeronautical sector is represented by the continuous research aimed to reduce the environmental impact of the next generation of aero-engines. This requirement is strictly related to the continuous monotonic increase of passengers and flights year after year and this requires the introduction of some forms of regulations to counteract the bad effects of flight travel on the environment. Between the different variables which play a role in this, those of major concern are generally identified in the emissions of CO_2 and NO_x and in the external noise produced by an engine. Common short and long-term goals have been set in a Strategic Research Agenda by the Advisory Council for Aeronautical Research in Europe (ACARE). The short-term targets are set for 2020, and between them the different companies in the sector have to ensure:

- 50% overall reduction of CO_2, which 20% is from the engine
- 80% overall reduction of NO_x, which 60% is from engine technology
- 50% overall reduction of noise

As an example of how major manufacturing companies are dealing with the defined objectives, in Figure 1.1 the trends which Rolls Royce's aero-engines have drawn since the imposition of the guidelines for 2020 are presented [1].

The design objectives to meet the requirements of reduced Specific Fuel Consumption (SFC), emissions and noise generally involve the design of aero-engines characterised by an increased By-Pass Ratio (BPR). This usually positively affects the performance of the engine, allowing to achieve a reduced Fan Pressure Ratio (FPR) and an improved propulsive efficiency [2], which is also related to an improved fuel burn [3]. An increase of BPR can be achieved in two ways: increasing the size of the whole engine, maintaining the dimensions of the core

engine; maintaining the dimensions of the whole engine and reducing the size of the core engine. The latter solution is obviously possible, but it requires levels of technology that are difficult to achieve in the short-term. Therefore, the former solution is more suitable for this kind of terms, but it brings few drawbacks that an engine manufacturer has to face. Between these, the major ones are the increased weight and size of the whole engine, especially when the section orthogonal to the direction of flight is considered. These two drawbacks can offset the benefits gained with an increased BPR generally leading to adverse weight effects, increased drag and SFC.

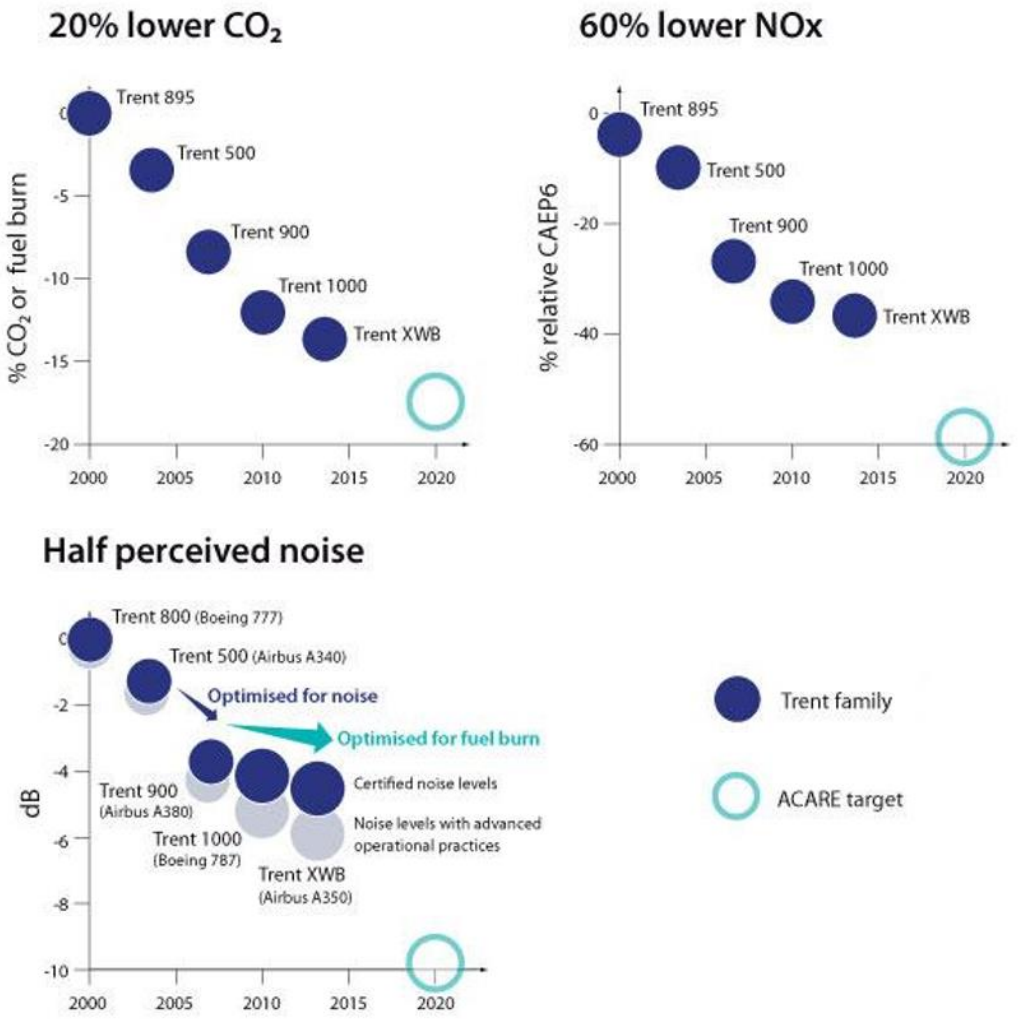


Figure 1.1: Future aero-engines - reducing the environmental impact [1]

In order to reduce the effects of these drawbacks, one solution considered by engine manufacturers is the design of shorter intakes, which technically allows

reducing weight and drag-related issues. Nevertheless, usually these benefits are obtained at the price of a reduced diffusion capability in the intake duct. This translates in lower Intake Pressure Recovery (IPR) and higher flow non-uniformities at the fan face, which adversely influence the overall efficiency and performance of the engine. A short intake design should provide optimal performance in all the phases of flight that an aircraft faces, almost at the same level of performance provided by the conventional length intake. Conditions such as take-off and climb pose at the intake designer a difficult problem to solve, since the achievement of optimal performance for all the conditions require a great number of compromises. Currently, the design of short intakes is carried out exploring different solutions manually, leading to the definition of non-optimal solutions. Therefore, in order to enhance the design process, the development of a procedure to quickly evaluate and generate optimal geometries for short intakes is mandatory. The process should automatically design and analyse different geometries, created on the base of the conflicting requirements that should be respected in all the conditions of flight.

1.2 Aim and objectives

The aim of the current research project is to create a surrogate model which provides a quick way to obtain optimal solutions for the design of the bottom line of a short intake. The objectives can be listed as following:

- Exploration of the design space of the bottom line geometry at incidence, aiming to improve the design space bounds derived from the past approach.
- Creation of a Latin Hypercube Sampling (LHS) DoE able to explore thoroughly the whole design space defined from the previous point.
- Exploration of the LHS DoE created, aiming to enhance its bounds and to filter undesired characteristics generated within it.
- Exploitation of surrogate modelling techniques for the creation of a metamodel able to generalise the behaviour of two metrics of interest.

LITERATURE REVIEW

2.1 Intake aerodynamics

The air intake system is the first component of a jet engine and it can be considered as a fluid duct which main task is to bring the free-stream air into the engine. In this process it has the fundamental objective to control the incoming air flow, ensuring the correct overall performance of the aircraft over a wide range of conditions [4]. The intake is designed to provide the appropriate amount of air to the compressor in order to achieve the level of thrust required in each phase of a flight. For subsonic podded engines, the incoming flow must reach the fan face with a moderate subsonic speed, and this imply that a method to reduce the speed of the free-stream has to be implemented. Considering as reference cruise flight conditions, the average aircraft speed is around $M = 0.85$. The maximum tolerable speed at the fan face for an engine as the one represented in Figure 2.1 is around $M = 0.5$ [5]. If this value is exceeded, this can lead to problems related mainly to compressibility effects. In any case, it is paramount that the required deceleration is carried out with the minimum amount of losses, ensuring the stability and the quality of the flow at the fan face [6].

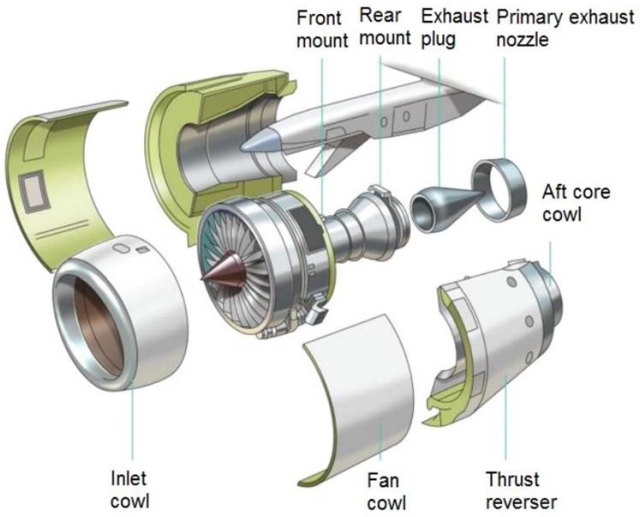


Figure 2.1: Typical nacelle elements of an engine [7]

2.1.1 Intake theory

Even though the air intake system can be considered as a simple aerodynamic duct, its design is a complex engineering issue, which requires the consideration of many compromises in terms of design and performance. As briefly stated above, it has to ensure the proper performance of the engine throughout all the operating conditions, from full thrust, when the aircraft is on the runway, to cruise, when the aircraft is in steady flight conditions. Considering for instance the former one, flow initially at rest has to be accelerated to the velocity required by the compressor. The typical shape of the stream-lines which characterise this condition is shown in Figure 2.2.

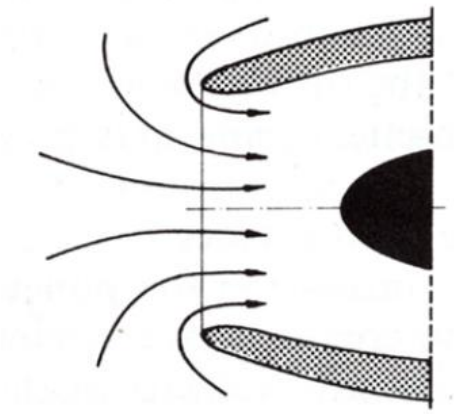


Figure 2.2: Flow field at static conditions [6]

As it can be figured out from the figure above, at full thrust conditions also air behind the intake lip is sucked inside the duct. In this situation the shape of the lip has an important impact on the behaviour of the flow. If it is too sharp, and so characterised by a strong curvature near the leading edge of the intake, separation inside the duct can be encountered due to the strong acceleration. In order to avoid this issue, the intake designer should solve the problem designing a thicker lip, characterised by a well-rounded geometry, which allows a gentle acceleration of the flow around the leading edge without encountering separation. Nevertheless, as it will be explained thoroughly later, even though this design is optimal for the condition considered, it is not desirable when the intake works in cruise

conditions. This is because the shape of an ideal stream-tube which drives the air stream in the engine is different in the two phases considered. As it will be clearer later, this also implies that the flow which enters the intake is characterised by a different behaviour in the region near the intake lip. This brief introduction gives an example of the difficulties encountered in the design of an intake system, highlighting the fact that different compromises must be found in order to find the optimal design which ensure correct performance in all the phases of flight.

In Figure 2.3 the basic concept of aerodynamic duct is reported, and it serves as an extreme simplification of the more complex aircraft engine. The intake spans from *section c* to *section f* and, as it can be inferred, it is mainly characterised by a diffuser shape. As stated by Seddon [8], the intake can be considered as a form of compressor since a pre-compression of the air stream occurs when this flows between the two considered sections. Along with the diffusion, the incoming flow is slowed down from the free-stream velocity to an acceptable value of velocity at the fan face. This value is determined and fixed by the engine manufacturer, and usually is set below $M_{FAN} < 0.6$. It allows to avoid mechanical and aerodynamic problems related to the coupling of the high speed at the fan blades tip and compressibility-related effects [5].

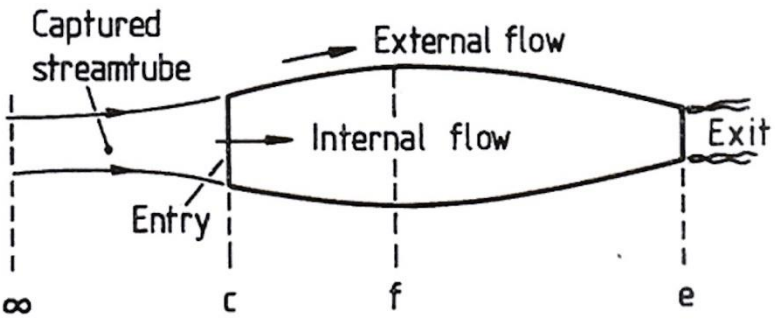


Figure 2.3: Aerodynamic duct in an airstream [8]

An important concept related to the aerodynamics of a fluid flow duct regards the concept of stream-tube. In particular, concerning the air intake system, the main interest regards the so-called pre-entry stream tube. Its shape varies from one flight phase to another, and it is mainly dependent on the air flow demand of the

engine at the different flight phases. An important parameter associated with the stream-tube concept is called mass flow capture ratio (MFCR). It is defined as the ratio between the area of the stream-tube at infinity and the designed area at the highlight of the intake, as expressed in Equation (2.1).

$$\mathrm{MFCR} = \frac{A_{\infty}}{A_{HI}} \tag{2.1}$$

Generally, associated with different values of this parameter, it is possible to identify four principal phases during a flight: take-off, climb, normal cruise and maximum cruise [5]. Depending from the value of MFCR, a representation of the different shapes that the pre-entry stream-tube can assume is given in Figure 2.4.

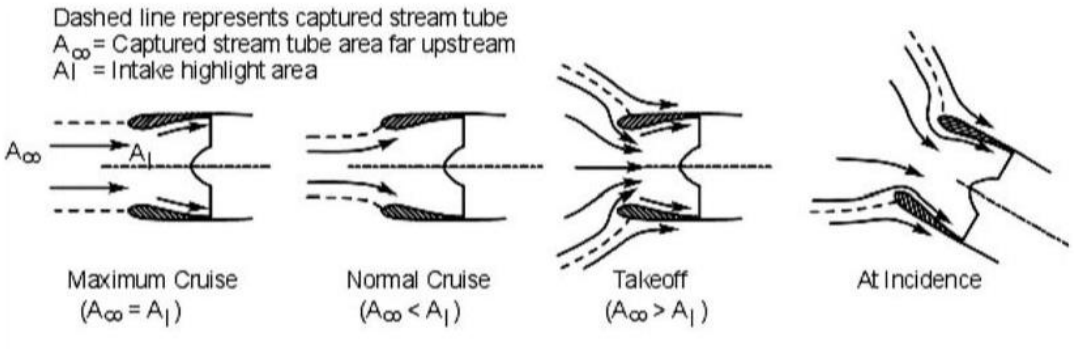


Figure 2.4: Mass flow demand in different flight phases [5]

For a MFCR greater than 1, typical range for low speed flight, the stagnation point is located on the nacelle external surface, as shown in Figure 2.5. Considering the airstream located near the boundaries of the stream-tube, when this reaches the region near the stagnation point it is forced to turn and enter the intake duct. Potential flow calculations yield that this lead to the development of very high velocities near the highlight point [8]. This becomes clear visually examining Figure 2.5, where the distribution of Mach number in different locations along the lip is reported. Considering hypothesis of inviscid flow, the Mach number, approaching the highlight point from the stagnation point, significantly increases until a shock wave is formed immediately after the leading edge, on the intake lip side. This occurs because the real flow is not able to deal with high adverse

pressure gradients related to high supersonic peak velocities [8]. The consequences caused by the shock wave comprise a rapid deceleration of the flow followed by the flow separation from the intake surface. If this separation is not properly controlled through, for instance, the shape of the intake duct, this could persist until the fan face, affecting the performance of the engine. In addition, in some of the cases in which the shape of the intake is not properly designed, boundary layer thickening can occur, leading to high losses in the duct.

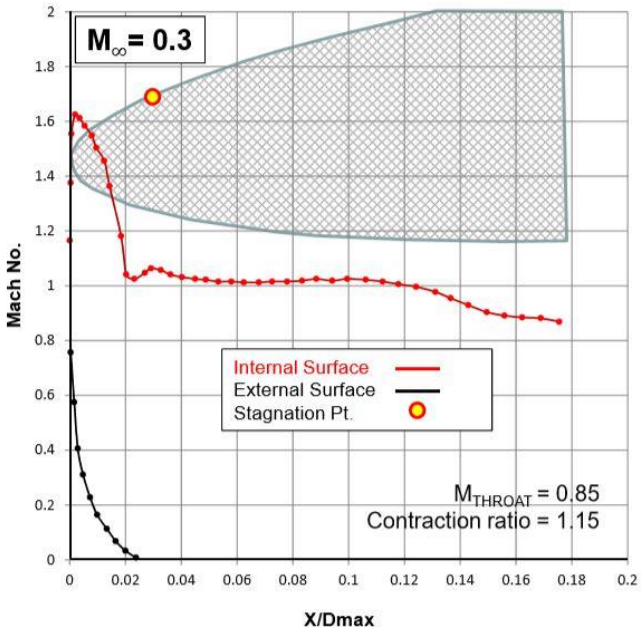


Figure 2.5: Mach number distribution on intake lip for MFCR > 1 [4]

Increasing the flight speed leads to the movement of the stagnation point towards the highlight point until the condition of maximum cruise is met, when the shape of the stream-tube can be assumed almost cylindrical. In cruise flight, characterised by a MFCR lower than 1, the stagnation point moves inside the intake, in a position generally located on the lip, downstream the highlight point, as shown in Figure 2.6. Analysing the Mach number distribution on the internal surface of the intake, it is clear that the phenomena which occurs in the previous considered condition are not present. The airstream velocity increases until the throat point due to the convergent profile of the lip, without the formation of shocks. In this situation the losses which can occur internally are mainly related to friction issues. In any case,

the aerodynamic problem which requires attention in this situation is related to the behaviour of the flow on the external surface of the intake. In fact, considering for example an intake characterised by a thick shape, the problem of the acceleration of the flow around the lip is now moved on the external side of the cowl. In this situation the main issue that a designer has to solve consists in the reduction of the cowl drag on the nacelle, caused by potential shocks related to the high acceleration.

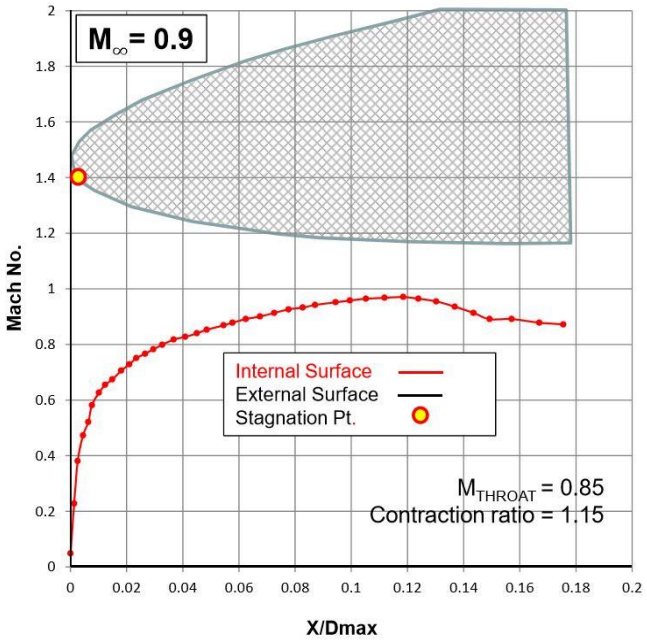


Figure 2.6: Mach number distribution on intake lip for MFCR < 1 [4]

Considering the intake throat, another important point that becomes paramount in the design of an intake is represented by the definition of the throat Mach number. Since the condition considered in the design phase is the maximum cruise one, the Mach number at the throat section in this situation should be maximised in order to achieve an acceptable value of Mach number at the fan face. Obert [9] reports that in order to prevent strong shockwaves on the intake walls the average Mach number at the throat should be below $M_{TH} < 0.8$.

In Figure 2.7 it is reported a study carried out by Seddon [8] on the influence of M_{TH} and MFCR on the lip losses encountered in an intake with a well-rounded lip and a fixed geometry, defined principally by the contraction ratio.

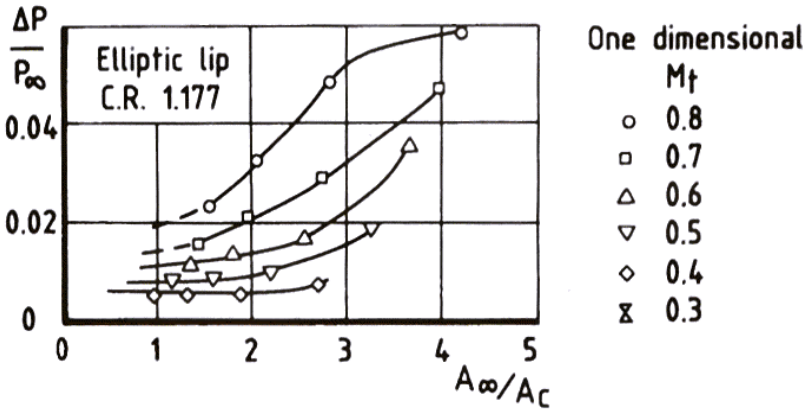


Figure 2.7: Influence of M_{TH} on total pressure loss [8]

The specification of the term "lip losses" is useful to distinguish the losses due to the shape of the lip, dependent also from the ratio between the highlight radius and the throat radius, defined as Contraction Ratio (CR), from those related to the phenomena which occur in the diffuser, often associated to skin friction. In intake design theory, lip losses are usually expressed as in Equation (2.2).

$$\left(\frac{\Delta P}{P_\infty}\right)_{LIP} = \frac{\Delta P_f}{P_\infty} - \left(\frac{\Delta P}{P_\infty}\right)_{DIFF} \tag{2.2}$$

2.1.2 Intake performance

In order to proceed to the discussion regarding the design of an intake system, it is useful to introduce two figures-of-merit with which the performance of this component is usually evaluated. For the purpose of the analysis carried out, the two metrics which have been analysed are the Intake Pressure Recovery (IPR) and the distortion coefficient DC_{60}, thoroughly described in the following two paragraphs. In this context they are also considered as aerodynamic performance metrics.

2.1.2.1 Intake Pressure Recovery

The value of the static pressure at the fan face is higher compared to the value in free-stream conditions. From a thermodynamic point of view, due to the high velocity of the incoming flow, the diffusion process which takes place across the

intake can be considered isentropic and adiabatic, leading to the preliminary assumption of constant total temperature. Nevertheless, the same conclusion cannot be considered valid for the total pressure, as shown in Figure 2.8. In fact, it is generally assumed that in the intake duct is difficult to limit the different phenomena or problems which could occur, such as for example the already mentioned separation phenomena and friction problems. The efficiency of an intake system is mainly related to the particular phase of flight in which it works and, for each one of these, various factors play important roles in determining the quality of the flow. For instance, as introduced in the previous paragraph, the shape of the lip determines the capability of the intake to deal with the incoming air stream and ensure that separation does not occur.

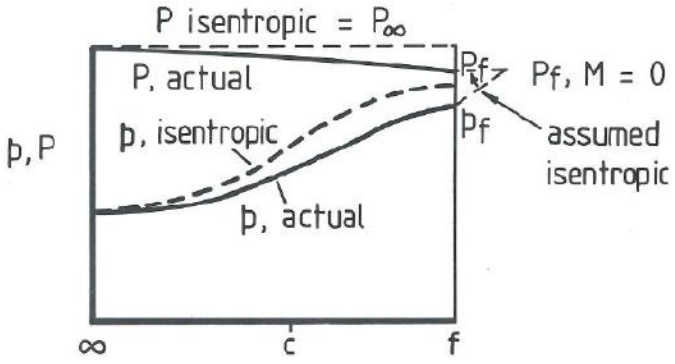


Figure 2.8: Process of intake pressure recovery [8]

The definition of efficiency of the diffusion process in the intake is given by Seddon [8], and it is expresses as in Equation (2.3). This expression is valid for compressible flow, but it is not largely used and other forms are usually preferred.

$$\eta = \frac{Work\ done\ in\ compression}{Kinetic\ energy\ available} = \frac{(\frac{P_f}{p_\infty})^{\frac{\gamma}{\gamma-1}}}{\frac{\gamma-1}{2}M_\infty^2} \qquad \textbf{(2.3)}$$

A more common and convenient way to define the efficiency of the diffusion process is presented in Equation (2.4). This is valid for high free-stream velocities, as well as for supersonic flows. Moreover, when intake systems are considered, this efficiency is well-known under the name of Intake Pressure Recovery (IPR).

$$\eta = \frac{P_f}{P_\infty} \tag{2.4}$$

Its effect on engine performance, especially in terms of thrust, is assumed to be directly translatable, as can be inferred in Equation (2.5) [8]. K is a factor which depends on the type of engine and on the flight speed.

$$\frac{\Delta\, Thrust}{Thrust} = \mathrm{K}\frac{\Delta P}{P_\infty} \tag{2.5}$$

IPR has therefore a determining effect upon the engine thrust level. As defined by Seddon [8], this is the resultant force which acts in the direction of cruise and which is produced on the aerodynamic duct by the internal flow. On the opposite side, intake drag is defined as the resultant force which acts in the opposite direction to the thrust and which results from the interaction between the external flow and the aerodynamic duct. Assuming steady flight conditions, it can be easily inferred that a reduction in total pressure in the internal intake duct, leading to a reduction of engine thrust, affects also the balance between thrust and the drag, requiring more fuel to counteract this outcome.

The factors which cause the reduction of total pressure in the intake duct depend on the operating condition considered [8]. As it has already been reported, for low speed flight and climb, characterised by MFCR > 1, the main factors which induce losses in the duct are shock waves, flow separation and the turbulent mixing related to this. In turn, for high speed steady flight, the main cause is principally the friction on the intake walls.

2.1.2.2 Distortion coefficient

This figure-of-merit is related with one of the main objectives that an intake system has to achieve: uniformity and stability of the flow at the fan face. The term "distortion" is used to identify the spatial non-uniformity of total pressure across the fan face. In the intake performance literature, the distortion coefficient is

usually identified as DC_θ, where θ indicates the angle of one of the sectors in which the fan face is subdivided. The distortion coefficient considered in the current analysis is identified as DC_{60}, since the total pressure distribution is evaluated on different sectors of 60 degrees on the fan face, as shown in Figure 2.9. This value for the angle is usually considered as a satisfactory minimum to obtain a good understanding of the level of distortion [8]. The distortion coefficient DC_{60}can be expressed in different forms depending on the country in which the engine is constructed, and the form more commonly used in the United Kingdom is expressed as in Equation (2.6). Generally, this parameter is given by the engine manufacturer to the engine designer as a constraint.

$$DC_{60} = \frac{\bar{P}_f - P_{60}}{q_f} \tag{2.6}$$

$\bar{P}_f$ is the area averaged total pressure at the fan face, P_{60} is the area averaged pressure in the sector with the lowest value of total pressure and q_f is the mean dynamic head corresponding to the whole fan face.

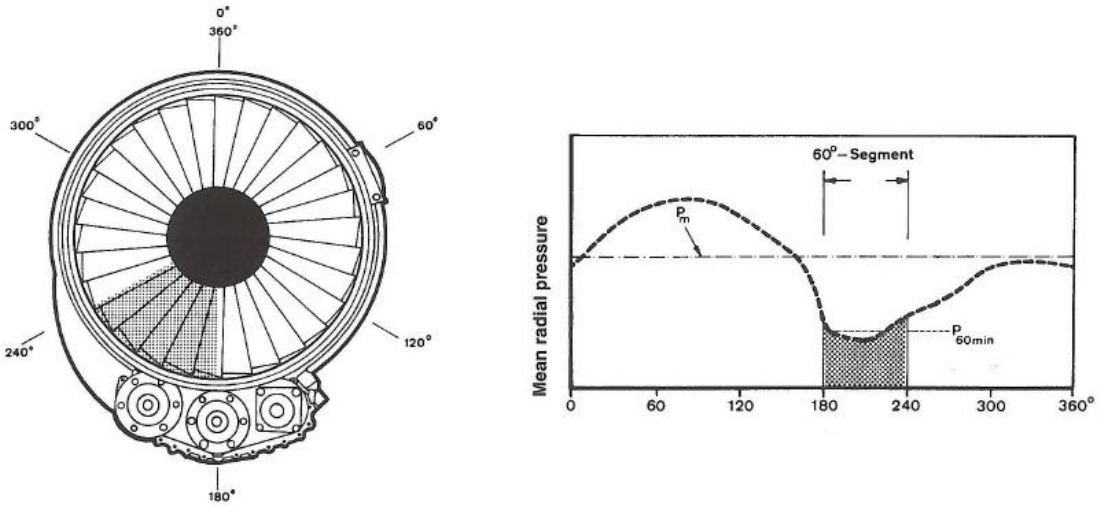


Figure 2.9: Distortion coefficient [6]

Between the different problems related to high values of distortion, the one of major concern is related to the surge of the compressor. The high rotational speed of this engine component, along with a non-uniform pressure distribution across

the fan face, can lead to failure due to excessive vibrational motion of the compressor blades.

The factors which lead to distortion at the fan face are kind of related to those which case a loss in total pressure. Nevertheless, by definition, the derivation of the distortion coefficient takes into consideration only pressure terms calculated at the fan face, and it evaluates the differences of total pressure between different sectors on this "surface". This means that if the air stream is able to re-attach following a weak shock-induced separation, the measured DC_{60} is low because there are not non-uniformities at the fan face. Though, at the same time, there is a marked loss in IPR, since the shock further reduces the value of total pressure. This case can be considered valid if the angle of attach of the aircraft is almost zero. Nevertheless, when considering incidence conditions attention should be paid in the design of both lip and diffuser. In fact, even if the shock occurring on the lip is characterised by weak intensity, if the diffuser is not properly design the boundary layer can grow and cause diffusion-induced separation, which in turn lead to high value of distortion and related issues.

2.1.3 Intake design

Kundu[5] provided some basic guidelines for the design of an air intake system, and the procedure was principally based on the maximum cruise flight condition. The shape of the stream-tube can be considered almost cylindrical, and the capture area at infinity can be assumed equal to the area at the highlight of the intake, as shown in the first picture on the left in Figure 2.4. Nevertheless, as briefly demonstrated in the first paragraph, the design of an air intake system involves the consideration of many compromises between the ideal condition of maximum cruise and the multiple off-design points. Between the large number of trade-offs that have to be considered, for instance, there is the one between IPR and drag since, as stated in the previous paragraph, the former has a direct impact on the engine thrust and therefore on fuel consumption. The ideal outcome from the design process is characterised by an intake which avoid boundary layer separation in the inner duct in all phases of flight and which is described by a high

value of IPR and by a low value of DC_{60}. The designer has to solve the difficult task of respecting the contrasting requirements defined for each phase of flight. From a generic consideration discussed by Seddon [10] in *Practical intake aerodynamics design*, the compromises should be found between:

- A thin lip required at cruise, condition characterised by a high flight Mach number which defines a stream-tube shape as the one pictured on the top of Figure 2.10. This shape aids to avoid shocks and induced drag over the nacelle surface. In fact, the shape of the stream tube is such that the air stream which flows both externally and internally the aerodynamic duct does not undergoes rapid acceleration or separation. The only issue which affect this condition is not an aerodynamic issue, but it represented by the friction on the inner walls of the duct, which can generally lead to pressure losses.
- A thick lip required at take-off, where the shape of the stream tube can be visualised in the bottom of Figure 2.10. The blunt shape of the lip allows to the air stream on the borders of the stream tube to flow inside the intake accelerating in a controlled way around lip, avoiding the rapid acceleration which could take place with a sharp lip and which could end up with shocks and related pressure losses and distortion.

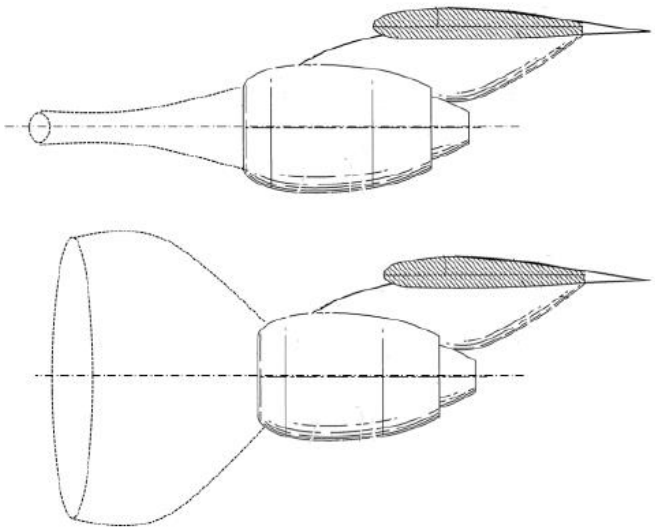


Figure 2.10: Stream-tube shape at cruise and take-off conditions [4]

In order to enhance the design process, it is useful to identify the two main sections of an intake system. This is formed by an initial convergent shape, called lip, which extends from the highlight point, the outer point of the engine casing, to the throat, the point of minimum area inside the intake duct. The second part is called diffuser and it is identified by a divergent geometry. The lip geometry can be defined by three main parameters which can be defined as in the following equations.

$$CR = \left(\frac{R_{HI}}{R_{TH}}\right)^2 \tag{2.7}$$

$$AR = \frac{L_{TH}}{R_{HI} - R_{TH}} \tag{2.8}$$

$$r_{IL} = f_{IL}\frac{(R_{HI} - R_{TH})^2}{L_{LIP}} \tag{2.9}$$

Moreover, in the design process the diffuser geometry is usually defined by parameters such as the curvature at the throat $\kappa_{TH}$and at the fan face κ_{FAN}. A particular point which is paramount to highlight at this stage is that a feasible diffuser shape is characterised by a single inflection point. This point becomes of crucial importance when the automatic creation of the different intake geometries will be considered. In fact, the potential generation of multiple inflections along the intake length is highly likely, and a solution to limit this problem has to be found. In Figure 2.11 it is reported a schematic representation of the different parameters which define the geometry of an intake, as specified by the parameterisation currently employed at the Cranfield University [11].

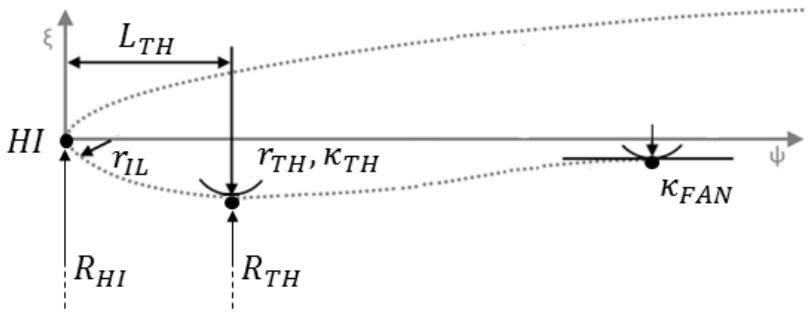


Figure 2.11: Intake design parameters [11]

When studies at incidence conditions are considered, as it will be demonstrated later, the CFD simulations require the use of a three-dimensional intake geometry in order to obtain feasible results which reflect the real aerodynamics effects which occur in the duct. Considering a real modern engine, in order to improve the performance at incidence, the cowl geometry is modified by two parameters [11] which remove the symmetry of the engine with respect to the mid-plane, parallel to the horizontal axis, as shown in Figure 2.12. The first parameter is the scarf angle, defined as the angle between the highlight plane, described by the inlet section of the intake, and the engine centre line. The second parameter is the droop, which in the context of this analysis is defined as an offset rather than an angle. It is represented by the vertical translation of the highlight centre and it is defined as the vertical distance between the highlight centre and the engine axis [12]. According to Obert [9] and considering momentarily the droop parameter as an angle, real aero-engines intakes are usually drooped of an angle comprised between 3 and 5 degrees.

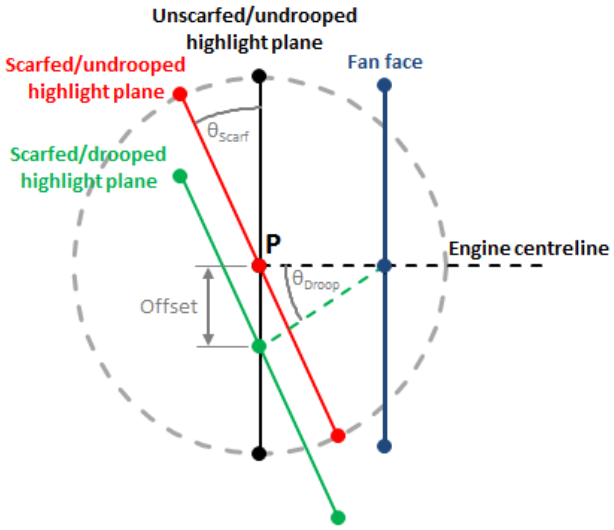


Figure 2.12: Scarf and droop [11]

2.1.4 Intake at incidence

The typical value of MFCR which characterise the ascent flight phase is usually greater than unity, as can be inferred from the shape of the stream-tube in the fourth picture of Figure 2.4. As during take-off, this is related to the low Mach

number and high power settings which characterise such flight condition. The stream-tube shape is no longer axisymmetric and the locus of the stagnation point is slightly moved on the nacelle surface. As shown in Figure 2.13, this is no longer located in one unique axial position.

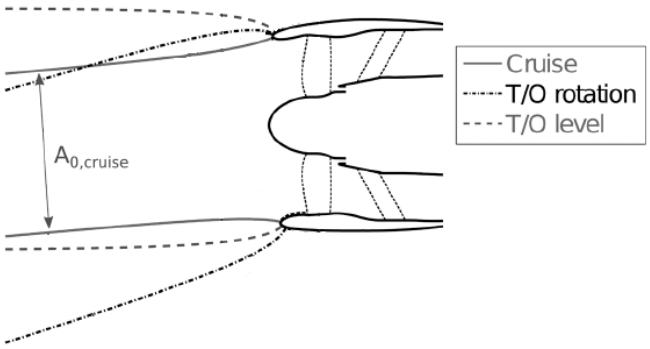


Figure 2.13: Stream-lines representation for cruise and ascent [13]

Considering the stagnation point located on the external surface of the bottom profile, at incidence this is moved downstream the highlight point. The air stream which flows around the bottom lip of the intake is therefore subjected to high acceleration due to the heavy stream line curvature. For the same reasons reported in Paragraph 2.1.1, this leads to a strong shock wave on the lip, as it is possible to see from the CFD results shown in Figure 2.14.

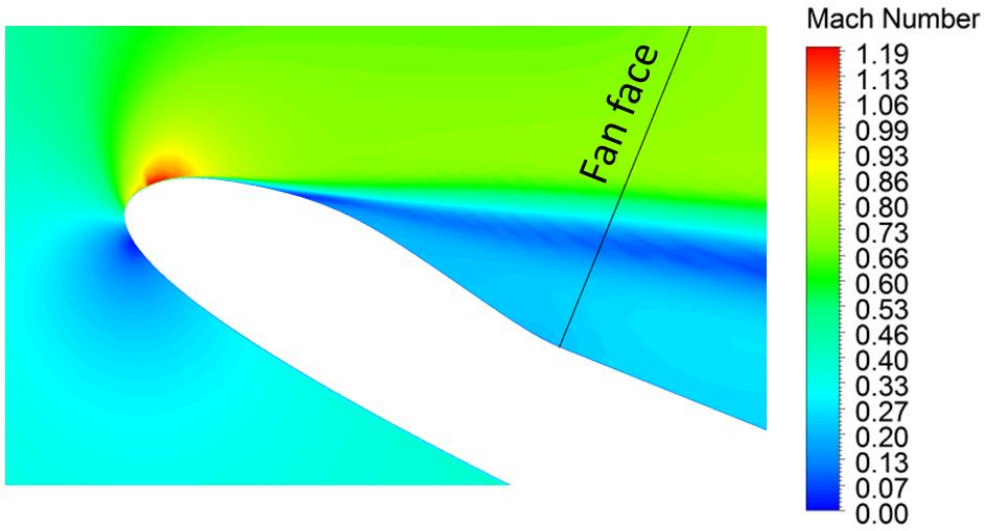


Figure 2.14: Mach number contours at high incidence [11]

Due to the high angle of the intake, if the bottom line geometry is not designed in order to delay or reducing the severity of the separation [8], the shock wave can prompt the formation of a growing region of adverse pressure gradient in the diffuser, which heavily affects the intake performance in terms of IPR and DC_{60}. In fact, separated flow is considered as the main cause of increased stagnation pressure loss and promotions of flow non-uniformity at the fan face, leading to the already reported problems of stability of the compressor and thus of the whole engine.

Seddon [8] identified that the geometric key driver which allows to control this phenomenon is represented by the contraction ratio, as demonstrated in the graph in Figure 2.15. As can be inferred, the effect of incrementing CR becomes more significant for high throat Mach numbers. Nevertheless, as it has already been demonstrated in Paragraph 2.1.1, high values of M_{TH} are directly related with an increase of lip losses.

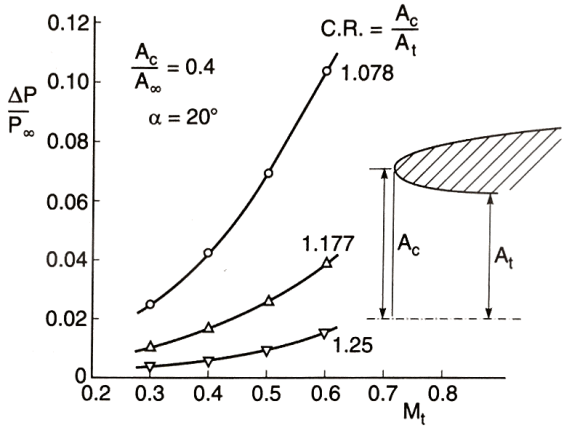


Figure 2.15: Effect of CR on intake pressure losses [8]

Another way useful to reduce the effects of incidence conditions is represented by the design of intakes with scarf and droop, as presented in the final discussion of Paragraph 2.1.3. In this way the incidence angle experienced by the engine is lower than the actual angle at which the aircraft flies. This leads to a reduced acceleration of the airstream around the intake bottom lip, which in turn reduces the possibility of separated flow inside the duct.

2.2 Design of Experiment

Designing an intake requires the study of the effects of a multitude of variables on the variation of the metrics of interest, as seen in paragraph 2.1.3. Each variable, interacting with one or more other variables, can influence the aerodynamic behaviour of the system in different ways. Even with the computational power available nowadays, it is not possible to explore all the possible combinations of potential designs and optimised strategies have to be established. The examination of each combination can require a total analysis time which tends to infinite and this is not an affordable time frame in terms of real world applications. Therefore, different techniques can be exploited to explore the design space of a problem of interest, obtaining feasible results even if the dimensionality of the problem is heavily reduced. As stated by Montgomery [14], the designer should carry out different experiments with two clear objectives: the determination of the most influential design variables in the problem; the identification of the bounds in which these variables influence the investigated output the most. Therefore, after the determination of the main design variables, specific bounds have to be found in order to start the design space creation. The outer surface of this bounded design space is known in spatial analysis as convex hull, defined as the smallest set which contains all the possible infinite design variables combinations in the Euclidian space. Montgomery uses the expression Design of Experiment, also known as DoE, to indicate the process of determination of the different experiments which have to be carried out in order to collect appropriate data and draw meaningful conclusions about the behaviour of the problem of interest. The set of experiments has to be determined in an intelligent way in order to explore thoroughly each part of the design space. In order to achieve this objective, a large number of different methods have been developed during the years. The selection of the methodology to exploit starts from the understanding of the available resources. Between them, the most important and crucial are represented by the computational power and time frame available. These two conditions determine the complexity of the DoE technique and the discretisation level at which the design space can be studied. Once all the experiments have been carried out, or all the simulations have been

run, it is possible to proceed to the next step of the procedure, which usually consist in the construction of surrogate model.

Two DoE techniques have been exploited in the context of this work, both for design space exploration and for design space creation. In particular, the techniques successfully employed are the Full Factorial (FF) and the Latin Hypercube Sampling (LHS). The theory behind them is explored in detail in the next two paragraphs.

2.2.1 Full Factorial

This method is one of the most common and intuitive strategies to explore the design space [15]. The number of experiments that are run with this technique is given by Equation (2.10).

$$N = \prod_{i=1}^{k} n_i \qquad i = 1, \dots, p \tag{2.10}$$

p represents the number of features or independent variables, or in alternative the number of dimensions of the problem, whereas n_i represents the number of levels for each variable. For example, considering the case in Figure 2.16b, the problem is defined by three independent variables, which determine the three dimensions, and for each variable the levels considered are two. Therefore, by definition, the total number of experiments is equal to $N = 2^3 = 8$.

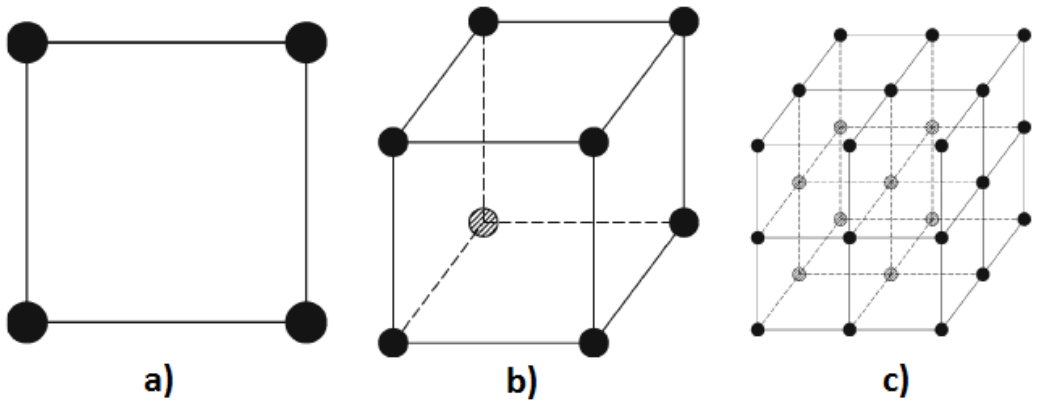


Figure 2.16: Full Factorial Designs examples [15]

The main advantage of this technique consists in the fact that it use efficiently the data without confusing the specific effects of the different parameters [15]. Since the data are distributed on a regular grid, Full Factorial (FF) allows to create the base for the creation of response surfaces based on polynomial interpolation of the data. Though, generally the quality associated to metamodels created in this way is poor and other interpolation techniques are preferred.

Although this method is easy to implement, its main downside is the exponential growth of experiments as the number of variables and levels is increased. For this reason, it is not suitable for high-dimensional design spaces or for design spaces where a high level of discretisation is required. Moreover, it is possible to remark upon the fact that the different points which identify the combinations overlap when they are projected on to the axes. Also, a common feature roughly belonging to all the approximation models that can be created is that they are more accurate nearby the point evaluated in the DoE process [16]. Thus, in relation to this point, a uniform spread of points across the design space which avoids this overlapping issue should be preferred. In fact, using a cheap DoE technique, such as the FF, is not always the best choice, because it usually translates in imprecise results and poor design space exploration [15]. The natural development of this concept can be found in the technique known as Latin Hypercube Sampling.

2.2.2 Latin Hypercube Sampling

Latin Hypercube Sampling (LHS) is one of the so-called Random DoE methods. Cavazzuti [15] suggests that this technique, as well as another technique called space-filling Sobol, over-perform other methods in the creation of a dataset for RSM purposes. Considering p dimensions and n design points, the design space is discretised into an orthogonal grid in which each axis is divided in n equal parts. The result is a multidimensional space subdivided into many equal sized hypercubes, called bins. From a conceptual point of view, the LHS is then populated placing in each bin a point, which represents a design variables combination which has to be simulated. As can be inferred from Figure 2.17, and assumed as one of the key requirements of the Latin Hypercube Sampling, each

sample must be unique in every direction parallel to the orthogonal coordinate system [16]. An optimal LHS is one which avoids the clustering of design points in certain locations, ensuring an optimal distribution of the samples across the design space. In this way, the behaviour of the metric examined can be studied thoroughly in all the regions of the design space. As an extreme example to give a better idea of this mentioned concept, also a simple distribution of samples along the diagonal of the design space satisfies the requirement imposed by this technique. Though, it clearly shows a strong correlation between the different dimensions and, in particular, most of the design space remains unexplored.

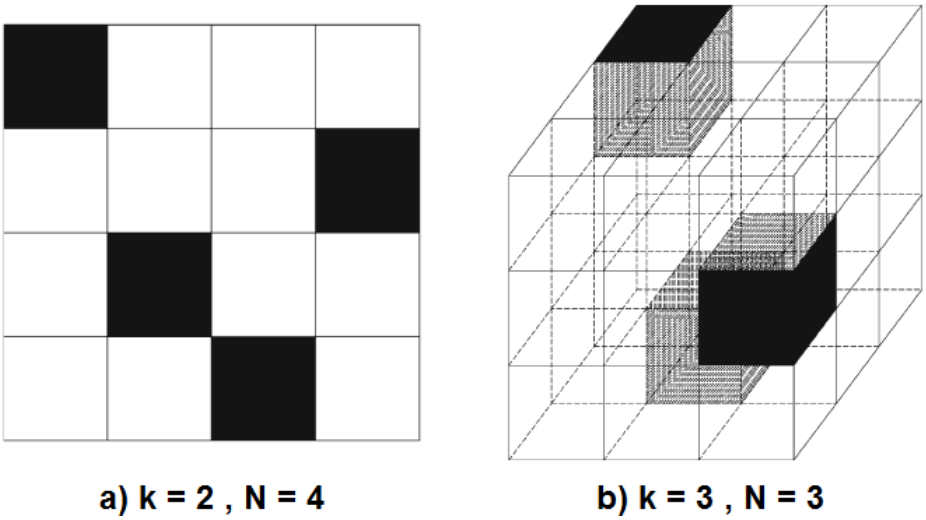


Figure 2.17: LHS designs examples [15]

This method does not suffer the collapse problem, because if one or more of the factors appear not to be important, every point in the design space still gives information about the influence of the other factors on the response.

One desired outcome required to the Latin Hypercube Sampling is to be space-filling, so the samples do not cluster in certain zones of the design space. In order to ensure the uniformity of the distribution of samples, conditions are set on the minimal distance between two consecutive design points, and the measure which is widely used to evaluate the uniformity of the sample is the maximin metric [16].

2.3 Surrogate modelling

Surrogate models, also known as metamodels, are often used after a DoE run to build a statistical model which aims to generalise the behaviour of the design space. Once the results from the DoE are collected, we have a multi-dimensional space characterised by different values of the response variable investigated. Nevertheless, all these different values are not interconnected and nothing can be said about the response variable in the locations of the design space outside the samples examined. The aim of surrogate modelling is to find the function $y = f(\boldsymbol{x})$ which represents the behaviour of the response in the design space [16]. The mapping of the function is based on the known values of the output variable in the locations specified by the DoE technique used. Another expression used as synonym for surrogate modelling is Response Surface Methodology (RSM). This name indicates that the relationship between the input and output variables can be represented as a hypersurface in the $\mathbb{R}^p$ dimensional space, identified by the p different features, or independent variables, that describes the problem.

2.3.1 Kriging

This method finds its origins in geostatistics, when Danie Krige, South African mining engineer, applied mathematical statistics to gold mining, in order to determine the location of gold reserves [16]. Kriging is a derivation of the Gaussian processes, and in particular it belongs to the family of the least squares algorithms, where it finds large use for processing highly non-linear responses. Given the set of input variables $\boldsymbol{X} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots, \boldsymbol{x}^{(n)}\}$ and the corresponding set of responses $\boldsymbol{Y} = \{f(\boldsymbol{x}^{(1)}), f(\boldsymbol{x}^{(2)}), \dots, f(\boldsymbol{x}^{(n)})\} = \{y^{(1)}, y^{(2)}, \dots, y^{(n)}\}$, where the term between brackets in the exponent stands for the *i-th* observation, the aim of the method is to obtain a response surface where is possible to find the value of the desired output for a generic $\boldsymbol{x}$. Kriging defines the response in that point as a weighted linear combination of the known response values, in the form

$$\hat{f}(\boldsymbol{x}) = \sum_{i=0}^{n} \lambda_i(\boldsymbol{x})\, y^{(i)} \qquad \textbf{(2.11)}$$

The term λ_i represents the weight given to each observation obtained through the DoE. The weights are obtained as solution of a system of linear equations which looks for the Best Linear Unbiased Estimator (BLUE) based on a stochastic model of spatial dependence quantified by a semivariogram γ, or by the average of the experimental responses μ, and by the covariance function [15]. The concept of variogram, represented as 2γ, is found in spatial statistics and it is described as in Equation (2.12), where E is the expectation and ν the expected value of $f(\boldsymbol{x^j})$. As briefly stated, the variogram represents the degree of spatial dependence of a stochastic process, and it is defined as the variance of the difference between the response values at two locations in the design space [17].

$$\begin{aligned} 2\gamma\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) &= \operatorname{var}\left(f\left(\boldsymbol{x}^{(i)}\right) - f\left(\boldsymbol{x}^{(j)}\right)\right) \\ &= \mathrm{E}\left[\left(f\left(\boldsymbol{x}^{(i)}\right) - \mu - f\left(\boldsymbol{x}^{(j)}\right) + \nu\right)^2\right] \end{aligned} \tag{2.12}$$

The average of the experimental responses is given by

$$\mu = \mathrm{E}[f(\boldsymbol{x})] = \sum_{i=1}^{N} \frac{f(\boldsymbol{x}^{(i)})}{N} \tag{2.13}$$

The covariance function is expressed as in Equation (2.14), and it is used to calculate the correlation between pair of variables.

$$\operatorname{cov}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) = \operatorname{cov}\left(f\left(\boldsymbol{x}^{(i)}\right), f\left(\boldsymbol{x}^{(j)}\right)\right) = \mathrm{E}\left[\left(f\left(\boldsymbol{x}^{(i)}\right) - \mu\right)\left(f\left(\boldsymbol{x}^{(j)}\right) - \nu\right)\right] \tag{2.14}$$

From Equation (2.12) and (2.14), assuming $\mu = \nu$, for any two points in the design space Equation (2.15) represents the correlation between the semivariogram and the covariance function.

$$\gamma\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) = \frac{1}{2}\operatorname{var}\left(f\left(\boldsymbol{x}^{(i)}\right)\right) + \frac{1}{2}\operatorname{var}\left(f\left(\boldsymbol{x}^{(j)}\right)\right) - \mathrm{c}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) \tag{2.15}$$

Many types of Kriging exist depending on how the average μ is computed [15]. Some of the most commonly employed types are:

- Simple Kriging $\mu(\boldsymbol{x}) = 0$

- Ordinary Kriging $\mu(\boldsymbol{x}) = \mu$
- Universal Kriging $\mu(\boldsymbol{x}) = \sum_{j=1}^{k} \beta_j x_{i,j}$

The Kriging variance $\hat{\sigma}^2$, called also Kriging error, is defined as

$$\hat{\sigma}^2 = \text{var}\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right) = \text{E}\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right) \tag{2.16}$$

In order to compute the weights λ_i the variance should be minimised assuming the unbiasedness condition in which it is assumed equal to zero. Avoiding to report the entire mathematical derivation behind this statement, it is found, for instance in case of ordinary Kriging, that the weights are given by

$$\sum_{i=1}^{n} \lambda_i(\boldsymbol{x}) = 1 \tag{2.17}$$

After some passages, from the mathematical extension of Equation (2.16) it is found that

$$\text{cov}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) \boldsymbol{\lambda}(\boldsymbol{x}) = \text{cov}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}\right) \tag{2.18}$$

and therefore the weights are given by Equation (2.19).

$$\boldsymbol{\lambda}(\boldsymbol{x}) = \text{cov}^{-1}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) \text{cov}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}\right) \tag{2.19}$$

The covariance terms are unknown and they have to be estimated through a semivariogram model [15].

In the current research project, the implementation of the Kriging method is done in Python. A better description of the underlying mathematical background is given in Paragraph 3.4.2, where the above introduction to the Kriging approach is adapted to the special needs of a programming language.

2.4 State of art

2.4.1 Short intake aerodynamics

The aero-engines currently operating on long-range flights are typically characterised by intakes with a ratio between their length and the fan diameter in the range $L/D = 0.5 \div 0.65$ [18]. As described in the introductory chapter, the current trend in the design of next generation aero-engines, along with the increase of BPR and therefore of the size of the engine, requires intakes characterised by a shorter geometry, aimed to avoid the issues related to the increased weight and drag. The analysis of the performance of the bottom line of a short intake was thoroughly studied by Peters [13]. It was found that the critical flight phase for this line was represented by climb conditions, and therefore, as a design guideline, the design process should be mainly oriented to the optimisation of the bottom line in this phase. The aerodynamic performance was analysed for different configurations, specifically at different ratios L/D and with the presence or absence of the fan rotor. The main findings showed that the shortest length which is possible to design corresponds to a ratio equal to $L/D = 0.25$. Below this value, the overall benefits gained in terms of performance are lost due to excessive fan efficiency penalties, which outweigh the benefits potentially gained from nacelle drag reduction.

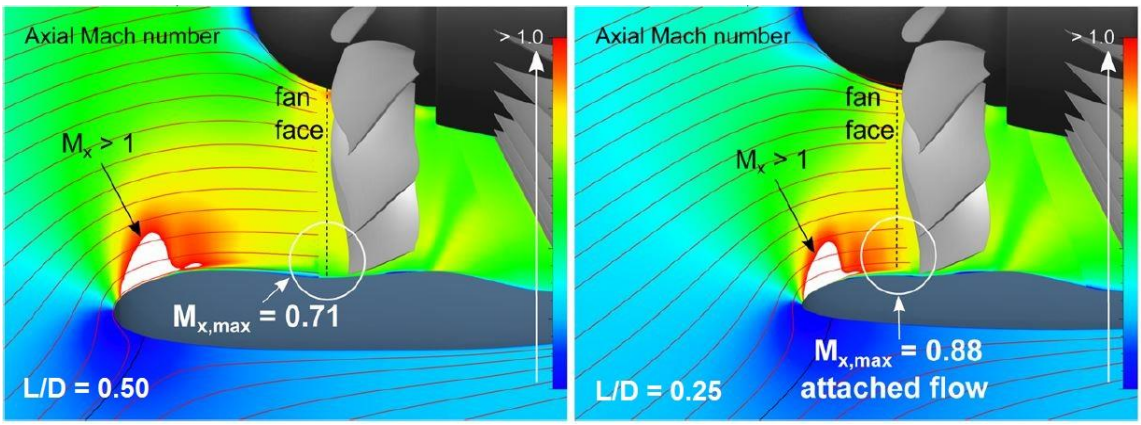


Figure 2.18: Mach number contours at incidence conditions [13]

In Figure 2.18 the Mach number distributions in the region near the bottom line at high incidence conditions are shown for the conventional and short length intakes.

As it can be inferred from the figures, in the former contour plot it is possible to see how the flow, after the initial shock formation at the intake highlight point, re-attaches and reaches the fan face with a lower axial Mach number . Though, the results obtained from the analysis of the short intake solution on the right show that, even though the axial Mach number is higher compared to the first case, the design is characterised by a separation-free inlet flow. Moreover, along with this optimisation achieved at climb conditions, this solution allows also to obtain a reduction of the nacelle drag by 16% at cruise. The guidelines given by Peters as conclusions of the described thorough study suggest that the optimal range in which a short intake should be designed is comprised in $L/D = 0.25 \div 0.4$.

Work on the short intakes design has been also carried out at the Cranfield University. Initially, a tool for the automatic design, meshing and simulation of two-dimensional geometries was designed, which exploited the iCST method [11] to construct aerodynamically-feasible geometries. The initial analysis aimed to validate and improve the tool, and the different tests were carried out on an intake characterised by a ratio $L/D = 0.35$ [19] in flight phases such as cruise and climb. Considering the latter condition, the tool was improved in order to better characterise the issues related to high incidence conditions, moving from inadequate two-dimensional models to more precise three-dimensional representations of the intake, as shown in Figure 2.19 by the Mach number distributions for 3D and 2D axi-symmetric simulated cases.

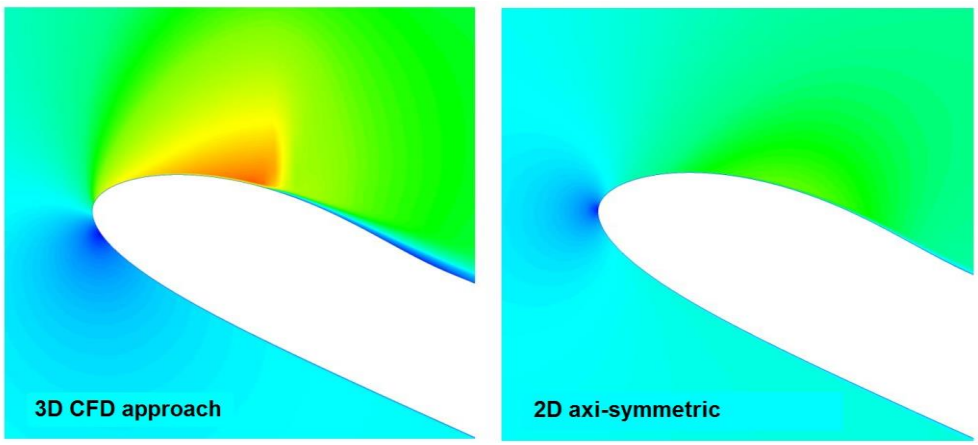


Figure 2.19: Mach number contours for 3D and 2D axi-symmetric designs [11]

2.4.2 Intake design optimisation

The study that followed what introduced in the previous section, aimed to find a way to solve the problem of determining the optimal geometries for the intake bottom line at incidence conditions, is presented. One of the approaches studied exploited a method based on Genetic Algorithms (GA) [20]. The intake examined was characterised by a three-dimensional geometry, in which features belonging to real engines were also included, such as the already described droop and scarf.

The 3D analysis was carried out on the conventional length intake, characterised by a ratio $L/D = 0.50$. The automatic Python script for the design, meshing and CFD simulation of the different three-dimensional cases was tested, and it provided good results in terms of computational efficiency and accuracy. The former one is particularly important in the framework of a study centred on optimisation. For each geometry, the whole process took 2 hours and a half on a 64 CPU machine, and this was the best result achieved with the imposed convergence strategy and a mesh size of almost six millions cells. Even though the mentioned amount of time can be considered a good achievement in terms of computational effort required by a single case with such number of cells, when GA are employed this becomes unfeasible. It was derived that, for optimal results, the optimisation routine should run for 50 generations composed by 50 individuals each. In other words, 2500 total geometries have to be created, meshed and simulated during the whole optimisation process. This translates in a total estimated duration of the process of almost eight months, amount of time required to obtain the optimal solution searched. Logically, this time frame is not suitable for real world applications. For this reason, a new method has to be designed in order to provide a tool for a quicker evaluation of the optimal geometries of the bottom line of a short intake.

One of the main key findings extracted from the GA optimisation is graphically reported in Figure 2.20. In the plot it is shown a close-up of the Pareto plot which was obtained after five generations of the optimisation routine. It highlights how well the method is able to optimise the geometries at each generation and obtain improved values of the examined aerodynamics performance parameters.

Moreover, this clearly shows a key feature of the whole study, which was already identified during the two-dimensional study and found confirmation in the three-dimensional approach: a linear relationship between the distortion coefficient and the pressure losses encountered in the intake duct is present. In fact, it is demonstrated that a reduction of DC_{60} implies an increase of IPR, which both lead to better intake performance. Nevertheless, as reported in Paragraph 2.1.2.2, this conclusion is not always verified. Therefore, particular attention must be paid and before drawing any conclusions, both aerodynamics parameters should be thoroughly examined. Anyway, for the preliminary analysis of the geometry of the bottom line, this finding allows to enhance the study of the intake performance, focusing the attention on a single aerodynamics performance parameter, which in the particular case of climb conditions can be identified in the distortion coefficient DC_{60}.

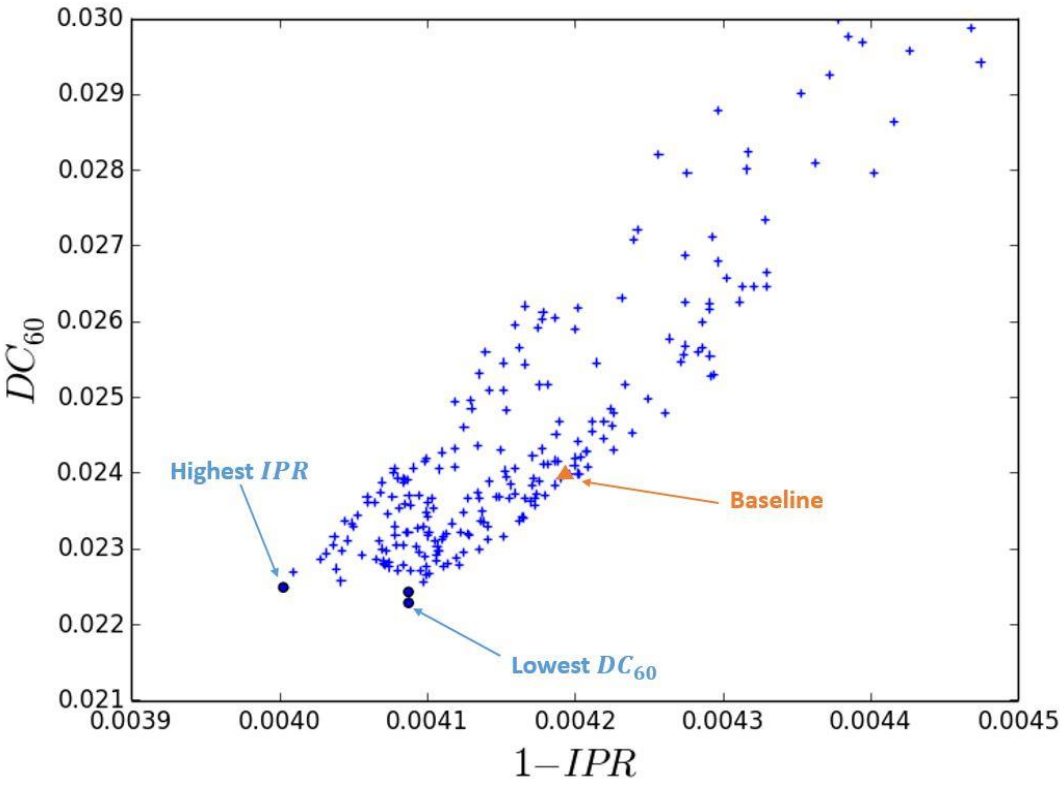


Figure 2.20: Pareto plot obtained after five generations of optimisation [21]

2.4.3 Design space exploration

The last step taken prior the beginning of current research project consisted in the initial analysis of the design space of the bottom line of a short intake. The flight condition in which this line was analysed was the climb condition, which can be assumed as the most challenging that this line can face, due to all the aerodynamics issues that can originate. The short intake design considered in the framework of the project was characterised by a ratio $L/D = 0.35$ [22], considered as a design trade-off following the thorough study of Peters [13].

The initial exploration was carried out exploiting a LHS DoE based on five independent design variables, namely CR, AR, f_{IL}, κ_{TH} and $f_{\kappa_{FAN}}$, which were used to describe the geometry of the bottom line as described in Paragraph 2.1.3. The metrics considered in the analysis were IPR, DC_{60} and $M_{max,isen}$, where the first two are considered as aerodynamic performance metrics, whereas the latter can be assumed as an indicator of performance. The design space bounds were initially set based on past experience and the initial population size of the DoE was set to 127 total samples. Eventually, the outcome of the exploration was characterised mainly by noisy results, between which geometries with multiple inflections and under-performing designs could be found. Due to these geometrical instabilities and poor associated performance, only 16% of the initial DoE geometries was considered aerodynamically satisfactory. This identified group of geometries was characterised by a value of distortion below 0.2, considered as a cut-off limit above which the performance is no longer generally acceptable [22]. The aim of the design space exploration was mainly related to find potential trends between design variables and metrics considered, in order to reduce the dimensionality of the problem and optimise the process of creation of new geometries. Nevertheless, the outcome did not allow to extract any trends and therefore it could be assumed that no correlations exist between the variables.

The research focused therefore on an alternative solution aimed to reduce the number of independent design variables involved in the definition of the geometry of the bottom line. The solution was found in the creation of an Python algorithm

which automatically calculates and defines the two curvatures at the throat and fan face points. This implies that just three of the five independent variables considered in the beginning of this analysis were maintained independent and variable in determined ranges. As an additional advantage of this solution, it is technically possible to avoid the issue related to the generation of multiple inflections in the profile shape, which is a particular problem when a reduced number of independent variables is used to create the geometry.

The discussion reported poses the foundations of the work carried out during this research project. The reduced dimensionality of the problem allows to speed up the process of creation of different geometries and, in turn, this allows a better and efficient exploration of the design space, which is paramount for the purposes of creation of a well-suited surrogate model.

METHODOLOGY

The aim of this paragraph is to give an overview of the methodologies employed to define, design and study the geometry of the bottom line of the short intake. The correct formulation of the problem is considered as a key point in the analysis and particular attention has to be placed in the selection of design variables, objectives and constraints.

3.1 Geometrical parameterisation

The use of geometrical parameterisation is a key driver when DoE analysis is involved. It allows to efficiently optimise the process of creation of each geometry in the design space by means of a parametric curve, reducing the number of design variables required [23] and therefore reducing the dimensionality of the problem. Between the different parameterisation techniques available, the one selected for the scope of the analysis is the Intuitive Class Shape Transformation (iCST) method, developed at the Cranfield University [23]. This parameterisation technique was born from the coupling of the Class Shape Transformation (CST) method, that has been proven to be very efficient in the geometrical optimisation of nacelles and intakes [24], and the parameterisation method for aerofoil sections called PARSEC [25], which adds intuitiveness to the design parameters for the definition of constraints, missing characteristic in the former method. The CST method is the best trade-off for 2D parameterisation, since, as stated, it permits to represent the shape of an aerodynamic profile with a limited number of design variables ensuring an efficient and rapid geometry definition [24].

3.1.1 2D parameterisation

In the CST method, a geometry is represented as the product of a class function $C(\psi)$, which defines the basic profile shape, and a shape function $S(\psi)$, as expressed in Equation (3.1). The additional term $\psi\Delta\xi_{TE}$ is required to modify the

ordinate of the end-point of the curve, since a condition imposed considers the starting and end-points of the profile placed at the same value of ordinates. As expressed in Equation (3.1), the parametric approach requires the non-dimensionalisation of the coordinates of all the points of the geometry, which are for this reason divided by the chord of the profile.

$$\xi(\psi) = S(\psi)C(\psi) + \psi\Delta\xi_{te} \qquad \xi = \frac{y}{c}, \psi = \frac{x}{c} \tag{3.1}$$

Kulfan [26] defines the class function $C(\psi)$ in the form defined in Equation (3.2).

$$C_{N_2}^{N_1}(\psi) = \psi^{N_1}[1-\psi]^{N_2} \quad for\ 0 \leq \psi \leq 1 \tag{3.2}$$

Between the different basic shapes that Equation (3.2) can describe, the one of interest in the framework of this research is the round-nosed aerofoil shape, obtained imposing $N_1 = 0.5$ and $N_2 = 1$, as shown in Equation (3.3).

$$C(\psi) = \sqrt{\psi}\,[1-\psi] \tag{3.3}$$

In Figure 3.1 it is possible to visualise how the CST method works. The shape function is commonly represented by Bernstein polynomials [26] which, through the use of different Bernstein polynomial coefficients, are transformed. Since these coefficient are not aerodynamically intuitive, the iCST method relates them to the design variables, and they can be analytically calculated from the these [23]. The output of the multiplication between the basic profile, described by the class function, and the transformed shape function is shown in the plot in the bottom of Figure 3.1. This demonstrates the capabilities of the iCST method to model smooth curves for the design of aerodynamic shapes.

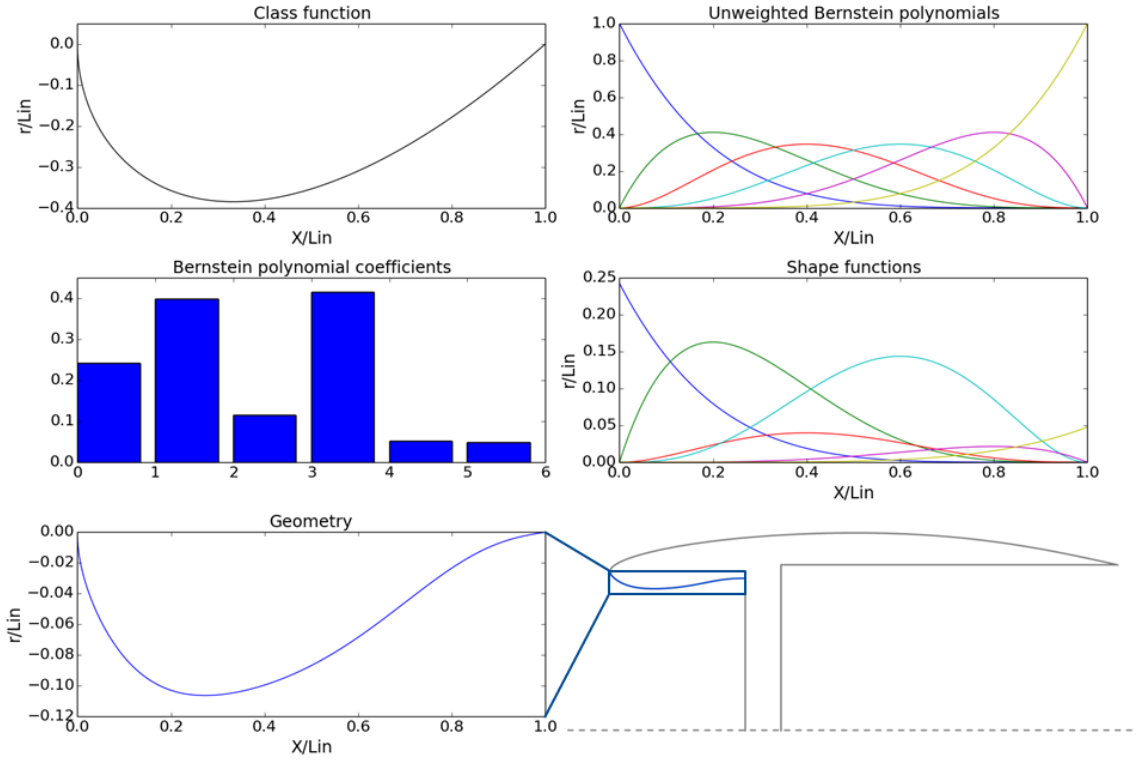


Figure 3.1: Intake CST parameterisation [11]

3.1.2 3D parameterisation

The engine which has to be simulated is characterised by a three-dimensional dropped and scarfed non-axisymmetric configuration, as it can be usually seen in modern aero engines. In order to generate such design, the two-dimensional iCST approach has been exploited in the representation of three-dimensional geometries and this task has been addressed in the past approach [12]. It is possible to design the three-dimensional engine cowl specifying the geometry of different sections in the azimuthal direction. To improve the efficiency of the process, the geometry of four main aero-lines is defined and the profiles are those of the top, side, control and bottom lines. Nevertheless, the three-dimensional geometry should be described by a greater number of aero-lines in order to improve the resolution of the final output. Through sensitivity analysis was found that in order to obtain accurate results the minimum number of cuts should be 256. Therefore, the geometries of all the aero-lines which are not specified by the

user are obtained from the interpolation of the constraints set on the user-defined aero-lines. A graphical representation of what discussed is shown in Figure 3.2.

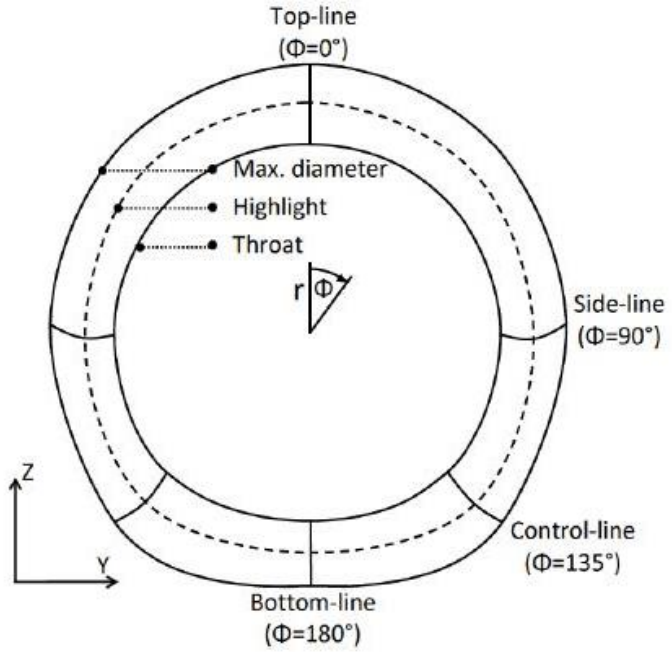


Figure 3.2: Azimuthal variations of the iCST constraints [12]

The focus of the current analysis is oriented only on the optimisation of the bottom line of the intake. Therefore, it was decided to fix the geometries of the other principal aero-lines, while varying the parameters which define the intake bottom line geometry.

3.2 CFD methodology

Computational Fluid Dynamics (CFD) is a powerful tool for the study of complex flow phenomena [27], as the ones which characterise the flow inside an intake at incidence. As previously stated, the aerodynamic properties which are considered of most interest to study this flight phase are IPR and DC_{60}. The software suite used for meshing, simulate and post-process the results was ANSYS, through the use of, respectively, ICEM 15, Fluent 15 and CFD-Post 17.1.

3.2.1 Meshing methodology

The first assumption to take into consideration in the CFD approach to this problem regards the size of the domain which has to be simulated. Since the engine can be considered symmetric to the vertical plane which contains the intake top and bottom profiles, it is possible to lighten the simulation analysing just half of the whole initial domain. Nevertheless, even though this assumption can be assumed as valid from a geometrical point of view, there are situations in which the flow-field inside the intake cannot be considered axisymmetric, such as in conditions of cross-wind. At incidence, though, it is possible to consider the internal flow-field as axisymmetric, and this strengthens the initial assumption of symmetry based on the geometry definition. The mesh was produced with ANSYS ICEM 15 through an automatic meshing tool written in Python and developed at the Cranfield University, which allowed to obtain a fully-structured multi-block mesh. Another important step in the early stages of the meshing procedure regards the correct setup of the distances between the geometry and the boundaries of the domain, in order to avoid the possible numerical errors that can occur in the area of interest and also to permit to the flow to stabilise before reaching the boundaries. For all these reasons, the domain considered was semi-spherical with a radius equal to 40 times the maximum diameter of the nacelle.

The mesh size employed in the current approach was the same as the one used in the past approach [21]. A mesh independence study was carried out to select the most efficient mesh size in order to speed up the simulation and decrease the computational effort required to complete a DoE run. Following the study, it was found that the results in IPR and DC_{60} were mesh independent and the optimal mesh size which was eventually selected had almost six million nodes with a $y^+ \leq 1$, usually imposed in order to obtain more accurate results.

3.2.2 Solver settings

The simulations were run on ANSYS Fluent 15. The solver setup is common to all the simulations run during the DoE. Particular attention has to be placed on this task, since it can affect the convergence of the simulations and lead to unwanted

problems or failed solutions. The simulations are considered as steady and the solver type for the analysis is an implicit density-based solver. The aerodynamics of the flow inside the intake in condition of flight at incidence may involve regions of separated flow for certain geometries. In order to solve turbulent motion, the use of turbulence models based on the Reynolds-Averaged Navier-Stokes (RANS) equations is mandatory. The selected turbulence model must be able to deal with high adverse pressure gradients and to predict the transition from laminar to turbulent flow. In fact, the growth of the boundary layer has to be adequately represented, since the shock-induced separation influences directly the aerodynamics parameters take into considerations for the analysis and it affects the performance of the whole engine. After these considerations, the model used for the analysis was the k-ω Shear Stress Transport (SST), one of the most valuable and accurate models available at the moment for the prediction of flow separation. It combines the advantages earned from another turbulence model, the k-ε model, in the outer free-stream region and the advantages of the k-ω model in the near-wall region. Air was modelled as a compressible gas and its density was calculated with the ideal gas law. The solution method used was an implicit scheme with flux calculations based on the RoE-FDS scheme and with gradient evaluation, required to compute velocity derivatives, calculated using the Green-Gauss node-based approach [23].

3.2.3 Boundary conditions

As previously discussed, since the engine has been cut in two symmetric parts with respect to the vertical plane, a symmetry condition has to be set on the plane where the engine lies. The boundary condition related to the pressure far field is defined on this plane and extended to the whole three-dimensional domain external the engine. This condition required the definition of static pressure, temperature, Mach number, velocity components and turbulence parameters. The pressure outlet condition was set to simulate the engine fan face, but as it can be visualised in the figure the condition was set downstream the nominal fan face. This was helpful to aid convergence, avoiding the influence that this condition could have on the real plane [23]. The pressure outlet condition required the

specification of static pressure, total temperature, target mass flow rate and turbulence parameters, as defined for the pressure far field condition. All the values required by these boundary conditions were derived by previous works carried out and which set the parameters for the operating condition in which the current simulated engine works. The last boundary condition set was the pressure inlet at the outlet of the engine. The required value of static pressure was obtained from a previous work carried out with Turbomatch at the Cranfield University.

3.2.4 Convergence criteria

The different geometries created within the DoE should be automatically meshed and simulated without encountering potential issues which can originate at any flow condition the engine may be involved in. This requires a solid convergence strategy, especially because for some geometries there is the possibility to experience large regions of separated flow, which can lead to a more difficult convergence. In the past approach [21] a unified convergence strategy has been developed, which is able to deal even with the most challenging cases maintaining the computational cost within acceptable limits. In order to improve the convergence, the under-relaxation factor for the specific turbulent dissipation rate is lowered to 0.05, while the other are maintained to their standard values. Each simulation has a standard duration of 3000 iterations, of which the first half is run with a first order discretisation scheme which then changes to second order for the remaining part. Moreover, during the first half of the simulation, the Courant number is increased from 1 to 50 at non-regular steps almost every 200 iterations. The same happens during the following 1200 iterations, but in this interval the Courant number varies between 1 and 30. The last 300 steps are used to achieve the prescribed mass flow rate, setting therefore a target mass flow as objective. The simulations are run on the Astral cluster, the High Computational Power (HPC) cluster available at the Cranfield University. Considering the selection of 32 CPU on the mentioned cluster, each simulation takes approximately 3 hours and a half to complete.

3.3 Design of Experiment

Design of Experiment (DoE) is defined as the process of planning an experiment in order to collect meaningful data and draw valid conclusions on the behaviour of a problem of interest [14]. DoE has increasingly become an important mean in conceptual design and Multi-Disciplinary Optimisation (MDO).

The current analysis focuses solely on the optimal design of the bottom lip, which geometry can be assumed defined principally by five design variables. Among these, just three are considered as independent variables: contraction ratio (CR), aspect ratio (AR) and a coefficient which sets the radius of curvature at the highlight, represented by f_{IL}, and obtained from Equation (2.9).

The other two remaining parameters, considered as design variables, are the curvature at the throat κ_{TH} and the curvature at the fan face κ_{FAN}. For sake of completeness, the curvature is generally defined as in Equation (3.4), where r is defined as the radius of curvature in the point where the curvature is calculated.

$$\kappa = \frac{\left|\frac{d^2R}{dx^2}\right|}{\left[1+\left(\frac{dR}{dx}\right)^2\right]^{3/2}} = \frac{1}{r} \tag{3.4}$$

Considering for instance the intake throat, in this location the first derivative of the radius of curvature is set to zero. The curvature is therefore given by Equation (3.5).

$$\kappa_{TH} = \left|\frac{d^2y}{dx^2}\right| = \frac{1}{r_{TH}} \tag{3.5}$$

The two curvatures mentioned can be considered as two “derived” independent variables, since they are implied in the definition of the bottom line geometry, but they are intrinsically dependent from the other three variables.

The non-dimensional length of the intake L/D, expressed as the ratio between the length of the intake itself and the fan diameter, is set to 0.35, which is a value comprised in the optimal range for short intakes studied by Peters [13]. It ensures

that the engine performance is maintained almost at the same level as in the conventional length intake case, usually characterised by a ratio $L/D = 0.5$, without jeopardising the operability and also leading to a general improvement in terms of weight and drag.

The use of dimensional values in the phase of creation of a design space could likely lead to failed or infeasible cases due to scaling issues [16]. Therefore, in order to avoid these problems, as a rule of thumb the parameters should be always exploited in their non-dimensional form after the normalisation into the unit cube, defined by the [0,1] range.

3.3.1 Design space bounds definition

Prior to obtain the final DoE that will be used to derive the RSM, an extensive investigation has to be carried out in order to find the optimal values for the design space bounds. Therefore, initially, the research focused on the improvement of the design space bounds used in the previous approach to the problem [21]. The initial objective was therefore to reshape these bounds, aiming to achieve a twofold result: eliminate regions of poor interest for the intake design, characterised solely by geometries which lead to poor aerodynamic performance; explore new regions that could be overlooked in the past analysis. Moreover, another important matter to take into consideration comparing the current work with the previous one is the number of independent design variables considered. This has been reduced from five to three, and therefore an investigation is due in order to review the possible changes caused by this dimensionality reduction.

3.3.1.1 Design space exploration through FF DoE

The investigation and determination of the new design space bounds have been carried out exploiting the Full Factorial (FF) DoE technique. As described in Paragraph 2.2.1, with this approach the design space is divided in a multi-dimensional regular grid and the responses are evaluated in its nodes. Starting from the bounds determined in the past approach for the three independent design variables considered in the current analysis, the design space is sliced in equal parts. Considering the variation of the bounds of one variable at a time, the

approach permits to expand uniformly these bounds until the region characterised uniquely by bad geometries is found. The last extreme feasible values of the considered variable determines the upper and lower bounds. As it will be discussed later in Paragraph 4.1, the determination of the bounds is based on the value of DC_{60}, which is the critical parameter of major interest in this analysis. The post-processing of the results obtained with the Full Factorial DoE will be carried out with the help of pair plots. Since there are just three independent variables in the problem, it is pretty easy to study the design space for the bounds evaluation. The limits for two of the three variables are set and kept fixed across all the plots, while, in regard of the left out variable, its value will vary from plot to plot. Therefore, the final output comprises a series of pair plots at different levels of the left out variable. From the visual inspection of these, aided by the visualisation of contour lines for different levels of DC_{60}, it will be possible to understand where the lower and upper bounds will be reached. As a rule of thumb, a region with a DC_{60} greater than 0.20 will not be considered. Since the design space considered can be viewed as a regular n-dimensional polytope, in the two-dimensional visualisation with pair plots the design space bounds are represented by vertical and horizontal lines. Hence, concerning the variables defined on the Cartesian axes, the values corresponding to the intersections of the vertical and horizontal lines passing for the extreme allowable values found and the axes define the bounds of the design space.

As described in Paragraph 2.2.1, along with the definition of the number of independent variables, the creation of a FF DoE requires the definition of the number of levels with which the interval of each variable is subdivided. The approach followed employed three independent variables and five levels. Therefore, the total number of samples analysed is equal to $3^5 = 243$. This corresponds to the initial size of the design space, which can be further expanded if the bounds are not found in the first run.

3.3.2 Design space creation

Once the bounds of the design space have been set for all the three independent variables, the final DoE can be created and populated. The sampling technique selected to accomplish this task is the Latin Hypercube Sampling method since, as described thoroughly in Paragraph 2.2.2, it is one of the best methods which allow to obtain a uniform distribution of samples which are well spread in the design space, especially thanks to the space-filling random distribution feature. Following the determination of the bounds of the design space through the FF DoE, it has been found that the amount of geometries created with a single inflection point, located in the diffuser, is around 38% of the total number of geometries created. The decision about the number of samples which should populate the LHS design space was based on the number of cases which ideally populate a FF design space composed by five variables and three levels. This means that the total number of cases which should be found in the final DoE is equal to 243. Therefore, considering the previous mentioned point regarding the generation of geometries with multiple inflection points, the number of cases which has to be simulated in order to populate the LHS design space is given by $243/0.38 = 639$, which is reduced and rounded to 625 total cases. Assuming that the same ratio between number of single inflection designs and total number of geometries remains, the final population of the DoE for RSM purposes will be equal to 243 samples.

3.4 Surrogate modelling

This paragraph is divided in two main sections, which will give an overview of the two investigated methodologies employed to create the metamodels. The first part focuses on the implementation of the Kriging method, established method for the creation of response surfaces from highly non-linear data and introduced in Paragraph 2.3.1; the second part will introduce a hybrid method which has been subject of investigation in the creation of surrogate models for intake and nozzles during this year at the Cranfield University. The methods to obtain the response surfaces are developed in Python 2.7.

3.4.1 Design variable selection

The geometries created during the LHS DoE are all based on three principal independent variables and two derived independent variables, as reported in the introduction of this paragraph. The approach to the creation of the metamodels focused on two different sets of variables. The first tested approach was based on three variables, namely the CR, AR and the radius of curvature at the highlight r_{IL}. The use of the latter rather than the use of the non-dimensional factor f_{IL}, used to obtain the different geometries, was considered more suitable for a real application based on surrogate modelling. In fact, considering the exploitation of the metamodels in real world applications, from a design point of view the definition of a radius can be considered more meaningful than the definition of a factor.

3.4.2 GPML Kriging Model

The mandatory step before the creation of the metamodel consists in the scaling of the explanatory variables $\boldsymbol{x}$ into the unit cube $[0,1]^p$, with p the number of dimensions of the problem. This step is important for two main reasons: it can avoid subsequent multi-dimensional scaling issues [16]; it permits to have the same degree of activity of the autocorrelation $\hat{\theta}_j$ parameter from problem to problem [15]. Objective of the current research is to construct a Kriging approximation model based on the LHS DoE obtained and then use this model as a surrogate for further analysis.

According to the computer experiments literature, the most popular method for generating functions are the Gaussian stochastic processes [28]. The Kriging model is obtained exploiting the Gaussian Processes class implemented in the scikit-learn module for Python [29]. In the environment of this programming language, the mentioned class is more commonly called Gaussian Processes for Machine Learning (GPML) and it consists of a generic supervised learning method designed for regression and classification problems [30]. Since the current work concerns a regression problem, the approach considered takes advantage of some of the potentialities implemented in GPML for regression, such as:

- Possibility to define confidence intervals which are used to refit the prediction in some region of interest.
- Possibility to exploit different regression and correlation models, along with the chance to define custom models if required.

3.4.2.1 GPML mathematical formulation

The mathematical formulation thoroughly presented in Paragraph 2.3.1 is reported in this section in a simplified way to support the setup decisions made in the creation of the Kriging metamodel.

Consider the unknown function f in $\mathbb{R}^p \rightarrow \mathbb{R}$ which models the behaviour of the design space obtained from the Latin Hypercube Sampling. GPML assumes that this function can be interpreted as the conditional sample path of a Gaussian process G [28], defined as in Equation (3.6).

$$GP(\boldsymbol{x}) = \sum_{i=1}^{p} g_i(\boldsymbol{x})\beta_i + Z(\boldsymbol{x}) = \boldsymbol{g}^T(\boldsymbol{x})\boldsymbol{\beta} + Z(\boldsymbol{x}) \tag{3.6}$$

The first term on the right member of the equation, g_i, represents the *i-th* known regression functions, while $\boldsymbol{\beta}$ is a vector of unknown regression coefficients. Z is defined as a zero-mean Gaussian process with a fully stationary covariance function shown in Equation (3.7).

$$\mathrm{cov}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) = \sigma^2 \,\mathrm{corr}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) \tag{3.7}$$

The σ^2 is the process variance and $\mathrm{corr}\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right)$ is the correlation matrix defined also in Paragraph 2.3.1. The objective is to find the Best Linear Unbiased Prediction (BLUP) of the sample path g condition on the observations [29], described by Equation (3.8).

$$\widehat{GP}(\boldsymbol{x}) = GP(\, \boldsymbol{x} \mid y_1 = f(x_1), \dots, y_n = f(x_n)\,) \tag{3.8}$$

This prediction $\hat{G}(\boldsymbol{x})$ is characterised by three main properties:

- It is given by a linear combination of observations $\widehat{GP}(\boldsymbol{x}) = a^T(\boldsymbol{x})\, y$

- It can be considered unbiased $\mathrm{E}\left[\mathrm{GP}(\mathbf{x}) - \widehat{GP}(\boldsymbol{x})\right] = 0$
- The MSE is minimum $\widehat{GP}(\boldsymbol{x})^* = \min \mathrm{E}\left[(\mathrm{GP}(\mathbf{x}) - \widehat{GP}(\boldsymbol{x}))^2\right]$

Eventually, the BLUP is defined as a Gaussian random variate with mean given by Equation (3.9)

$$\mu_{\hat{Y}}(\boldsymbol{x}) = g^T(\boldsymbol{x})\hat{\beta} + r^T(\boldsymbol{x})\gamma \tag{3.9}$$

and variance expressed as in Equation (3.10).

$$\sigma^2{}_{\hat{Y}}(\boldsymbol{x}) = \sigma^2{}_{Y}\left(1 - r^T(\boldsymbol{x})\, corr^{-1}\, r(\boldsymbol{x}) + u^T(\boldsymbol{x})(F^T corr^{-1} F)^{-1}\, u(\boldsymbol{x})\right) \tag{3.10}$$

Equations (3.9) and (3.10) introduce from a mathematical point of view the terms which are found in the setup of the Kriging model. The most important and fundamental terms that should be mentioned are:

- Correlation matrix, defined by the correlation function and the θ parameter

$$\mathrm{corr}_{ij} = \mathrm{corr}\left(\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right), \theta\right), \qquad i, j = 1, \dots, n \tag{3.11}$$

- Vector of cross-correlations between the point where the prediction is made and the points in the DoE

$$\mathrm{r}_i = \mathrm{corr}\left(\left(\boldsymbol{x}, \boldsymbol{x}^{(i)}\right), \theta\right), \qquad i = 1, \dots, n \tag{3.12}$$

- Regression matrix

$$\mathrm{G}_{ij} = g_i\left(\boldsymbol{x}_j\right), \qquad i = 1, \dots, p, \qquad j = 1, \dots, n \tag{3.13}$$

The definition of the correlation and regression functions cannot be made in advance, but it requires a series of empirical test to select the more suitable [29]. Once these functions have been selected, the procedure focuses on the setup of the various required parameters. These can be determined and fixed for the whole duration of the process, but usually it is preferred to calculate them through a Maximum Likelihood Estimation (MLE) technique. This turns the estimation problem into a global optimisation problem and evaluates the auto-correlation

parameters searching for the optimum values within determined bounds set by the user.

3.4.2.2 Kriging model setup

The setup involves the selection of many different parameters and settings, leading to a great number of possible available combinations. Nevertheless, according to the design space analysis consequent the DoE, some of them could be discarded in advance due to expected poor feasibility. In any case, it is required to the user to test the remaining combinations and find the more suitable for the needs of the problem. The documentation available [29] for the model setup is pretty thorough, but this does not allow to avoid the empirical tests to find out the optimal settings. For sake of clarity, in the following discussion the names of the different functions and parameters that are found in the GPML module are identified with a different font, specifically Courier New, to facilitate the reader to identify them in the code.

The first two mandatory choices that have to be taken regard the selection of the correlation and regression functions.

In the GPML module different correlation functions are implemented, but the ones of interest for the majority of the problems are, specifically, the `absolute_exponential`, `squared_exponential`, `linear` and `cubic` functions. The inputs required by the functions are common between all of them, and they comprise:

- Auto-correlation parameter θ, passed as a single value or as a vector of values $\theta = [\theta_1, \theta_2, \dots, \theta_p]$, each one dependently determined by its corresponding predictor $\boldsymbol{x}_p$.
- A vector containing all the componentwise distances between each pair of samples in the design space. In other words, for each dimension all the distances between each possible pair of observations are evaluated. The total number M of cross-distances $d_k^{(ij)}$ possible in one single direction is given by Equation (3.14), where n represents, as usual, the total number of samples.

$$M = \frac{n\,(n-1)}{2} \tag{3.14}$$

Each correlation function is characterised by its own expression. Considering the functions of greatest interest, which can be identified as the first two mentioned above, they are expressed by the following equations:

- `Absolute_exponential` (Ornstein-Uhlenbeck stochastic process)
  For a pair of samples identified as i and j, the m^{th} correlation term determined by this correlation function is given by Equation (3.15).

$$r_m(\theta, \boldsymbol{x}) = \exp\left(\sum_{k=1}^{p} -\theta_k \left|d_k^{(ij)}\right|\right) \tag{3.15}$$

- `Squared_exponential` (infinitely differentiable stochastic process)
  For a pair of samples identified as i and j, the m^{th} correlation term determined by this correlation function is given by Equation (3.16), where the main difference with the previous one is that each cross-distance is elevated to two.

$$r_m(\theta, \boldsymbol{x}) = \exp\left(\sum_{k=1}^{p} -\theta_k \left|d_k^{(ij)}\right|^2\right) \tag{3.16}$$

The output $r(\theta, \boldsymbol{x})$ is an array of the length M, where each term is represented by $r_m(\theta, \boldsymbol{x})$, which, as presented in the above reported equations, represents a correlation-weighted sum of the distances in each dimension between a generic pair of samples. After several test, it has been demonstrated that for the current problem of interest the `absolute_exponential` function allows to obtain the best results and it can therefore assumed as the correct type of correlation between the samples. As stated on the guide [29], this is often also the case when, from the study of the dataset obtained, the original experiment is known to not be smooth and therefore infinitely differentiable.

The built-in regression models available in the GPML module are, specifically, the "`constant`", "`linear`" and "`quadratic`" functions, which are linear regression

models which represent, respectively, zero-, first- and second-order polynomials. The only input required by all of them is the predictors vector $\boldsymbol{X} = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \dots, \boldsymbol{x}^{(n)}\}$. The resulting output is a vector of ones, in the case of the "*constant*" regression function, or a matrix containing the values of the functional set. Sometimes it could happen that the mentioned functions implemented in the module do not work properly with the dataset available. In this situation it is possible to derive a custom regression function, designed and coded as the default models. The first rule to respect in the creation of the new function is that the number of terms in the regression model must not be greater than the number of observations in the DoE. This ensures that the underlying regression problem is not underdetermined.

Of at least the same importance, along with the two fundamental functions above mentioned, also the type of optimiser to use. As discussed in the paragraph regarding the mathematical formulation for the GPML, the estimation problem is turned into a global optimisation problem and this has to be solved exploiting one of the two optimisation functions implemented in the GPML module. They are identified by the names "fmin_cobyla" and "Welch". The former one is an optimisation algorithm contained in the SciPy library for Python [31], and generally it aims to minimise a function exploiting the Constrained Optimisation BY Linear Approximation (COBYLA) method [32]. The latter method has been developed by Welch et al. [33] in order to lower the computational cost involved in the MLE of the correlation parameters for high-dimensional problems. For the purpose of this analysis both method have been tested. The better results are obtained with the COBYLA algorithm and it is therefore selected as the optimisation algorithm for the model creation.

After the selection of the different functions to exploit, there are some important parameters that have to be adequately selected. The first that will be considered is the auto-correlation parameter θ. This can be input as a single value, which is therefore the same for all the dimensions, or as a vector of length p, in which it is possible to define a different value of θ for each feature. Of primary importance, the GPML requires the definition of an initial value of the auto-correlation

parameter, or set of values, commonly identified with the term θ_0 or `theta0` in the code, in order to initialise the MLE of the same parameter. Though, this estimation starts just if also the values, or vectors of values, for the lower and upper bounds are set. These are identified respectively as θ_L, or `thetaL`, and θ_U, or `thetaU`. When they are not specified, their default value is `None` and the MLE does not start, setting θ_0 as the fixed and general value of the auto-correlation parameter for the whole estimation. After several tests, two important points have been found:

- As it will be reported in Paragraph 4.4, optimal solutions have been obtained both from the definition of a fixed auto-correlation parameter and from the use of the MLE for its determination.
- The optimal values used for the creation of the different surrogate models reported in Paragraph 4.4 are given in Table 3.1. These are given as results of different tests which have proven that a narrower interval should be maintained around the θ_0 values. The two main reasons associated with this are: increasing the size of the interval generally requires an increase of the number of MLE cycles within the interval, which in turn leads to very high computational effort for a single evaluation; if a low number of cycles is imposed (around 200) the results obtained exploiting the MLE on a large interval are characterised by a large error.

Table 3.1: Auto-correlation parameter values definition for MLE

	CR	AR	r_{IL}	κ_{TH}
θ_0	1	1	1	0.5
θ_L	0.9	0.9	0.9	0.4
θ_U	1.1	1.1	1.1	0.6

In case the limits of the θ parameter are set, the estimation of the auto-correlation parameter is performed, but it requires the definition of an additional parameter which determines the number of times the MLE will be performed. The starting point of the process is always the value, or values, of θ_0. The following value of θ is determined randomly in the interval between θ_L and θ_U, according to a log-uniform distribution, and this is repeated for the number of restarts specified. In the code this parameter is defined as `random_start`, and it has to be set carefully in order to find the optimal θ within the range, otherwise, as reported earlier, it is highly likely to obtain wrong results. For the creation of the metamodels reported in Paragraph 4.4 a number of restarts equal to 200 has been set, trade-off between computational cost and accuracy. Moreover, it has been found that increasing this number above 200 does not produce significant improvements, but instead it slows down the construction of the model.

Last but not least, another important parameter is the so-called `nugget`. This term can be a single value, valid for all the samples included in the design space, or it can be defined individually for each sample, in a vector of length n. Cavazzuti [15] defines it as a noise parameter, since it is usually exploited when noisy data are involved. In this case, it is possible to specify the nugget as the variance of the noise for each point. Considering for example the `squared_exponential` auto-correlation function, the term is defined as in Equation (3.17) [29].

$$\text{nugget}_i = \left[\frac{\sigma_i}{y_i}\right]^2 \tag{3.17}$$

Basically, if its value is zero, the Kriging method is defined as interpolating and the response surface exactly pass through the DoE data, condition which is often difficult to achieve; if its value is different from zero the Kriging method becomes an approximating method, and the parameter allows smooth predictions from noisy data, as it can be seen in . In the code, the `nugget` is added to the diagonal of the covariance matrix, acting as Tikhonov regularisation term, which is a method of regularisation of ill-posed problems [29]. Put simply, in the resolution of a system of linear equations, such as $A\boldsymbol{x} = b$, the OLS approach can lead to situations

with no solutions or situations identified by more than one solution. Since the OLS method searches for the minimum sum of squared residuals $\min\|A\boldsymbol{x} - b\|^2$, a regularisation term is added as in Equation (3.18).

$$\min\|A\boldsymbol{x} - b\|^2 + \|\Gamma\boldsymbol{x}\|^2 \tag{3.18}$$

The term Γ is the Tikhonov matrix, usually given by $\Gamma = \alpha I$, I the identity matrix. The `nugget` definition is not a straightforward task, since it involves a lot of tests to find the optimal solution. The most feasible, in the context of this work, has been found in the selection of a different value for each sample. Also, after several tests, it has been decided that a hybrid solution provides the best results. This solution considers two different nugget values for two different regions in the design space:

- A nugget value equal to machine precision ($\sim 10^{-16}$) is assigned to the samples characterised by a value of DC_{60} in the range of interest.
- A nugget value equal to the value of the metric is assigned to all the samples outside the range of interest.

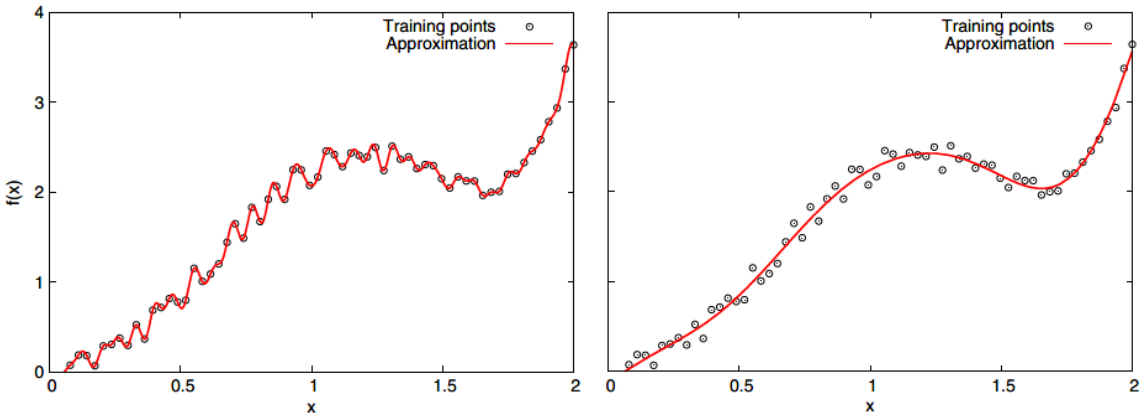


Figure 3.3: Nugget effect [34]

An alternative tested consisted in the definition of a different nugget value for each sample, values defined as fractions of each metric value. Nevertheless, this solution showed good outcomes in the beginning, but it was then overcome by the above mentioned solution.

RESULTS AND ANALYSIS

The discussion presented in this chapter follows the order in which the process of creation of a metamodel for design purposes should be carried out. The outcome of the design space exploration for the determination of its bounds is presented and it represents the base of the current work. Next, the result obtained from the creation of the dataset within the bounds previously determined is presented in the Design space bounds determination paragraph, along with the thorough preliminary analysis carried out on the DoE. Another important step in the analysis is represented by the geometrical study of the shape of the intake, which is reported in the Lip design analysis paragraph. Eventually, the process of creation of the response surfaces through the two methodologies described in Paragraph 3.4 are presented in the Surrogate modelling section, along with their validation using the Leave-One-Out (LOO) method.

4.1 Design space bounds determination

The preliminary exploration of the design space for the design of short intakes has to be carried out in order to establish the bounds within which only feasible geometries can be created. This exploration allows indeed to identify regions characterised solely by geometries with non-desired features, such as multiple inflections along the intake length, or characterised uniquely by design which lead to high distortion and significant losses. The reduction of dimensionality of the problem, along with the automatic determination of the other two derived independent variables, has required the identification of new different bounds in comparison to those found during the previous year. The determination of the design space bounds, as briefly reported in Paragraph 3.3.1.1, is based on a Full Factorial (FF) DoE. This allows to explore the design space in an ordered way, analysing the results on different slices and allowing a uniform expansion of the bounds. The initial step in the exploration involved the exploitation of a FF defined

by three levels which, along with the three independent variables defined, determined an initial population of 27 geometries. For each geometry the other two derived independent variables have been calculated using the approach already described in the introduction of Paragraph 2.4.3. The method succeeded in the creation of a smooth geometry from the highlight to the throat, but in some cases it led to geometries characterised by multiple inflections in the diffuser. This could be mainly related to the fact that for certain values of the three main independent variables a smooth curvature distribution can be obtained on the lip, since the three parameters technically describe the shape it should present. Nevertheless, in order to respect the constraints at the throat and at the fan face points, and also considering the fixed geometry parameters ahead the throat point, the behaviour of the curve along the intake length can be distorted and characterised by multiple inflections. This provides an example of what stated earlier. In fact, after a proper analysis of the whole design space, the conclusions can lead to the removal from the design space of the region characterised by this behaviour, specifically because it is mainly characterised by practical non-feasible geometries. After the exploration carried out with the first FF, the bounds were not found and the design space had to be expanded adding new levels to the original hypercube. The study took several steps, until a final six levels FF DoE was created. This was composed by 216 geometries, but after the different steps taken to reach this design space some of them were filtered due to the problem of multiple inflections, reaching a total of 150 geometries. In Figure 4.1 the contour plots for IPR and DC_{60} are presented. These plots are useful to provide a preliminary idea of the behaviour of the metrics across the design space, but care must be paid since they have not to be intended as response surfaces. They are the result of a simple linear interpolation of the values in the dataset and not the result of more complex RSM method such as Kriging. Nevertheless, from the figure below is possible to extract the information needed to define the bounds of the design space that will be used for RSM purpose. The values found on the different subplots in the figure are obtained from the non-dimensionalisation of the original values against the final values established.

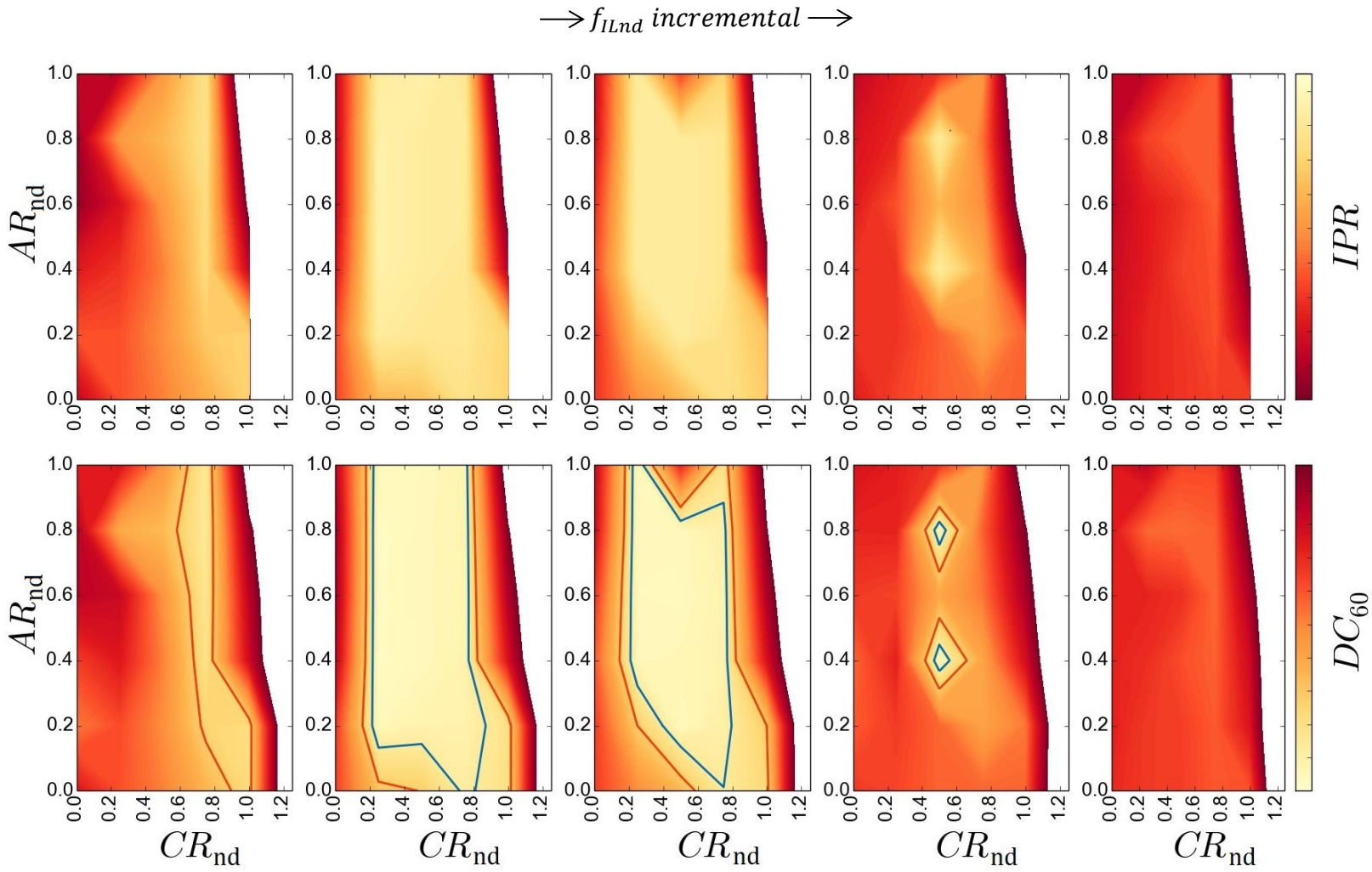


Figure 4.1: Design space exploration for bounds determination

In the figure reported, the distributions of the metrics of interest are given at different levels of a selected independent variable, which in this case is represented by f_{ILnd}. Moreover, the figure above is mainly divided in two parts:

- On the upper row the distribution of IPR for different slices of the 6x6x6 hypercube is presented. In order to ease the representation, it has been decided to report only five slices along the direction represented by f_{ILnd}.
- On the bottom row the distribution of DC_{60} is presented. The red and blue lines within the plots represent two critical values iso-lines.

The determination of the bounds of the design space is based on the analysis of the plots of the bottom row, since it has been assumed that, or a preliminary analysis, the distortion coefficient DC_{60} is the most crucial parameter to analyse and control when the performance of the intake bottom line at incidence is studied. The bounds are determined based on a limit value of this coefficient equal to $DC_{60} = 0.2$, which has been defined as the admissible limit under which the geometries can be considered aerodynamically satisfactory [22]. Therefore, the analysis of the bounds focuses on the area described by the red iso-line in Figure 4.1. As can be inferred from the fourth subplot in the bottom row, the number of geometries related to a distortion coefficient within the limit established is narrowing. For this reason it has been decided to fix the upper bound for the f_{ILnd} slightly above the value it has in the considered slice. The same is valid for the determination of the lower bound of this variable, as can inferred from the first subplot, where the number of feasible geometries is drastically reduced.

The other bounds for the remaining two independent variables can be easily extracted from the figure above. Nevertheless, difficulties have been encountered in the determination of the upper limit of the aspect ratio. In order to determine it in the most efficient and quick way, it has been opted for a random selection of samples located in different slices for values of AR_{nd} above those represented in Figure 4.1. in this way the computational effort required by the generation and simulation of 36 new geometries, corresponding to a new slice of AR_{nd}, has been avoided. Though, it is highlighted that this is not the correct way to proceed, but it

was the quickest given the limited amount of time available. The outcome of the exploration is plotted in Figure 4.2. The post-processing of these results can highlight a kind of outlier in correspondence of the slice defined by the higher value of AR_{nd} in Figure 4.1. Nevertheless, it is shown that increasing the aspect ratio above the limit already examined lead to higher values of DC_{60}. For this reason, the previous limit has been selected as the upper bound of the design space for AR_{nd}, since no improvement of the convex hull shape can be derived increasing its value.

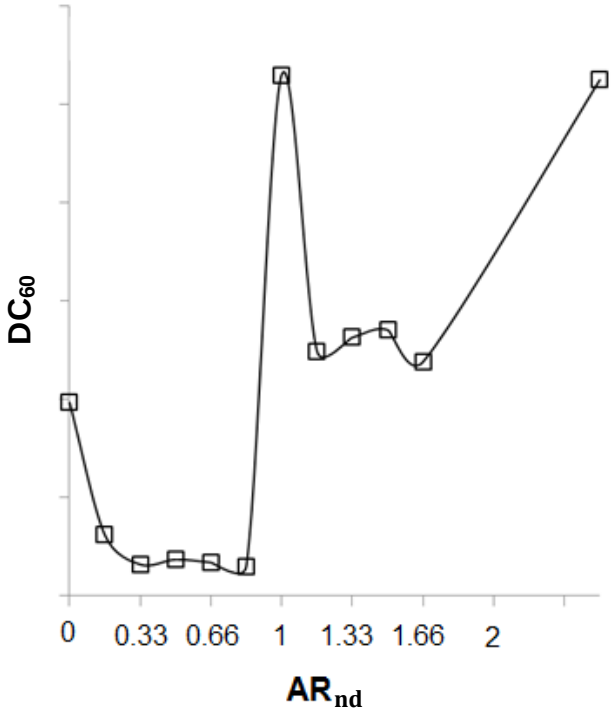


Figure 4.2: Upper bound determination for AR_{nd}

4.2 Design space exploration

An introduction to the preliminary work done on the design space obtained after the LHS DoE is given. As discussed in the introduction of Paragraph 3.3, the different samples which populate the design space are geometrically defined by three independent variables, namely CR, AR and f_{IL}, and two derived independent variables, namely κ_{TH} and κ_{FAN}. In the discussion presented in this chapter, a

slightly different set of independent variables is considered. In place of f_{IL} the main independent variable which will be used is r_{IL}, which represents the radius of curvature of the intake profile at the highlight point. r_{IL} is characterised by a better "physical" meaning compared to the coefficient used for the design space creation. Nevertheless, the use of f_{IL} for this purpose has been certified to be feasible, but for RSM purposes it is more indicated to use variables with a physical meaning, which can be easily defined from a designer point of view. The set of geometrical parameters considered is expressed in the non-dimensional form after being normalised in the unit cube. The metric of interest used to evaluate the performance of the bottom line at incidence are IPR and DC_{60}. In the current report, the latter one will be used predominately to analyse the performance of the bottom line at incidence, both because it has been demonstrated that optimising the geometry for this metric leads to better results and because there is a linear relationship between IPR and DC_{60}, which preliminary allows to focus on just one single aerodynamics performance parameter.

4.2.1 LHS result

Following the discussion of Paragraph 3.3.2, an automatic Python script was responsible for the creation, mesh and simulation of each one of the 625 intake design initially planned. After the creation of each geometry the shape was analysed to check the number of inflection points along the intake length. Each design was automatically examined and, in case more than one inflection point was detected, it did not pass to the following phase of meshing and simulation. In order to facilitate the post-processing and delete all the undesired geometries from the dataset, they were marked with a DC_{60}value equal to 1. As stated in Paragraph 3.3.2, it was expected that the number of geometries available after the creation of the LHS DoE was around 243, number ideally correspondent to a FF composed by five variables and three levels. Eventually, the Python script detected 285 geometries with multiple inflection points, reducing the number of samples available from 625 to 340, yet increasing the number of expected designs of almost 100 cases. After a preliminary inspection of the design space through scatterplots and three-dimensional visualisation, the objective of creation of a space-filled

design space can be considered achieved, thus providing a good insight in the behaviour of the dataset and, especially, in the sensitivity of the metrics of interest to variations in the design parameters. Though, in some situation a high amount of samples can be seen as a drawback since, if a highly non-linear behaviour of the measured parameters is found, the applicability of RSM techniques becomes difficult and alternatives, advanced solutions and trade-offs have to be found.

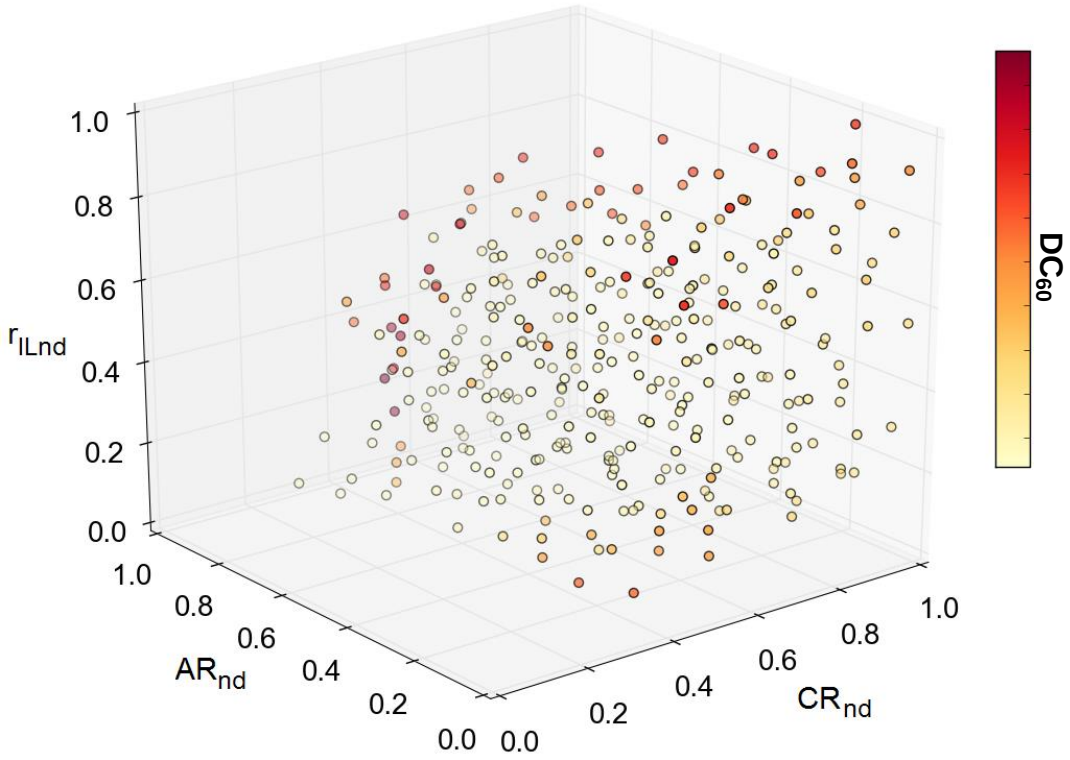


Figure 4.3: 3D visual representation of the design space

In Figure 4.3 the three-dimensional visualisation of the design space obtained is shown. Each dot represents a geometry, defined by the three main independent variables, namely CR_{nd}, AR_{nd} and $r_{IL\mathrm{nd}}$, and the different colour associated to each one is related to the calculated value of DC_{60}. As it will become clearer with more precise representations, at a first examination it can be noted that the design space is not perfectly contained in a regular structure, but instead the shape of the convex hull is pretty irregular. Moreover, the preliminary analysis reveals that the geometries characterised by a low value of CR are no longer present in the design

space, and this is principally due to the fact that a correlation between the low value of such design variable and the non-desired formation of multiple inflections in the intake is found. Due to the linear relationship between IPR and DC_{60}, the three-dimensional scatterplot describing the distribution of the former has a similar distribution as the one of the latter. Therefore, it has been considered of poor meaning to report the plot for IPR in the current discussion.

4.2.2 Geometric correlations

Following the mention of the sensitivity of the metrics of interest to the variations of the considered design parameters, one way to graphically analyse this is through Parallel Coordinates (PC) plots. This graphical aid is a very useful way for visualise and analyse multivariate data and it is mainly employed in this research to help to identify certain trends in the dataset. In fact, PC plots allow to graphically detect patterns, but, especially with high dimension design spaces, it requires few tests to find the most suitable combination of design parameters which can highlight them. One important pre-requisite of this kind of plot is to have data normalised in the unit cube, since PC is based on the linear combination of consecutive pairs of variables [35]. Therefore, the scaling of the data can also be viewed as an advantage in terms of improvement of the quality and significance of the PC outputs. In the PC plot each vertical line identifies one single design parameter and it represents the non-dimensional interval [0,1] in which this lies. Each geometry is then represented by a polyline which connects the values of its design parameters on the vertical axes.

In the PC plot of Figure 4.4 the 340 considered design variables are drawn and each geometry is coloured depending on its associated value of DC_{60}. On the horizontal axis it is also possible to see which combination of the design variables has been considered as the most meaningful for the current analysis in order to highlight potential trends. Though, as can be inferred, no particular patterns can be extracted from the analysis. One point that is worth noticing regards the clustering of some of the high DC_{60}cases in a small interval of values of κ_{TH}. In other words, it appears that geometries which lead to high values of $DC_{60}$are also characterised by

a certain narrow interval of curvatures at the throat. This issue raised should be investigated in an enhanced analysis which includes also κ_{TH} as main independent variable.

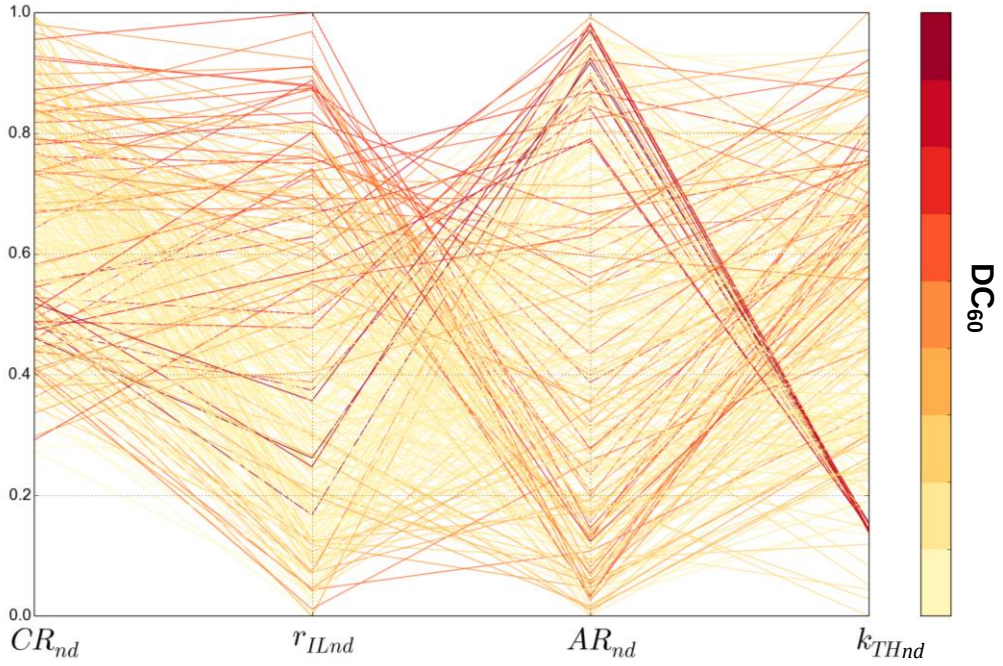


Figure 4.4: Parallel coordinates for distortion coefficient distribution

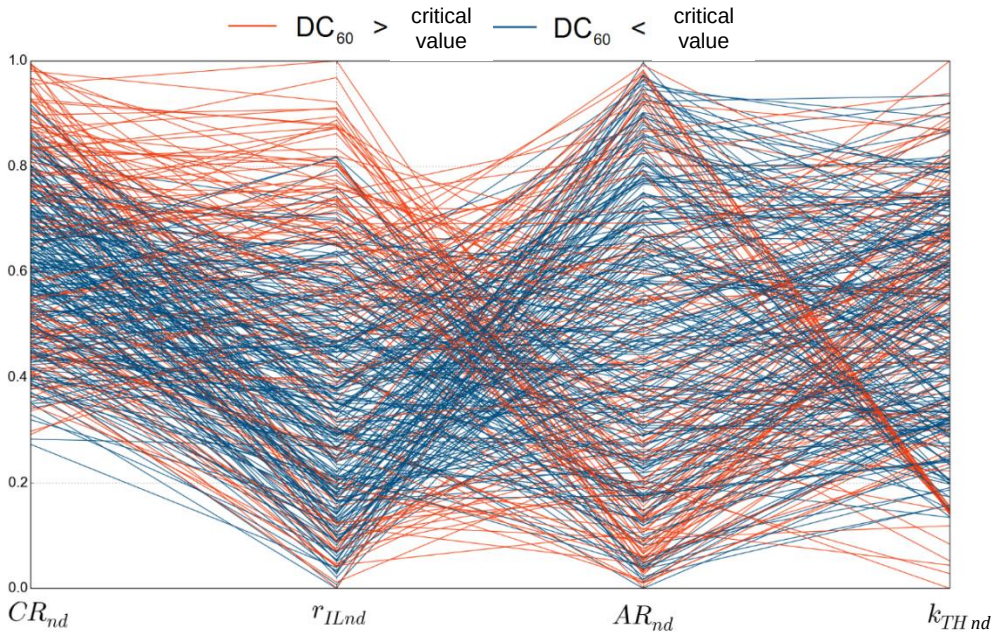


Figure 4.5: Parallel coordinates coloured for RoI

In Figure 4.5 the study based on PC plots is enhanced classifying geometries by colour, according to their associated value of the distortion coefficient metric. Red polylines mean that the geometries are associated to values of the metric above a determined critical value, below which is defined the Region of Interest (RoI). In the previous figure, based on the colour map describing the entity of DC_{60}, it was not possible to extract specific trends between the design parameters and the metric considered. Yet, the thorough analysis of the figure above allows to draw some interesting conclusions. Focusing on the two design parameters on the left of the PC plot, namely the non-dimensional variables $CR_{nd}$and r_{ILnd}, it is possible to notice that in the upper region of both intervals there is the presence of designs associated only to undesired values of DC_{60}. As already stated, it is not possible to remove samples from a dataset only because they possess a value of the metric of interest higher than the desired output. Nevertheless, in this case, since these regions of high $CR_{nd}$and high r_{ILnd} have been thoroughly explored through different simulations, it is possible to remove the samples from the dataset, leading to the reshape of the convex hull and to the re-definition of the design space bounds.

The same conclusions can be obtained also through the analysis of the scatterplots which define the relationships between the different pairs of variables. In Figure 4.6 it is shown one of the scatterplot obtained, whereas in Appendix A.1.1 is possible to find the whole set of correlations studied. The preliminary visual inspection of the scatterplot reported leads to the same results derived from the analysis of the PC plot for the same set of two variables. In the scatterplot the presence of regions in which the geometries are characterised solely by a value of DC_{60} higher than that researched for the bottom line design is made clear. Following the visual analysis, the geometries in these regions are extracted and filtered out from the dataset. In addition to the filtering based on CR_{nd} and r_{ILnd}, a further investigation showed that also f_{ILnd}, related tor_{ILnd} by Equation (2.9), allowed to filter some of the geometries due to a region characterised by the same behaviour highlighted above. The final amount of samples which were possible to remove from the dataset was equal to 29 total cases. Therefore, the filtering based

on the refinement of the design space bounds through the analysis of the correlations between the different variables allowed to pass from an initial dataset composed by 340 cases to a dataset of 311 total cases.

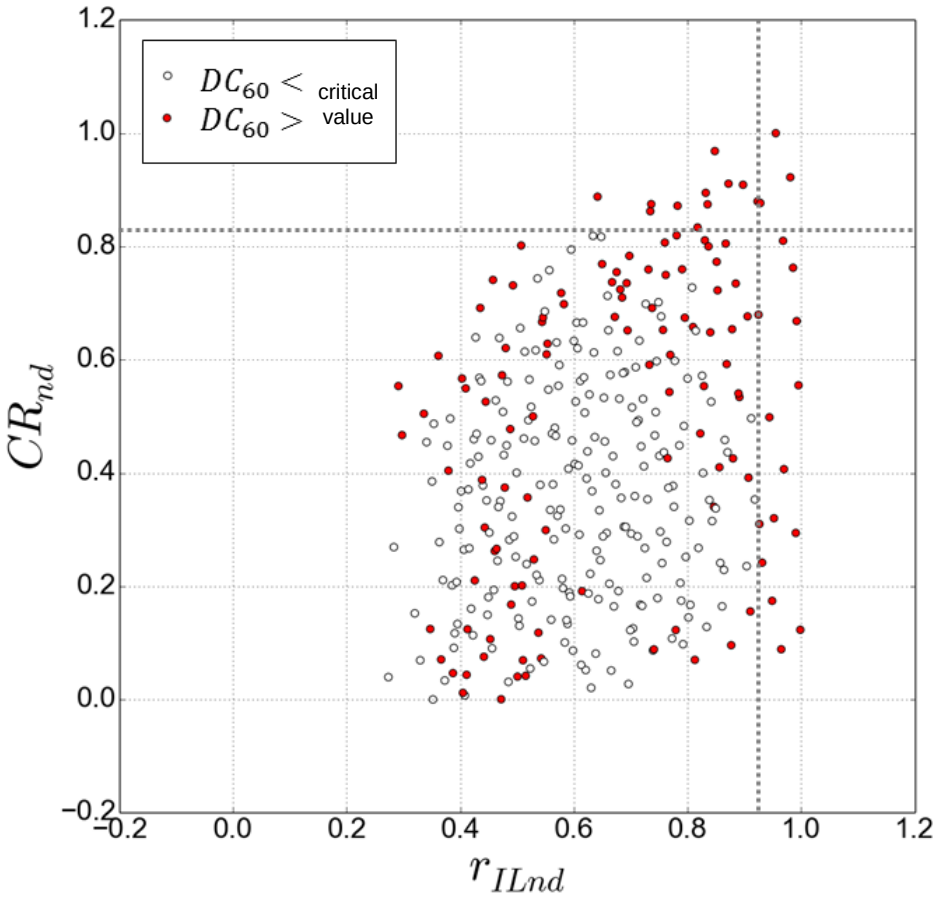


Figure 4.6: Geometric correlations scatterplot

The new bounds found are given in Table 4.1, and they constitute the new design space bound of the filtered dataset.

Table 4.1: Design space bounds refinement

CR_{nd}	0.9185
r_{ILnd}	0.9697
f_{ILnd}	0.8188

4.3 Lip design analysis

One of the main requirements to meet in the design of the intake lip is the maximum incidence angle capability at which separation occurs and leads to reduced performance of the whole engine [36]. Therefore, one point to ensure in the analysis at incidence conditions is the effective design of the intake lip in order to avoid separation inside the intake duct. For this reason, one procedure which, in the context of the current analysis, can give a good insight about the aerodynamics of the bottom lip of a short intake is the thorough study of the shape of the geometries created. In order to facilitate the analysis, the shape of each profile has been analysed separately for lip and diffuser design. After a preliminary study, it is possible to state that the shape of the diffuser does not tell much about the aerodynamic behaviour of the geometries, especially those which lead to high values of distortion at the fan face. Though, for further analysis, the profile of the diffuser can play an important role in the reduction or growth of the boundary layer resulting from the diffusion-induced separation. In turn, the aerodynamic behaviour of the different designs becomes very clear examining the shape of the lips. For this reason, the study reported in this section focuses solely on the examination of this part of the profiles, which extend from the highlight point to the throat point of the intake.

The two geometrical parameters used to analyse the profiles are the radius of curvature, identified by r, and the rate of change of the radius of curvature, identified by the acronym ROCOR, which corresponds to the first derivative of r in each point on the intake profile. All the geometries created are characterised by different combinations of values of the design parameters, especially regarding AR, for the purposes of this analysis. Therefore, the position of the throat in the relative intake reference system is different for each geometry. For this reason, the non-dimensionalisation of the intake length is necessary to ease the comparison between the geometries. In order to do that, the position of a point on a generic profile projected on the horizontal axis has been divided by the length of the lip of the same profile, and it is identified by x/L_{LIP}. Another important operation to

improve the comparison process is to non-dimensionalise the radius of curvature, and this is done dividing it by the radius at the highlight R_{HI}, which is common and fixed for all the geometries.

In Figure 4.7 and Figure 4.8 the distribution of the non-dimensional radius of curvature and non-dimensional ROCOR across the non-dimensional lip length are shown for all the 340 geometries in the dataset. Each design is coloured based on the fact that its related value of $DC_{60}$is greater or lower than a determined value of interest. As an aid for the following analysis, the twofold coloration is useful to have a quick understanding of which designs are going to be considered and filtered.

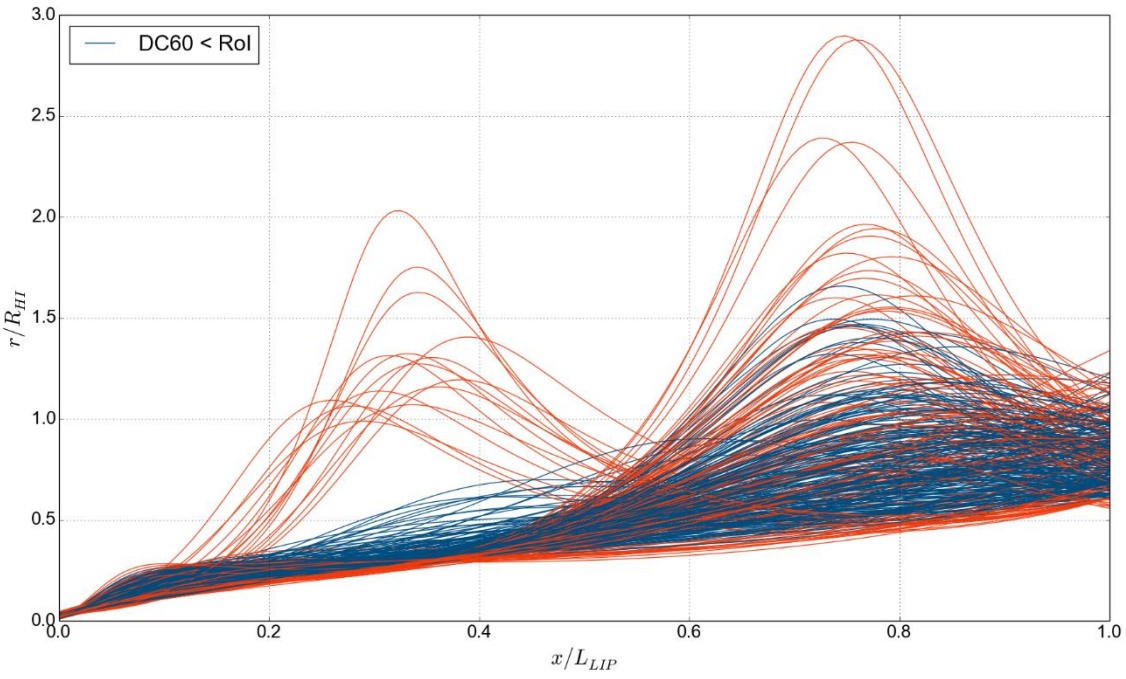


Figure 4.7:Non-dimensional radius of curvature distribution on lip profile

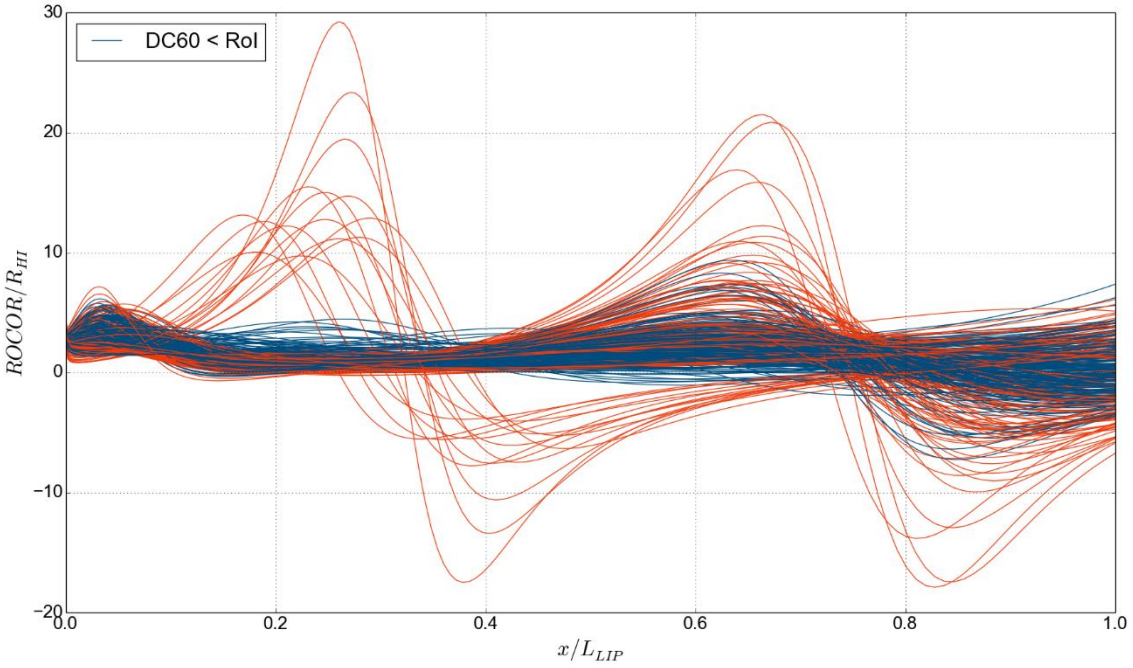


Figure 4.8: Non-dimensional ROCOR distribution on lip profile

Red-coloured geometries are related to a DC_{60}value higher than the range of interest, in accordance with what initially stated, and therefore, since their design is not desirable, it is possible to filter them out from the dataset. As mentioned in the introduction of Paragraph 3.3, the ideal objective of the intake design process for the bottom line at incidence conditions is to design a lip profile with a linear lip curvature distribution [11]. This is required in order to avoid the rapid acceleration and associated shock-induced separation which can occur with both rapid reduction and increase of curvature near the highlight region. In fact, controlling the lip curvature distribution allows to limit the strength of the shock resulting from the acceleration and possibly avoid the separation [13]. The research reported in this paragraph focuses therefore in the selection of the optimal design which allows avoiding the phenomena described earlier. As can be inferred from the figures, and described in detail later, the twofold coloration generally highlight the fact that geometries which do not respect what stated are also characterised by a related high value of distortion. Blue-coloured geometries, all related to a value of DC_{60} comprised in the range of interest, can be preliminarily assumed acceptable for intake lip design. They are in fact

represented by a smooth increase of radius of curvature near the highlight. Nevertheless, there are some exceptions between these geometries, likely related to the design of the diffuser, which can easily lead to distortion at the fan face.

In order to improve the outcome of the analysis, the lip profile has been divided in fore-lip, which has been defined as the first half of the non-dimensional lip length, and in aft-lip, defined as the second half of the lip. This separation between the two parts of the lip can be easily assumed after a preliminary analysis of both figures, where it is clear that some geometries are characterised by an increase of radius of curvature in the first part of the lip, near the highlight point, while others are characterised by a marked increase in the region near the throat point.

4.3.1 Fore-lip design analysis

The initial focus of the analysis was placed on the fore-lip design. Based on the data available after the DoE, the geometries corresponding to the red-coloured distributions in Figure 4.7 were extracted and analysed. They are reported in Table 4.2, where each one of these geometries is represented by the maximum value of non-dimensional radius of curvature and the corresponding position along the non-dimensional lip length, as well as by the associated value of distortion at the fan face. The highest value of the latter is equal to $DC_{60} = 0.587$. In Appendix A.2 the locations of the geometries considered in the fore-lip design analysis are identified in the design space.

Table 4.2: Fore-lip analysed designs

$\mathbf{Max\ r/R_{HI}}$	$\mathbf{Position\ max\ r/R_{HI}}$	$\boldsymbol{DC_{60}}$
1.195	0.378	0.504
1.065	0.276	0.249
1.752	0.336	0.334
1.324	0.325	0.538
1.159	0.354	0.472
1.406	0.384	0.466
1.139	0.297	0.569
2.032	0.319	0.352
1.093	0.255	0.233
0.990	0.286	0.294
1.315	0.308	0.587
1.627	0.336	0.438
1.071	0.331	0.512
1.305	0.348	0.486
1.274	0.336	0.543

Along with the extraction of these considered “bad” geometries, a design space exploration was also exploited. In Figure 4.9, the design which leads to the worst value of distortion is shown. In order to give a clearer idea of what will be discussed, the design reported corresponds to the fifth profile from the top in Figure 4.7, considering just the fore-lip. The axial component of the velocity is represented in the plot below, useful to highlight regions characterised by reverse flow. From the figure it is easy to identify a long flat profile in the lip design. This is related to the rapid increase of the radius of curvature near the highlight which, as can be inferred, easily trigger the separation of the air stream from the surface. Generally, as a rule of thumb for a good intake lip design, the increase of radius of curvature should be linear and the profile obtained results well-rounded and suitable to be employed in the flight condition considered.

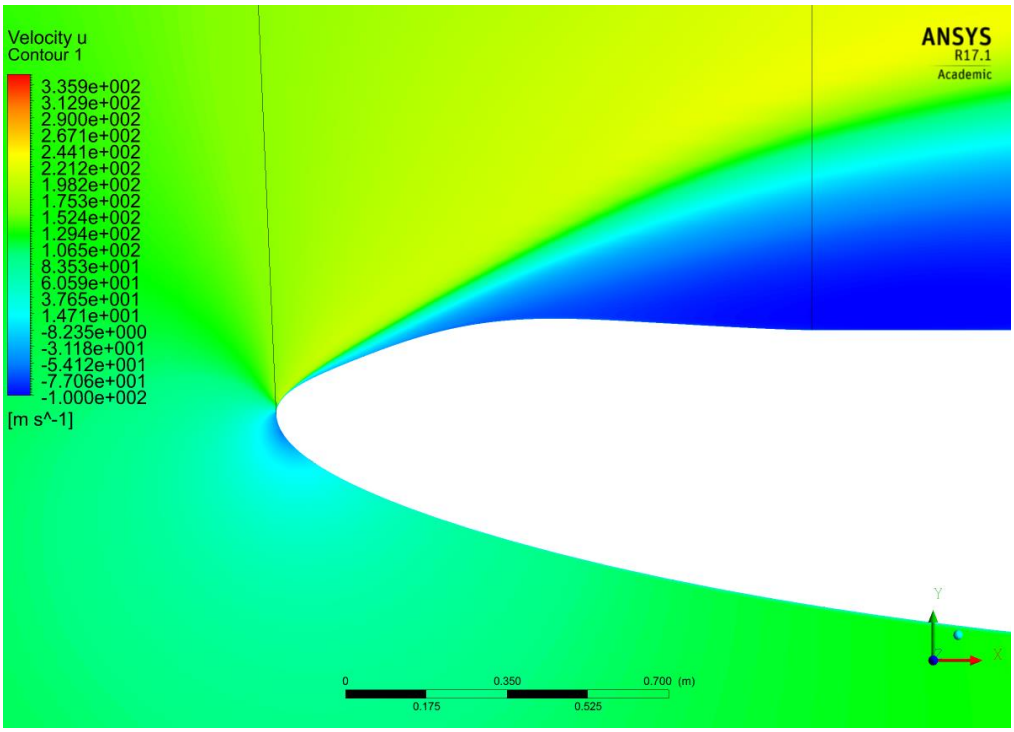


Figure 4.9: Fore-lip analysis - worst design – axial velocity

4.3.2 Aft-lip design analysis

Analysing the radius of curvature distribution on the aft-lip, it is possible to see that, for few geometries, the increase of r is more marked than it was in the fore-lip. As reported earlier, the geometries characterised by rapid increases of r lead to flat regions along the lip profile. As reported in Table 4.3, from the analysis of the aft-lip design it is possible to extract 11 geometries which do not respect the requirements expressed in the beginning. Each design extracted is characterised by both a rapid increase of non-dimensional radius of curvature near the throat point and by a related high value of DC_{60}. Eventually, the highest value found for the distortion parameter is equal to $DC_{60} = 0.425$. It corresponds, on the second half of the lip, to the second profile from the top drawn in Figure 4.7, specifically identified by a maximum non-dimensional radius of curvature equal to $r = 2.877$.

Table 4.3: Aft-lip analysed designs

$\mathbf{Max\ r/R_{HI}}$	$\mathbf{Position\ max\ r/R_{HI}}$	$\boldsymbol{DC_{60}}$
1.697	0.757	0.360
1.965	0.757	0.415
1.943	0.766	0.394
2.877	0.748	0.425
1.735	0.766	0.330
2.371	0.748	0.348
1.906	0.766	0.407
2.898	0.740	0.291
1.804	0.783	0.339
2.391	0.722	0.180
1.822	0.740	0.311

In Figure 4.10 it is possible to have a better idea of what discussed previously regarding the requirement of a smooth increase of radius of curvature while designing an intake lip. Considering the “bad” profiles for the aft-lip design in

Figure 4.7, if only the distribution of r until the end of the fore-lip is considered, the geometry reported in the figure below presents exactly the design which is desired to achieve. Though, after the fore-lip, as can be inferred from Figure 4.7, the radius of curvature markedly increases and it translates in the flat region prior the throat point, as it is clearly visible in Figure 4.10. As highlighted in the isentropic Mach number distribution in Figure 4.11, the sudden change in profile causes a shock formation upstream the flat region, which triggers the shock-induced separation of the air stream from the duct surface.

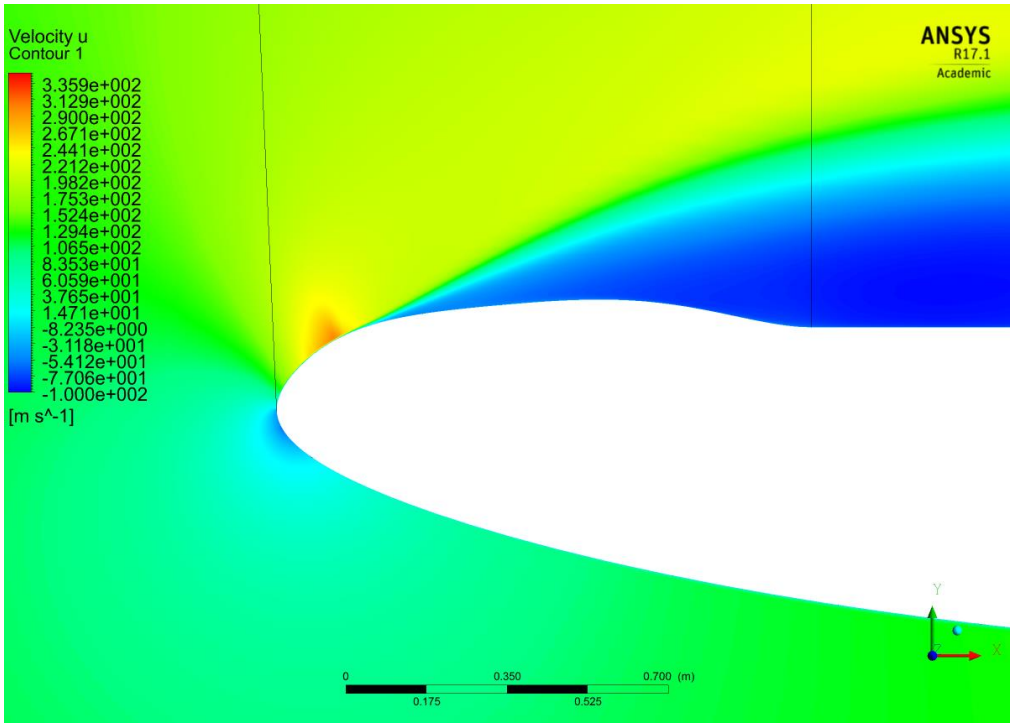


Figure 4.10: Aft-lip analysis - worst design – axial velocity

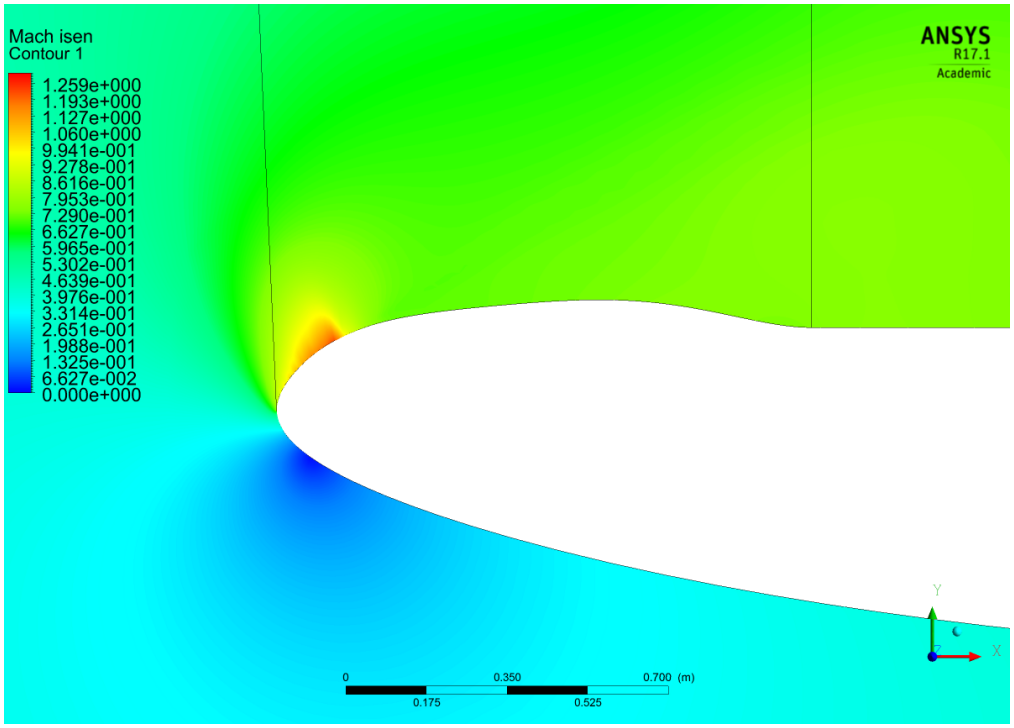


Figure 4.11: Aft-lip analysis - worst design – Isentropic Mach number

4.3.3 Filtering based on lip design analysis

The work carried out to analyse the intake lip design showed that some of the geometries generated are not in accordance with the objectives of the current research. It can be inferred that geometries characterised by a non-monotonic increase of radius of curvature along the lip length generally lead to distortion at the fan face. As a rule of thumb, usually it is not correct to filter samples from the initial population of a design space, but this should be avoided just if the filtering is based on undesired values of the metric evaluated. Though, in this situation, since the geometries analysed in the previous paragraphs are not suitable for intake design, they can be removed from the dataset. Therefore the filtering is based on geometrical undesired features, and not simply on undesired metrics, which is more acceptable in the context of creation of a response surface which aims to generalise the behaviour of the metrics of interest.

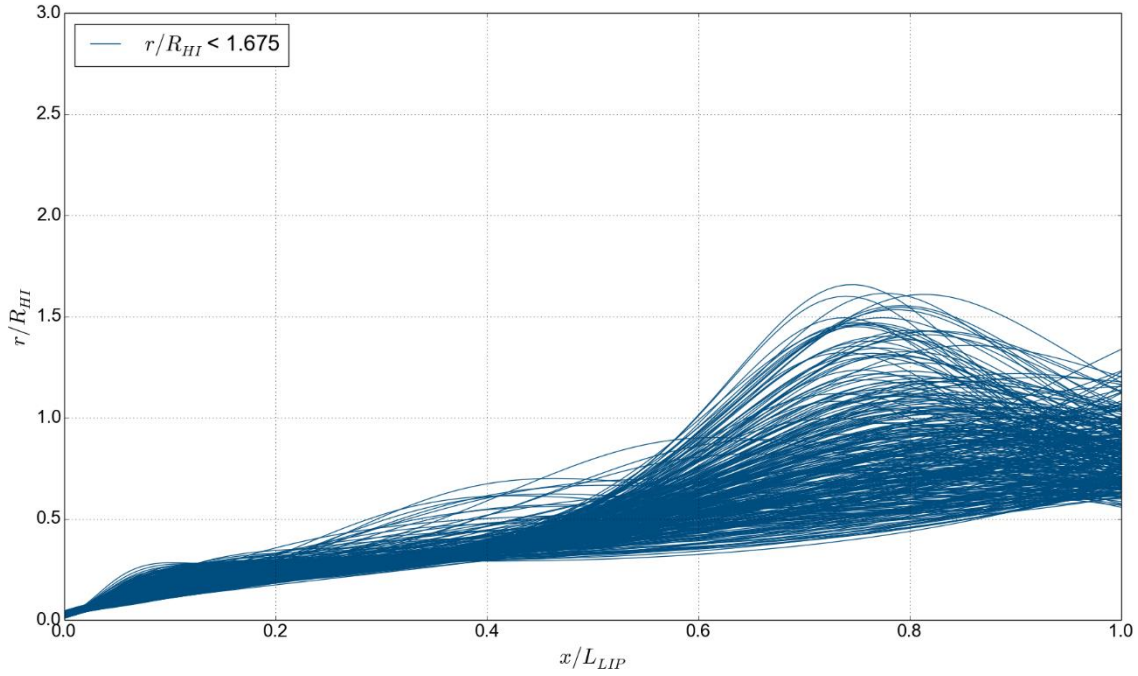


Figure 4.12: Filtered non-dimensional radius of curvature distribution

In Figure 4.12 the non-dimensional radius of curvature distribution on the lip is presented. As a general remark, it is possible to assert that the profiles of the fore-lip respect the requirement of smooth and monotonic increase of radius of curvature, while the same is not valid for the aft-lip. For this part of the lip it is clearly visible that there is a group of geometries designed with a non-linear increase of r. Nevertheless, they are characterised by a low value of distortion which ensure a correct performance of the engine, and therefore, even if the shape is not the one desired, they can be preliminarily considered acceptable.

4.4 Surrogate modelling

The results presented in this paragraph represent the key point of the entire work, aimed to obtain a surrogate model to quickly evaluate the optimal design of the bottom line of a short intake at incidence conditions. A thorough analysis of all the results obtained from the validations of the different models tested is reported. Prior to present the work carried out to obtain the metamodels, it is paramount to assert that the behaviour of the initial dataset created is highly non-linear, highlighting the difficulties in the design of a short intake. A lot of effort has to be put in order to improve the design space, for instance through filtering or other kind of conclusions, and obtain a surrogate model as more general as possible. Nevertheless, the results obtained are encouraging, but still not optimal for use in real world applications. In the following analysis all the issues and problems solved are analysed, as foundation for future improvement.

4.4.1 Kriging model

The creation of a reduced order model is just the tip of the iceberg of the whole process, since it is not possible to employ the model until the validation process confirms its reliability. In order to assess it, cross-validation is performed every time a model is created, even if it is usually a very computational expensive procedure. This is paramount since this method helps to make an informed decision on whether the model obtained is reliable of not for the exploitation in a real world application, and usually this decision is based on statistics parameters which serve as diagnostics. In the context of this work, the cross-validation of the different models created is based on the concept of Leave-One-Out (LOO), in which each sample participates to the assessment of the quality. The basic idea behind this technique consists of two steps: initially, a surrogate model is created without one sample, removed temporarily from the dataset; subsequently, the value of the metric in the position of the sample left-out is obtained from the metamodel and it is compared against the actual value of the metric in the same location. These steps

are repeated for all the points in the design space, and the outcome gives useful information about the quality of the obtained model.

The two summary statistics used to assess the quality are the Root Mean Squared Error (RMSE) and the Pearson r coefficient [37]. The former is defined as in Equation (4.1), and it is commonly employed when the relationship between predictions and real values is subjected to investigation. From a statistical point of view, it can be considered as the sample standard deviation of the differences between the predicted metrics $\hat{y}^{(i)}$ and the real metrics $y^{(i)}$. From a practical point of view, it can be viewed as a measure of how well the model predicts the real values. The main advantage for which the RMSE is evaluated consists in the fact that it is expressed in the same units as the data. Therefore, it is easier to understand the entity of the error that characterises the metamodel in correspondence of the known locations in the design space, and it gives an idea of the level of error which can be encountered when predictions are made outside these locations. Considering Equation (4.1), since the cross-validation is carried out within the design space bounds and especially just using the dataset obtained, the term between brackets is well-known as "residual". Attention should be paid on the use of different terms, because, for instance, this term is different from the "prediction error", which generally represents the error for a prediction made outside of the known values [37].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{4.1}$$

The Pearson r coefficient is a common representation of the correlation between two different variables, which in the context of this work are represented by the two sets of predicted and real metrics.

For the scope of this research, the ideal values that the two statistics considered should have are:

- RMSE $\cong$ 0, which indicates that the model perfectly predicts the observations $y^{(i)}$ in the known locations obtained from the DoE.

- r ≅ 1, which reflects the same concept stated in the previous point, and which highlight the fact that there is perfect correlation between predictions and real values. Though, the critical analysis of this term should be carefully considered along with the other considered statistics.

In addition to the summary statistics, which give a quantitative representation of the quality of the model, two plots are usually provided after the cross-validation to visually assess the quality of the metamodel created. The first one is the scatterplot of the predicted values plotted against the real values. Ideally, a 1:1 correspondence is desired, but usually it is more likely that the linear fitting between the two quantities measured is characterised by a slope lower than 45 degrees. Nevertheless, this is an known characteristic of the Kriging method, since generally it tends to under-predict large values and over-predict small values, within the specified dataset of interest [37]. The second plot considered the scatterplot of the residuals values plotted against the predicted values of the metric. Ideally, the points plotted on this plot should lie on the horizontal axis, corresponding to a residual value of zero. Nevertheless, since it is not expected that a generic prediction $y^{(i)}$ corresponds exactly to the real value in the same location, the different points are spread in the plot. In few occasions, the inspection of this plot allows to extract trends, which lead to the re-definition of the surrogate modelling technique that should be employed. In order to ease the readability of the following discussion, only the former plot described is reported, while the second one can be found in Appendix B for all the models considered.

The initial study, aimed to create a metamodel for intake design purposes, focused on the creation of a model based solely on the three main independent variables already defined in the description of the project and which are identified by CR, AR and r_{IL}. Nevertheless, as it will be shown, it has been found that a careful increase of the dimensionality of the problem allows to improve the quality of the metamodel created. In fact, increasing the dimensionality of the design space adding the design variable κ_{TH} has shown significant improvements in the surrogate model results. Nevertheless, it has to be reminded that this variable is a "derived" independent variable, and therefore its influence cannot be examined

outside the values automatically calculated during the geometry creation at the DoE stage. Even though this mathematical constraint is present, the implementation of this additional variable is still valid. Though, when the metamodel created from these four variables set will be queried, κ_{TH} should be derived in the same way it was derived in the initial DoE. In other words, while the three main variables will be selected within their determined ranges, κ_{TH} will be determined from the same least square approach described in Paragraph 2.4.3.

For sake of clarity, the discussion regarding the results commences analysing the outcomes obtained considering the non-filtered design space. This is mainly done for two reasons: the study started from the mentioned condition, and therefore it shows the initial point of the analysis which led to the thorough study carried out and described in the previous paragraphs; it serves as a meaningful measure of comparison with the results obtained from the subsequent filtered design space.

As demonstrated in Paragraph 3.4.2.2, the number of settings which control the creation of a surrogate model is high, and different configurations have to be tested to find the optimal settings. During the testing phase of the current study a large number of these have been tested and in the following sections the most significant results found for the most significant configurations are reported.

In the following discussion the results are analysed and compared with the statistics parameters described earlier. Along with them, one additional parameter employed, which gives a better understanding of the quality of a model in relationship with the specific considered dataset, is the ratio RMSE / $range$, in which the range is the simple difference between the maximum and minimum value of the metric considered. Even if the consideration of the other two parameters is paramount, as suggested by Forrester [16] the ratio provides an improved indication of quality. As it will be found in the following sections, the post-processing of the results focuses also on the analysis of the different statistics in two different ranges. These are mainly identified as the “overall” one, where the quality of the whole model is assessed, and the “RoI” one, which defines the region with geometries characterised by a value of DC_{60} below a critical value.

Concerning the three variables model approach, the results reported are divided into two parts: User-Defined Auto-Correlation Parameters (UDACP), in which the auto-correlation parameter θ has been fixed for all the different MLE evaluations in the cross-validation process; MLE-Defined Auto-Correlation Parameters, in which the auto-correlation parameter is calculated through the Maximum Likelihood Estimation at each evaluation of the LOO process. This allows also to understand the limits of the employed method since, as it will be demonstrated, the MLE does not provide feasible results and it requires further investigation to assess its potential.

In any case, especially following the realisation that the intake design space is highly complex, the results which will be obtained from a surrogate model obtained with the current method employed have to be post-processed considering also the results of the corresponding cross-validation.

4.4.1.1 Three variables – non-filtered dataset

USER-DEFINED AUTO-CORRELATION PARAMETERS (UDACP)

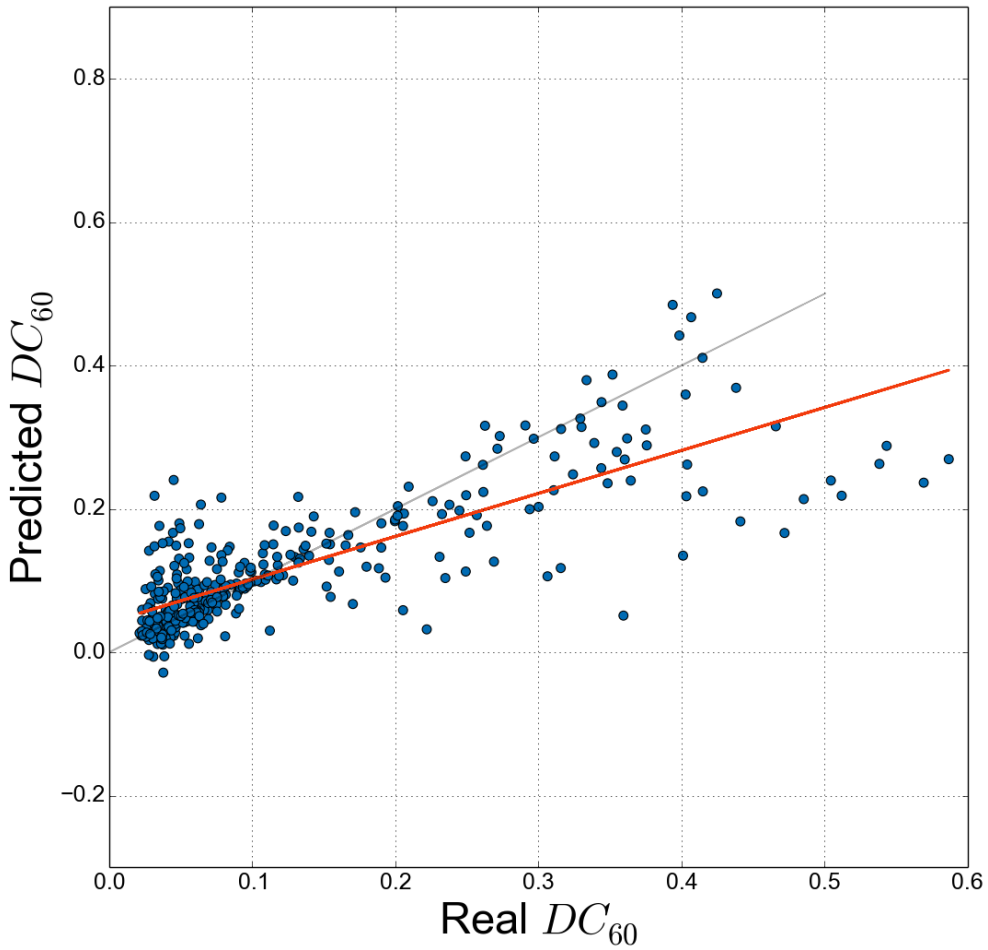


Figure 4.13: Predicted vs real - 3 variables model - non-filtered – UDACP

Table 4.4: 3 variables model - non-filtered - UDACP

Overall	RMSE	0.075
	RMSE / *range*	0.132
	Pearson′s r	0.801
RoI	RMSE	0.052
	RMSE / *range*	0.520

The first model analysed is based on the three main independent variables which have been defined earlier and, in order to avoid confusion, are identified by CR, AR and r_{IL}. As described at the end of the introductory paragraph, the analysis reported in this section focuses on models obtained through UDACP. It was found that optimal results for the three variables models can be found assigning a value of θ common to each feature considered, and this value has been set to 1. In Figure 4.13 the result of the cross-validation procedure is presented exploiting the main scatterplot described in the previous paragraph. As an aid to the visual inspection of this plot, Table 4.4 is provided and it contains all the summary statistics used to evaluate the quality of the model created. The grey line in the plot represents the ideal linear correspondence between predictions and actual values of the metric considered, where, ideally, all the points should lie. The red line is a basic linear regression of the data in the plot. As stated earlier, one of the summary statistics of major interest considered in this work is the Pearson r coefficient. Often to this parameter is also associated the slope of the red line, which gives a rough indication of how far the data are from the 1: 1 correspondence. Though, as it has always to be reminded in this kind of analysis, the slope does not give a proper indication of the quality of the model and it must always examined along with another more significant statistics parameter.

As it becomes clear from the visual analysis of the scatterplot, the distribution of the points in the plot "almost" finds agreement with what reported in the main introduction of this paragraph. In fact, it can be inferred that for low values of actual DC_{60} the prediction is over-estimated, while the region outside the range of interest is a general characterised by an under-estimation of the predictions of the metric considered.

In Table 4.4 the different summary statistics allow to have a better idea of the quality of the model. Concerning the quality of the predictions in the range of interest, the RMSE describes a model which is not acceptable. In fact, it is highlighted that the value given from the surrogate model for a generic query in a location within the design space bounds can lie in an interval of $DC_{60} \pm 0.052$.

MLE-DEFINED AUTO-CORRELATION PARAMETERS

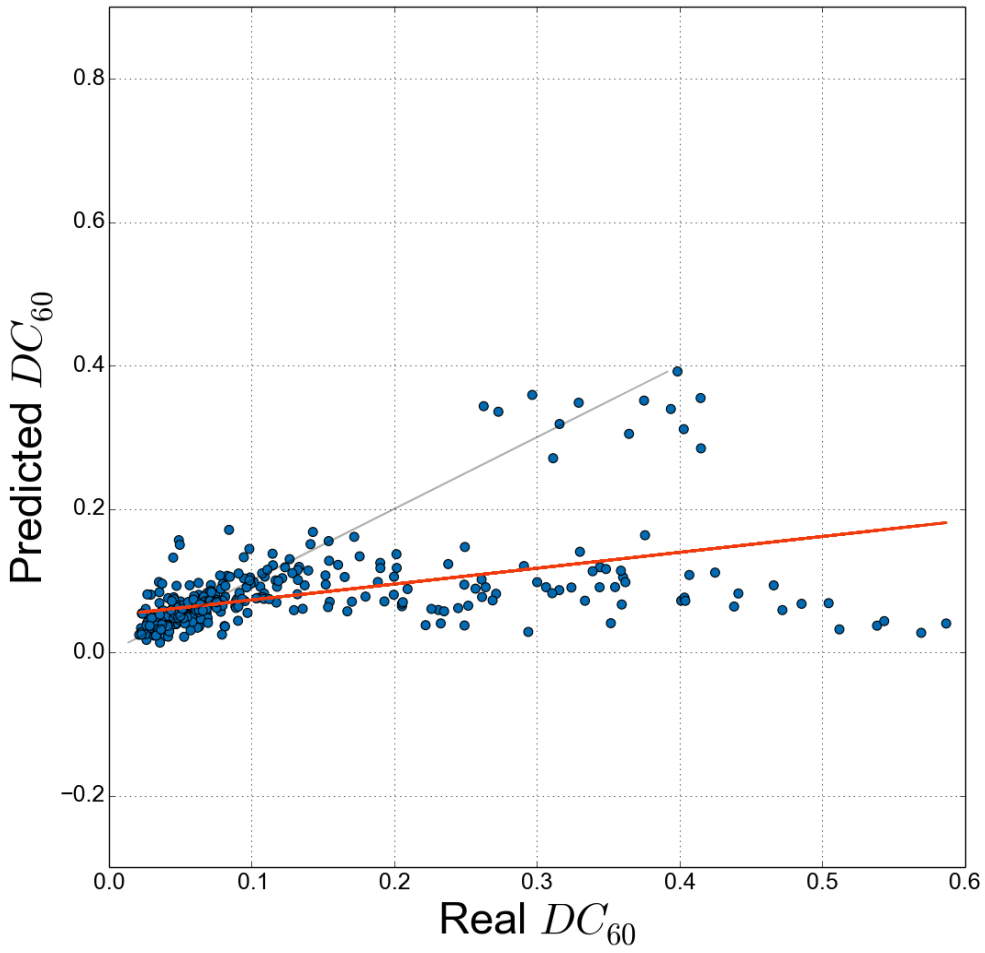


Figure 4.14: Predicted vs real - 3 variables model - non-filtered– MLE

Table 4.5: 3 variables model - non-filtered - MLE

Overall	RMSE	0.119
	RMSE / *range*	0.211
	Pearson's r	0.459
RoI	RMSE	0.029
	RMSE / *range*	0.290

In this section the results obtained from the cross-validation of a Kriging model in which the auto-correlation parameters are calculated through MLE are presented. As it is possible to understand examining Figure 4.14, the behaviour of the model changes when the parameters are not user-defined. In particular, except for few cases out of the total 340 designs, the surrogate is assumed to have a stiffer behaviour compared to the previous analysed case. This seemingly describes a quasi-flat response hypersurface, likely characterised by various spikes across it, related to the variability of the predicted values reported in the figure below.

A first comparison between the statistics in Table 4.4 and Table 4.5 shows a notable improvement in the quality within the region of interest, but a general worsening of the quality of the overall model. This is mainly related to what reported previously, since a stiff model is not able to well generalise the behaviour of a highly non-linear dataset. It approximates only the low values of the metric considered, since eventually it is not able to generalise the behaviour of potential increases of the metric in some zones. In any case, it is notable the reduction of the RMSE for the region characterised by low values of the metric of interest, but this does not still allow to consider the model as completely acceptable.

4.4.1.2 Three variables – filtered dataset

USER-DEFINED AUTO-CORRELATION PARAMETERS (UDACP)

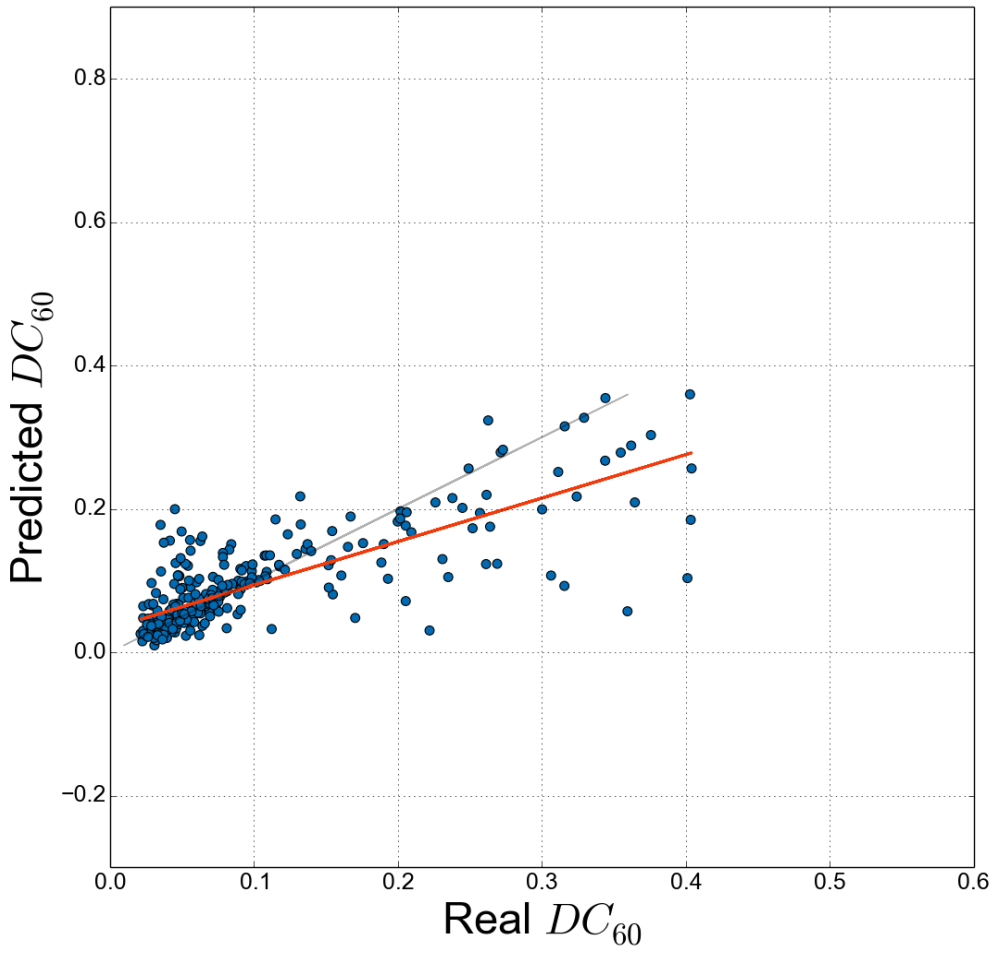


Figure 4.15: Predicted vs real - 3 variables model - filtered– UDACP

Table 4.6: 3 variables model - filtered - UDACP

Overall	RMSE	0.054
	RMSE / *range*	0.141
	Pearson's r	0.731
RoI	RMSE	0.036
	RMSE / *range*	0.360

In this section the main effects related to the filtering of some of the geometries out of the dataset are studied and the results are presented following the scheme of the previous section. The filtering considered is the one resultant after the initial design space exploration, as described in Paragraph 4.2, and after the lip design analysis, as described in Paragraph 4.3, which allowed to discard some of the geometries analysed during the DoE. Also in this case the UDACP approach employed the same common auto-correlation coefficient defined in the previous section, and set to 1. From the plot reported in Figure 4.15 it is not completely possible to understand the improvement achieved after the filtering, but important information are derived examining Table 4.6. According to the data obtained after the cross-validation, there is a general improvement of the whole metamodel quality, with an overall reduction of the error associated to the predictions. Nevertheless, the reduction of the Pearson correlation coefficient, in comparison with the non-filtered dataset case, denotes that the settings used and maintained constant in the two situations lead to a reduction of the predictive capabilities of the model. Therefore, further investigation on the variation of the settings from one model to the other has to be carried out in order to reduce the stiffness of the model created, which is not able to well approximate regions of high DC_{60}.

MLE-DEFINED AUTO-CORRELATION PARAMETERS

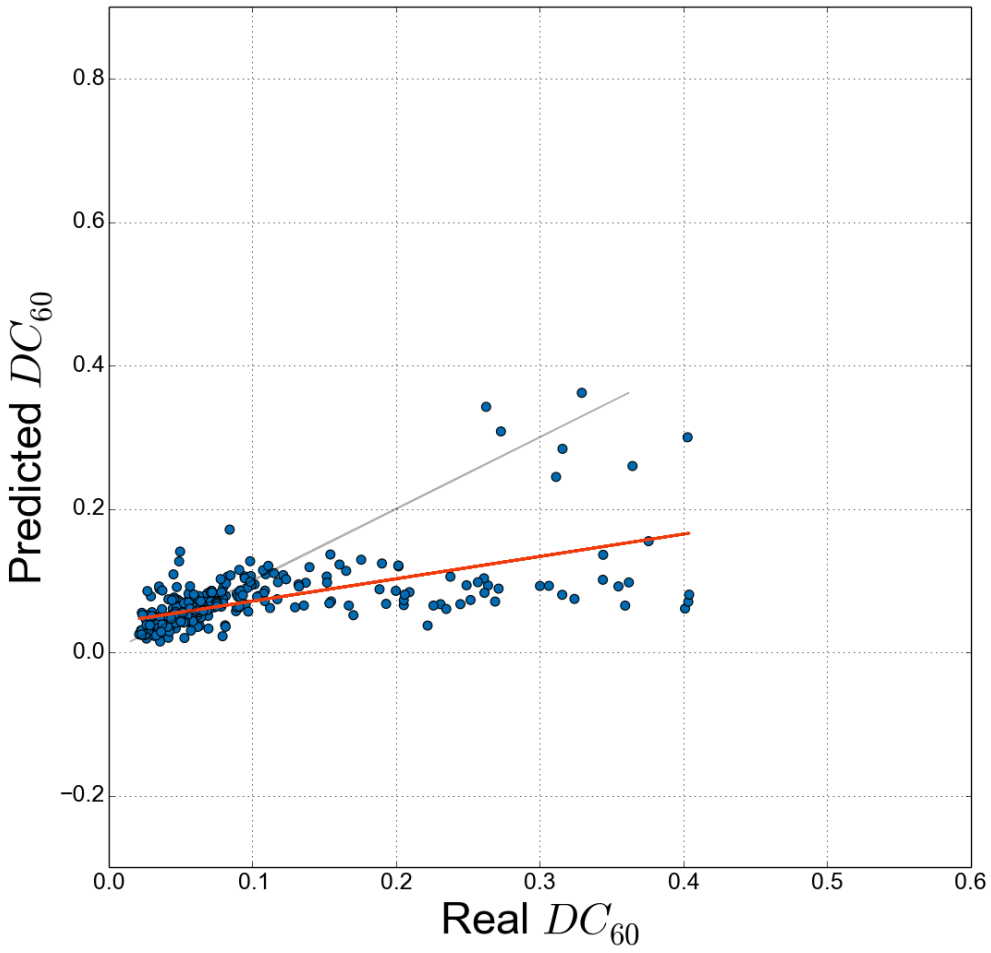


Figure 4.16: Predicted vs real - 3 variables model - filtered– MLE

Table 4.7: 3 variables model - filtered - MLE

Overall	RMSE	0.075
	RMSE / *range*	0.196
	Pearson's r	0.595
RoI	RMSE	0.024
	RMSE / *range*	0.240

The automatic estimation of the auto-correlation parameters has the same effect on the filtered model created as it had on the non-filtered case. This can be considered as a trivial conclusion, but it was expected that an improvement of the design space bounds and the filtering of some of the undesired geometries would eventually lead to an improved model, where the characteristic stiffness highlighted earlier was no longer present. Though, as it is demonstrated in Figure 4.16 and Table 4.7, this conclusion is not verified and, as suggested in the previous section, a modification of the different settings of the Kriging model should be carefully considered when the filtered case is employed.

4.4.1.3 Four variables – non-filtered dataset

MLE-DEFINED AUTO-CORRELATION PARAMETERS (UDACP)

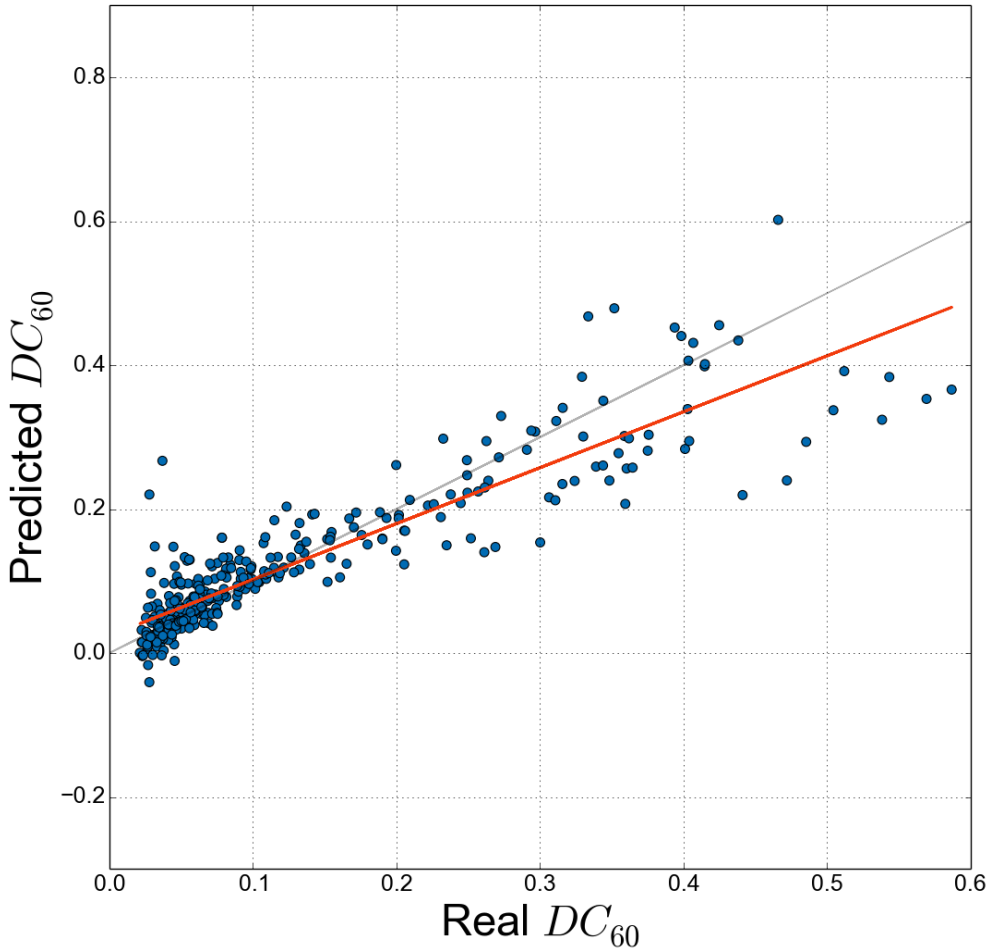


Figure 4.17: Predicted vs real - 4 variables model – non-filtered– MLE

Table 4.8: 4 variables model – non-filtered - MLE

Overall	RMSE	0.054
	RMSE / *range*	0.095
	Pearson′s r	0.901
RoI	RMSE	0.043
	RMSE / *range*	0.430

The discussion concerning the use of four variables instead of the standard main three starts from the analysis of the model created using the complete dataset without geometries filtered out. Moreover, for this increased dimensionality case, the focus is solely on the automatic calculation of the auto-correlation parameters through the MLE method, which is a key feature of the Kriging method and it is exploited to obtain improved results. From Figure 4.17 and Table 4.8 it is possible to clearly notice that there is a marked improvement in the quality of the metamodel created. Focusing the attention on the range of interest, the spread of the predictions in this region is less pronounced compared to the non-filtered dataset case based on three variables. Nevertheless, in this zone of the plot, characterised by $DC_{60} < RoI$ there is the unexpected presence of some samples which predicted values are above $DC_{60} > 0.2$. Even if the statistics parameters achieved a marked improvement, these outliers require a further investigation to understand the underlying reasons for such behaviour. Therefore, at this stage the model cannot be considered acceptable, but it allowed to reach a level of overall quality which was not possible to achieve in the smaller dimensionality case.

An enhanced analysis of the current non-filtered increased dimensionality metamodel has been carried out. Eventually, as in an optimisation routine, the achievement of the reported results has led to the necessity to exploit a thorough analysis of the design space in order to understand the underlying driving behaviour of the intake design dataset, which was represented by the one reported in the first paragraphs of this chapter. The initial investigation, which has led to the consequent thorough analysis, focused on the analysis of the points which can be identified as outliers in Figure 4.17. It has been found that this two points, which are situated externally the cloud of points in the region of interest, if they are analysed in a three-dimensional visualisation of the design space it is possible to see that they are located near a region of high values of DC_{60}, situated near the borders of the convex hull. Therefore, the possible explanation for such result is that the response hypersurface in this region tends primarily to approximate the high values of the metric, returning a value of the same magnitude when the surrogate is queried in the considered region. For this reason, the question which

arises is if it is possible to remove them or the neighbour samples from the dataset in order to improve the quality of the metamodel in the generalisation of the behaviour in this region. The answer, as already stated, lies in the first paragraphs of this chapter and in the final model reported in the next section.

4.4.1.4 Four variables – filtered dataset

MLE-DEFINED AUTO-CORRELATION PARAMETERS

The filtering of undesired geometries from the dataset for the four variables based model, as can be inferred from Table 4.9, allows to obtain the best level of quality achieved so far. As an example of the effect of the filtering, the two geometries considered in the previous section, characterised by an associated actual value well below $DC_{60} < RoI$, are now located in the cloud of points of the region of interest and this is due to the removal of some of the designs related to high DC_{60} which, in the previous case, constrained the response hypersurface in this region.

The current surrogate model is still considered scarcely acceptable, but few conclusions can be derived:

- The RMSE value compared with the range of the metric indicates that the overall error is slightly below 10%. According to Forrester [16], below this percentage the model can be considered a reasonable global model.
- The over-prediction of the different metric in the region above $DC_{60} > 0.2$ has been eliminated, and in the filtered case there is accordance with what stated in the beginning of this paragraph regarding under- and over-predictions.

From the Pearson's coefficient is also possible to derive the so-called coefficient of determination r^2. This is usually employed as statistic in linear least squares regression, but it can also be applied to different models with a slightly different meaning. In the context of this work, it represents a measure of how well a real metric can be constructed from the predicted observations [38]. Regarding the values that this statistic should have, Forrester [16] reported that for values of

$r^2 > 0.8$ the surrogate has good predictive capabilities. For the case under analysis this is equal to $r^2 = 0.794$, and therefore, along with the statement in the first bullet point, it can be assumed that the surrogate model created is a feasible foundation for future improvements.

According to the results reported in Table 4.9 for the RMSE in the region of interest, if the surrogate model is exploited in a real world application it is almost certain that it can well predict values of the metric included in the interval $0.028 < DC_{60} < 0.072$. In fact, it can be assumed that, since the RMSE of the data contained in the RoI is equal to 0.028, the real metric value lies in the interval $\hat{y} \pm 0.028$, with $\hat{y}$ value of the predicted metric obtained querying the surrogate.

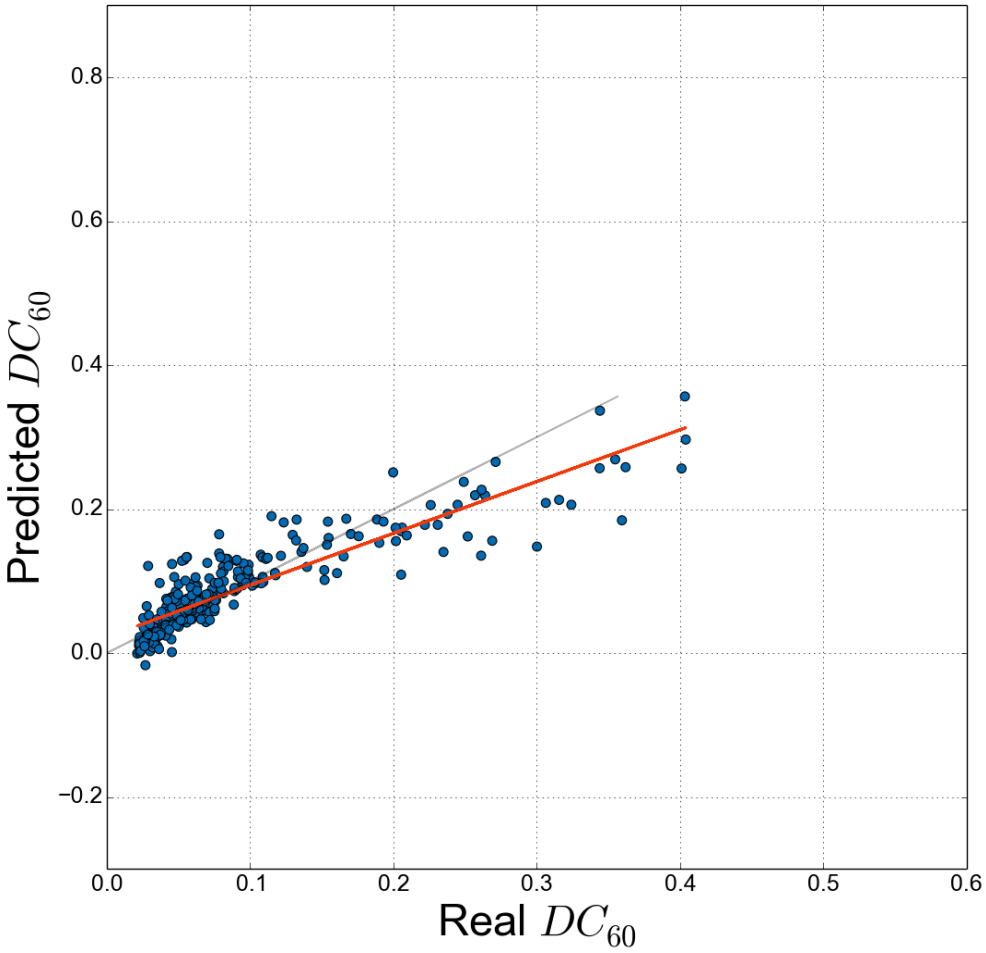


Figure 4.18: Predicted vs real - 4 variables model - filtered– MLE

Table 4.9: 4 variables model – filtered - MLE

Overall	RMSE	0.036
	RMSE / $range$	0.095
	Pearson's r	0.891
RoI	RMSE	0.028
	RMSE / $range$	0.280

4.4.1.5 Comparison results

The summary of the results obtained from the creation of the different surrogate models is given and reported in Table 4.10. The models based on three variables show a general lower value of correlation between predictions and actual values and therefore their predictive capability, also possibly expressed by the coefficient of determination, is certainly inadequate. Though, in the case of the filtered dataset for the three variables models, the best and lowest result for the RMSE within the region of interest is achieved. Though, this model, considering both correlation coefficient and overall model quality, expressed by the ratio RMSE/$range$, is not feasible for the purposes of the current research.

Nevertheless, after different tests, the major interest has been moved to the four variables model, since clear improvements are obtained when the dimensionality of the problem is increased. The overall quality of this enhanced model is generally better than the one related to three variables models and, moreover, when the filtered dataset is considered the RMSE in the range of interest reaches almost the same value as the one associated to best case found in the previous models. The additional advantage earned with this model is that the RMSE evaluated on the complete dataset is at its lowest and it is considered within the acceptable limits established by Forrester [16], as stated in the previous paragraph.

Table 4.10: Comparison between Kriging models created

Variables	Dataset	Overall			RoI	
		RMSE	$\frac{\text{RMSE}}{range}$	r	RMSE	$\frac{\text{RMSE}}{range}$
CR, AR, r_{IL}	Non-filtered MLE	0.119	0.211	0.459	0.029	0.290
	Filtered MLE	0.075	0.196	0.595	0.024	0.240
CR, AR, r_{IL}, κ_{TH}	Non-filtered MLE	0.054	0.095	0.901	0.043	0.430
	Filtered MLE	0.036	0.095	0.891	0.028	0.280

CONCLUSIONS

Scope of this chapter is to provide to the reader a summary of all the work carried out during this project. The main findings will be thoroughly summarised in order to build the foundations of the future work on the argument. Always focusing on the next steps, few recommendations and suggestions are given with the intention of guiding future candidates towards the complex achievement of creation of a surrogate model for intake design.

5.1 Project summary

The principal objective of the project was the construction of a reliable surrogate model for the future exploitation in the optimal design of the short intake bottom line. The whole analysis focused on this line and to the aerodynamics behaviour when it is exposed to the worst flight condition which it can face. This condition is represented by the climb phase, in which the high incidence angle can lead to potential pressure losses and non-uniformities at the fan face, in the situation in which the bottom line is not properly designed. These aerodynamics issues affect irremediably the performance of the engine, leading to surge and to all the potential mechanical damages which can take occur when the flow interacts with the fan rotor and with the compressor. In order to create a reliable surrogate model, different steps were required before the achievement of acceptable results. The project started from the foundations built by the outcome of the initial design space exploration carried out for the bottom line at incidence [22]. The result led to the decision of opting for a reduced dimensionality problem, moving from five independent design parameters, characterised by specific ranges, to three main independent variables and two automatically derived parameters. The ranges of the new reduced design space had to be found, and an initial exploration carried out by means of a Full Factorial (FF) Design of Experiment (DoE) was exploited. It allowed to efficiently determining the convex hull of the hypercube which was

going to be populated by all the time- and computationally-affordable design combinations of the three main design parameters. After an analysis of the different surrogate modelling techniques available in literature, the one which was selected was the Kriging method, efficient solution when highly non-linear problems are subject of investigation. In fact, as preliminary shown in previous approaches to the problem, the intake bottom line behaviour when exposed to incidence condition is highly non-linear. Along with the selection of the surrogate modelling technique, the proper selection of the DoE method used to populate the design space on which the creation of the final metamodel is based is crucial. The decision about which technique was best for this purpose was pretty straightforward and eventually the Latin Hypercube Sampling (LHS) was selected, due to its high reliability accompanied by enhanced space-filling capabilities [15]. This particular characteristic of the LHS method led to the generation of a well-spread distribution of samples across the entire design space, allowing to analyse the behaviour of the intake bottom line at incidence conditions for a large number of possible combinations of the design variables. In accordance with what found in the initial analysis of the intake bottom line aerodynamics [21], since the two aerodynamics performance parameters evaluated, namely Intake Pressure Recovery (IPR) and distortion coefficient DC_{60}, are linearly related, it was decided to focus the attention solely on one of the two metrics. The one which showed the worst behaviour in the aerodynamic analysis was the DC_{60}, and for this reason it was selected as the main metric to study. After the LHS DoE, the design space exploration showed that the dataset obtained was characterised by a highly non-linear distribution of the metric of interest. This was also demonstrated by the fact that the application of the surrogate modelling technique selected was ill-suited for such dataset. A thorough analysis of the design space was therefore mandatory, and this required the refinement of the design space bounds and the geometric analysis of the intake shape, with main focus oriented to the study of lip geometry. This combined analysis allowed to filter some of the undesired geometries and poor meaning regions out of the initial dataset. At this stage it was possible to optimise the surrogate modelling technique selected and adapt it to the filtered design space. Due to the high complexity of the surrogate modelling technique

selected, related to the high number of possible combinations of the different settings, a large number of tests was required to find the optimal settings which allowed its proper operation. Nevertheless, the creation of the reduced order model based on the three main independent variables selected has been demonstrated to be barely acceptable. Eventually, an important result was obtained: a careful increase of the dimensionality of the problem, from three to four design parameters leads to improved results, demonstrated by a better overall quality of the metamodels created. This conclusion, along with the whole discussion reported in this document, should serve as a starting point for the improvement of the employed method and the potential achievement of the objective of creation of the surrogate model for the bottom line design.

5.2 Main findings

As conclusion of the work, few points can be highlighted as main findings and they can be summarised as following:

- After the completion of a DoE run, a thorough analysis of the design space through scatterplots of the correlations between the different variables has to be carried out. This can lead to a refinement of the design space which is not possible to achieve at the stage of determination of the design space bounds. Considering an intake geometry, in the specific case of intake at incidence the analysis of the lip design is mandatory, since it is a key feature which drives the aerodynamic behaviour of the air stream which flows inside the intake duct. The methodologies implemented are not completely able to smooth out the radius of curvature distribution of the lip, and therefore a manual inspection is due, in order to identify the potential geometries characterised by an undesired shape.
- The Kriging method is a complex method to implement, and it requires a rigorous strategy which allows to determine the settings which lead to optimal solutions, or metamodels. Based on the achievements reached at the end of this project, further investigation has to be carried out in order to

explore new combinations of settings, which can over-perform the ones tested during this work.

- The three variables based models have shown an overall poor quality in terms of generalisation capabilities. The quality of a metamodel was assessed through Leave-One-Out (LOO) cross-validation, and two different summary statistics were used to describe the found results quantitatively. The best result was achieved when the filtered dataset was considered, and when the auto-correlation parameters were defined by the user and set as fixed for all the cycles of the cross-validation. The overall quality can be described by RMSE $= 0.054$. Though, it has been demonstrated that when the auto-correlation parameter is defined at each cycle of the cross-validation by a method known as Maximum Likelihood Estimation the overall quality decreases, but the quality for the range of interest of the metric, set at $DC_{60} < RoI$, improves and reaches the best values found at RMSE $= 0.024$. This can be assumed as a good quality results, but it has been shown that the associated Pearson coefficient is very low and it highlights poor predictive capabilities of the model [16].
- The four variables based models have shown an overall better quality in terms of RMSE compared to the previous discussed models. Considering the filtered dataset, the overall quality reaches RMSE $= 0.036$, while in the range of interest it was almost equal to the lowest found in the three variables model. Nevertheless, according to Forrester [16], if the ratio RMSE$/range = 0.095$ is evaluated, it is possible to assert that with the current model it was reached a reasonable level of prediction capabilities, which Forrester sets until a maximum acceptable value of RMSE$/range =$ 0.10.

5.3 Future considerations

In the path of achievement of the main objective, the current project has explored a numerous variety of possible ways in which the surrogate modelling technique selected could be applied. Nevertheless, many of these have led to inexact

conclusions and many others have led to cross-validations which highlighted the creation of metamodels affected by very high errors in the predictions.

The suggestions which are derived from the work carried out during this year can be summarised in few points:

- The results showed that the increase of dimensionality of the problem led to improved results. In the context of this research, the additional design parameter was added to the dataset using the values automatically calculated during the DoE creation. In fact, as thoroughly discussed, the additional variable is not allowed to vary within determined ranges, but it has to be considered only for the values obtained from the automatic calculations. Following the same strategy, and assuming that the behaviour of the metamodel is the same, it should be tested the addition of a fifth variable to the already increased dimensionality problem. Few tests have been carried out during the project, trying to introduce as additional variables average diffuser angle and curvature at the fan face κ_{FAN}. Nevertheless, this has not generally led to improved results, and, in some cases, they even worsened the original result.
- Accordingly to the previous bullet point and to the final results obtained in the current research, it is suggested to employ the curvature at the throat κ_{TH} as an additional main independent variable, instead of calculating it through the least square approach described. The possibility to vary its value within determined bounds could allow to explore more precisely the design space, with the added potential benefit of enhancing the design of the lip. Moreover, since the addition of the automatic calculated κ_{TH} leads to substantial improvements of the quality of the metamodel created, it is expected that if the bounds of the same variable could be enhanced through the same procedure exploited at the beginning of the current research, this parameter can more efficiently participate to the improvement of the metamodel. Nevertheless, the potential benefits related to the addition of a new independent parameter to the original three variables set can be overweigh by the computational time required to obtain the same number

of samples as the one obtained at the beginning of the project. In fact, this problem is well-known as 'curse of dimensionality' [16], and it requires a thorough analysis of the various relationships between metrics of interest and design variables to examine the actual feasibility. Though, the clear demonstration obtained in terms of quality improvement following the addition of κ_{TH} suggests that it could be worthy to explore this possibility.

- Further analysis of the literature available regarding the creation of surrogate models, it appears that interesting results have been obtained with the exploitation of Neural Networks. The mathematical background of this method is even more complex than the one which lies behind the Kriging method, and it offers a robust method which is worthy to be examined. The MATLAB® implementation is thoroughly documented [39] and its reliability has been well reported in different papers. Solutions for programming languages such as Python are available as extensions of the SciKit module [40], the same class of modules in which the Gaussian Processes module exploited for the current research is contained.

REFERENCES

[1] J. Whurr, "Future civil aeroengines architectures and technologies", 10th European Turbomachinery Conference, Lappeenranta, Finland, 2013.

[2] A. C. Nicholas, “Preparing for the future: reduction gas turbine environmental impact - IGTI Scholar Lecture,” *Journal of Turbomachinery*, vol. 132, no. 4, p. 17, 2010.

[3] J. E. Green, “Civil aviation and the environmental challenge,” *The Aeronautical Journal*, vol. 107, no. 1072, pp. 281-300, 2003.

[4] D. Nalianda, *Propulsion Systems Performance and Integration notes, Vol.1*. Cranfield University, 2017.

[5] A. K. Kundu, *Aircraft design*, 1st ed. Cambridge University Press, 2010.

[6] K. Hünecke, *Jet engines: Fundamentals of theory, design and operation*. 1997.

[7] Rolls-Royce, *The jet engine*. John Wiley & Sons, Inc., 1996.

[8] J. Seddon and E. L. Goldsmith, *Intake aerodynamics*, 2nd ed. Blackwell Science Ltd, 1999.

[9] E. Obert, *Aerodynamic design of transport aircraft*. IOS Press, 2009.

[10] J. Seddon and E. L. Goldsmith, *Practical intake aerodynamic design*, 1st ed. Blackwell Scientific Publications, 1993.

[11] “Restricted Report #4.”

[12] “Restricted Report #2.”

[13] A. Peters, Z. S. Spakovszky, W. K. Lord, and B. Rose, “Ultrashort nacelles for low fan pressure ratio propulsors,” *Journal of Turbomachinery*, vol. 137, no. 21001–1, 2015.

[14] D. C. Montgomery, *Design and analysis of experiments*, 7th ed. Wiley, 2009.

[15] M. Cavazzuti, *Optimization methods: From theory to design*. Springer, 2013.

[16] A. I. J. Forrester, A. Sóbester, and A. J. Keane, *Engineering design via surrogate modelling*, 1st ed. Wiley, 2008.

[17] Noel A. C. Cressie, *Statistics for spatial data*, Revised ed. John Wiley & Sons, Inc., 1993.

[18] B. MacIsaac and R. Langton, *Gas turbine propulsion systems (AIAA Education Series)*, 1st ed. John Wiley & Sons, Inc., 2011.

[19] "Restricted Report #5."

[20] K. Deb, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Inc., 2001.

[21] "Restricted Report #3."

[22] "Restricted report #1."

[23] R. Christie, A. Heidebrecht, and D. MacManus, "An automated approach to nacelle parameterisation using Intuitive Class Shape Transformation curves," in *ASME Turbo Expo 2016: Turbomachinery Technical Conference and Exposition*, vol.139 (6, GTP-16-1227), 2016.

[24] M. Albert and D. Bestle, "Aerodynamic design optimization of nacelle and intake," in *ASME Turbo Expo 2013: Turbine Technical Conference and Exposition*, no. GT2013-94857, 2013.

[25] H. Sobieczky, "Parametric airfoil and wings," *Recent Development of Aerodynamic Design Methodologies*, pp. 71–87, 1999.

[26] B. M. Kulfan and J. E. Bussoletti, "Fundamental parametric geometry representations for aircraft component shapes," *11th AIAA/ISSMO multidisciplinary analysis and optimization conference*, AIAA 2006-6948, pp. 1-45, 2006.

[27] C. Hirsch, *Numerical computation of internal and external flows, vol.1: the fundamentals of Computational Fluid Dynamics*. John Wiley & Sons, Inc., 1988.

[28] T. J. Santner, B. J. Williams, and W. I. Notz, *The design and analysis of computer experiments*. Springer, 2003.

[29] "Gaussian Processes - Scikit-learn Python module." [Online]. Available: http://scikit-learn.org/0.17/modules/gaussian_process.html. [Accessed: 05-Mar-2017].

[30] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

[31] "SciPy.org." [Online]. Available: https://scipy.org/. [Accessed: 20-Feb-2017].

[32] M. J. D. Powell, "Direct search algorithms for optimization calculations," *Acta Numerica*, vol. 7, pp. 287–336, 1998.

[33] W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, and M. D. Morris, "Screening, predicting, and computer experiments," *Technometrics*, vol. 34, no. 1, pp. 15–25, 1992.

[34] J. Ollar, C. Mortished, R. Jones, J. Sienz, and V. Toropov, "Gradient based hyper-parameters optimisation for well conditioned kriging metamodels," *Structural Multidisciplinary Optimisation*, no. 55, pp. 2029–2044, 2017.

[35] R. Moustafa and E. J. Wegman, "Multivariate continuous data - Parallel Coordinates," in *Graphics of large datasets: visualising a million*, Springer, 2006, pp. 143–156.

[36] A. Pierluissi and C. F. Smith, "Intake lip design system for gas turbine engines for subsonic applications," in *49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, AIAA 2011-1130, 2011.

[37] ArcGIS-ArcMap, "ArcGIS Desktop Website." [Online]. Available: http://desktop.arcgis.com/en/arcmap/. [Accessed: 24-Mar-2017].

[38] B. S. Everitt, *Cambridge dictionary of statistics*, 2nd ed. 2002.

[39] H. Demuth and M. Beale, *Neural Network Toolbox - For use with MATLAB®*, Release 13. 2002.

[40] "SciKit Neural Networks website." [Online]. Available: http://scikit-neuralnetwork.readthedocs.io/en/latest/index.html. [Accessed: 19-Jul-2017].

APPENDICES

Appendix A

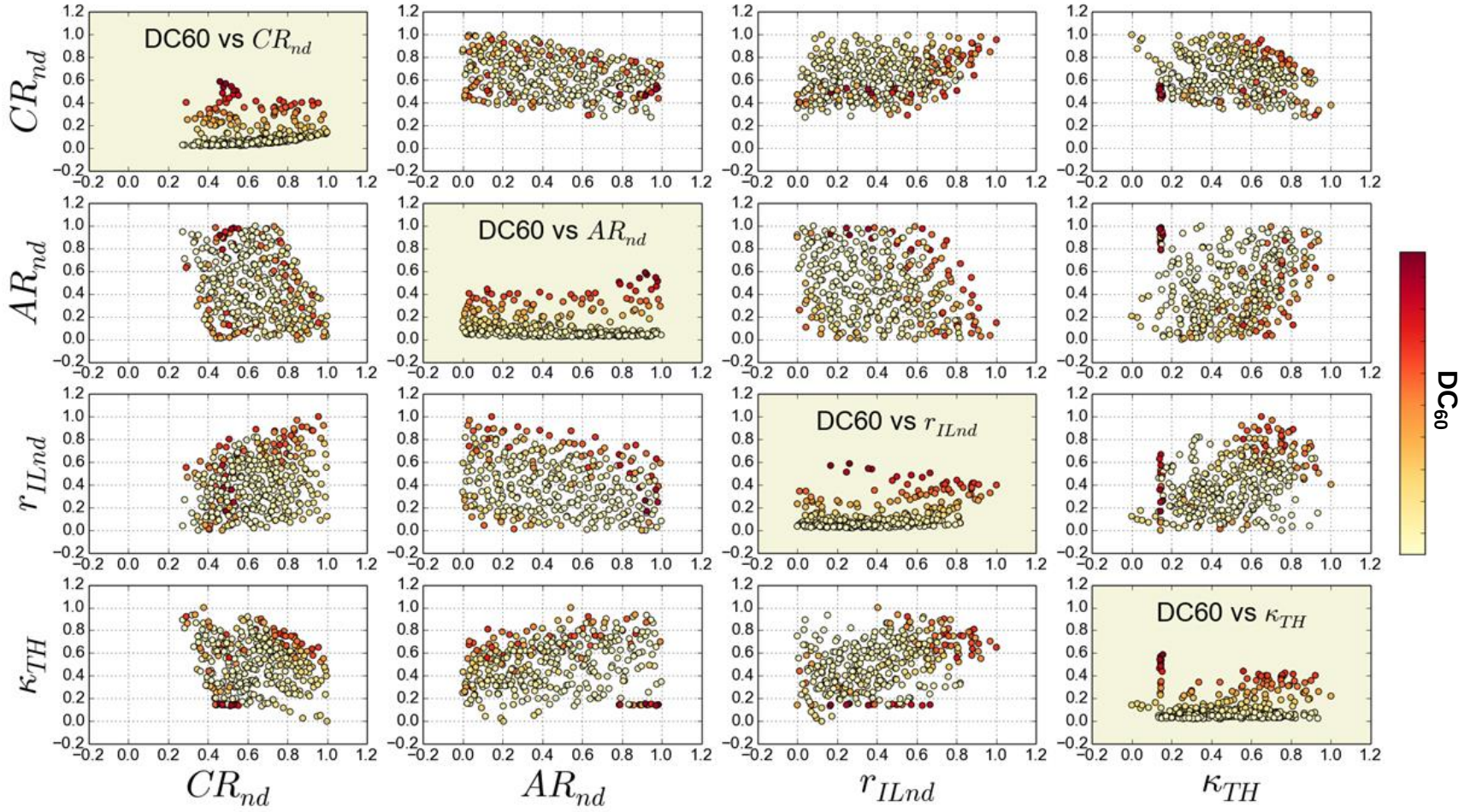
A.1 Design space exploration

The material contained in this appendix has a twofold objective:

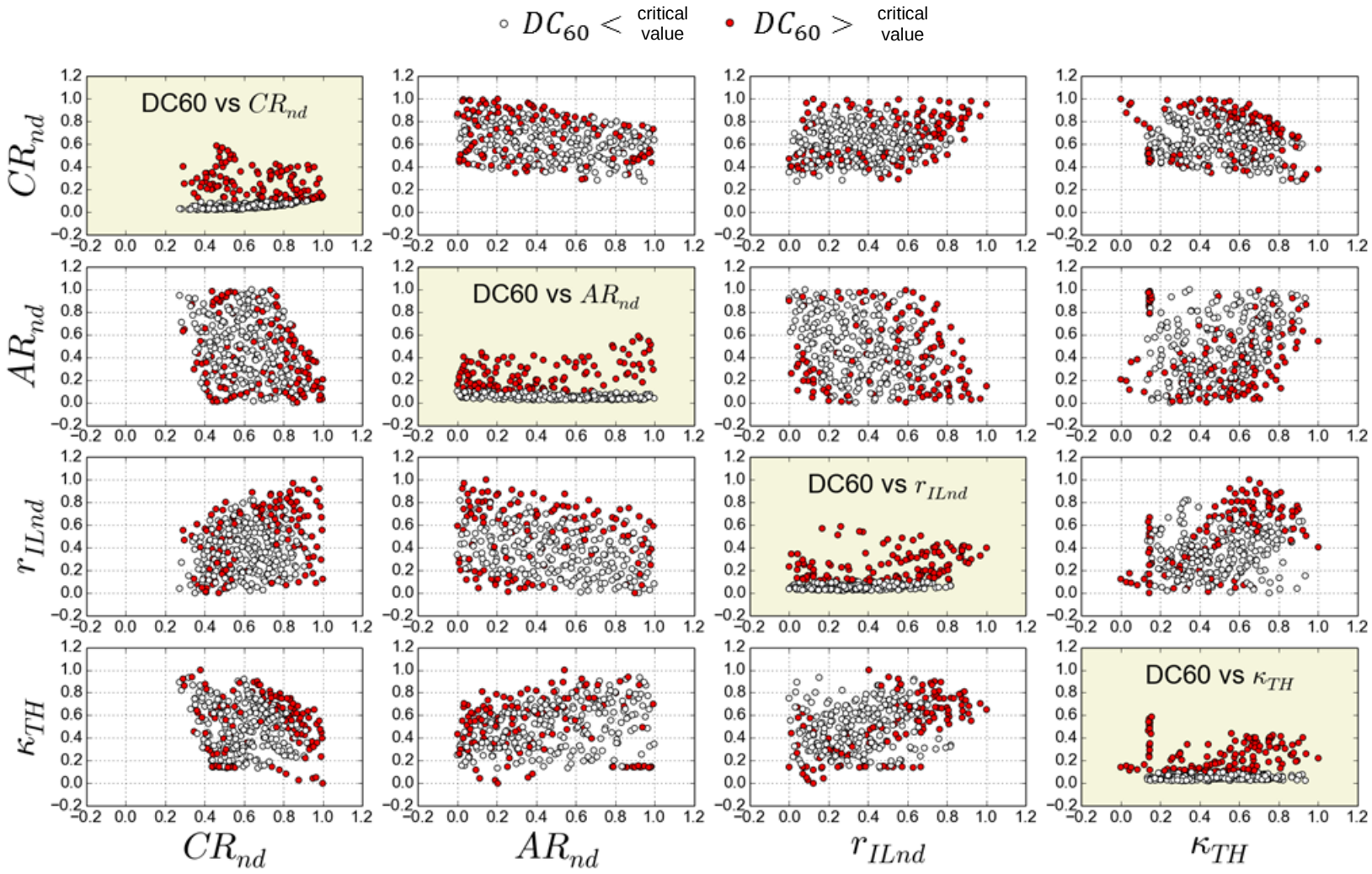
- It can be used for future analysis.
- It is useful in order to have a deeper insight in the work carried out in the context of the geometry analysis.

It mainly consists of plots and tables, which are an integration or extension of the arguments discussed in Chapter 4.

A.1.1 Geometric correlations analysis

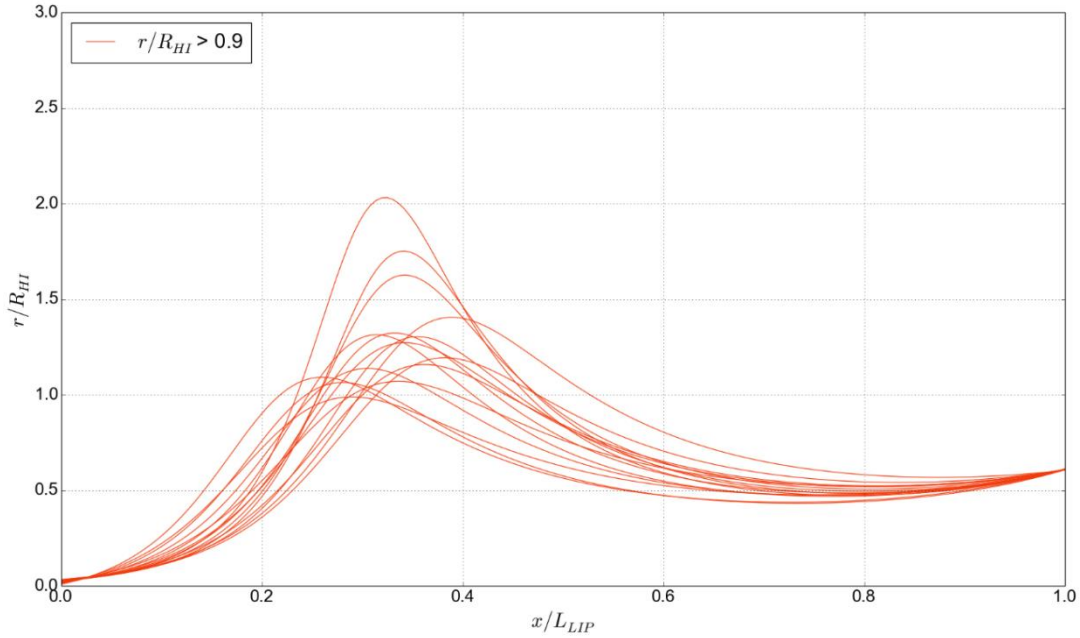


Geometric correlations scatterplots

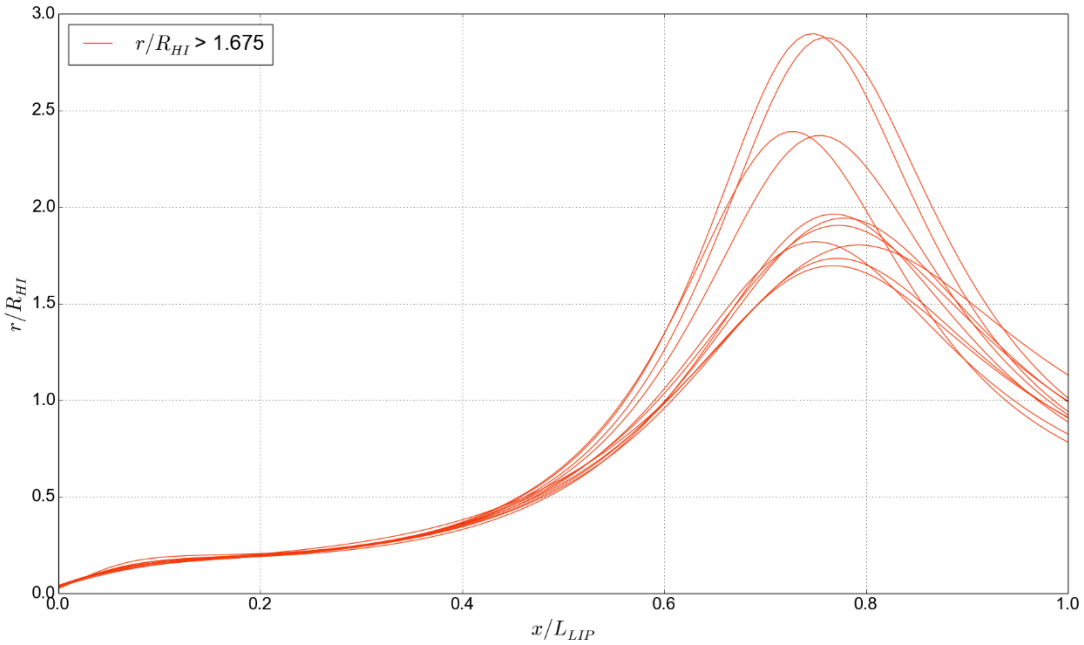


Geometric correlations with differentiation based on RoI

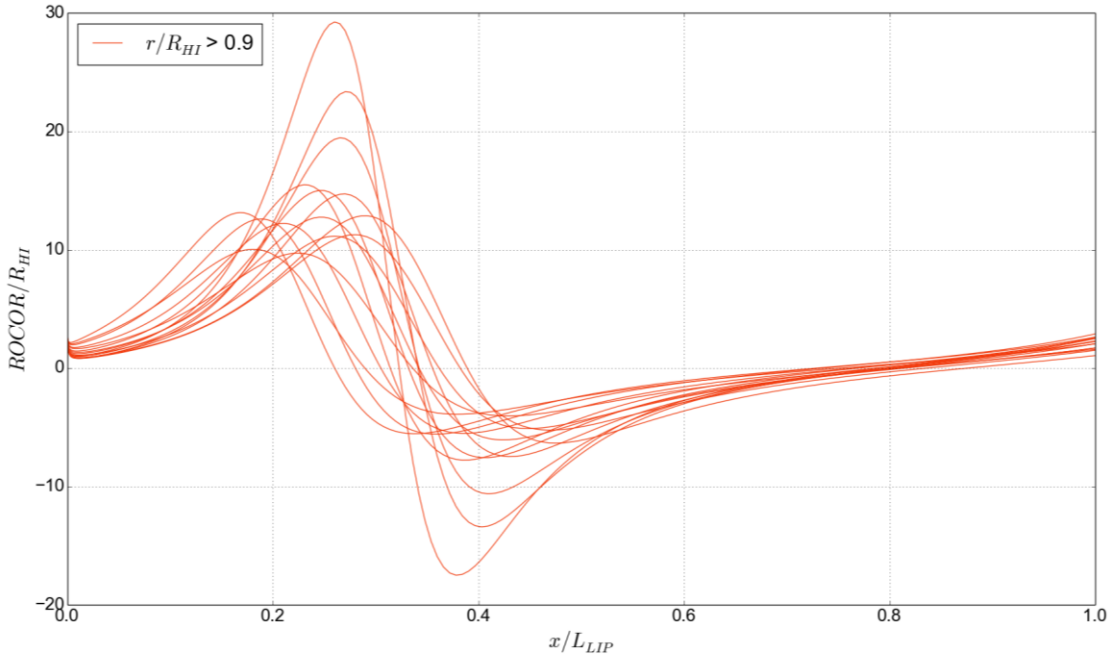
A.2 Lip design analysis



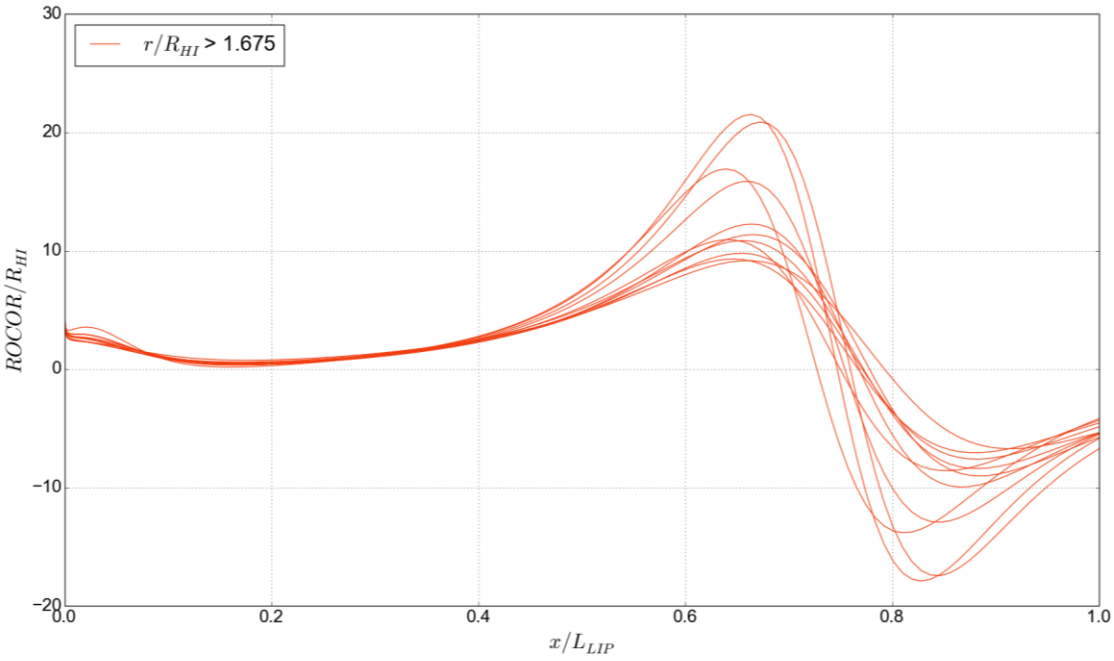
Fore-lip undesired geometries – radius of curvature distribution



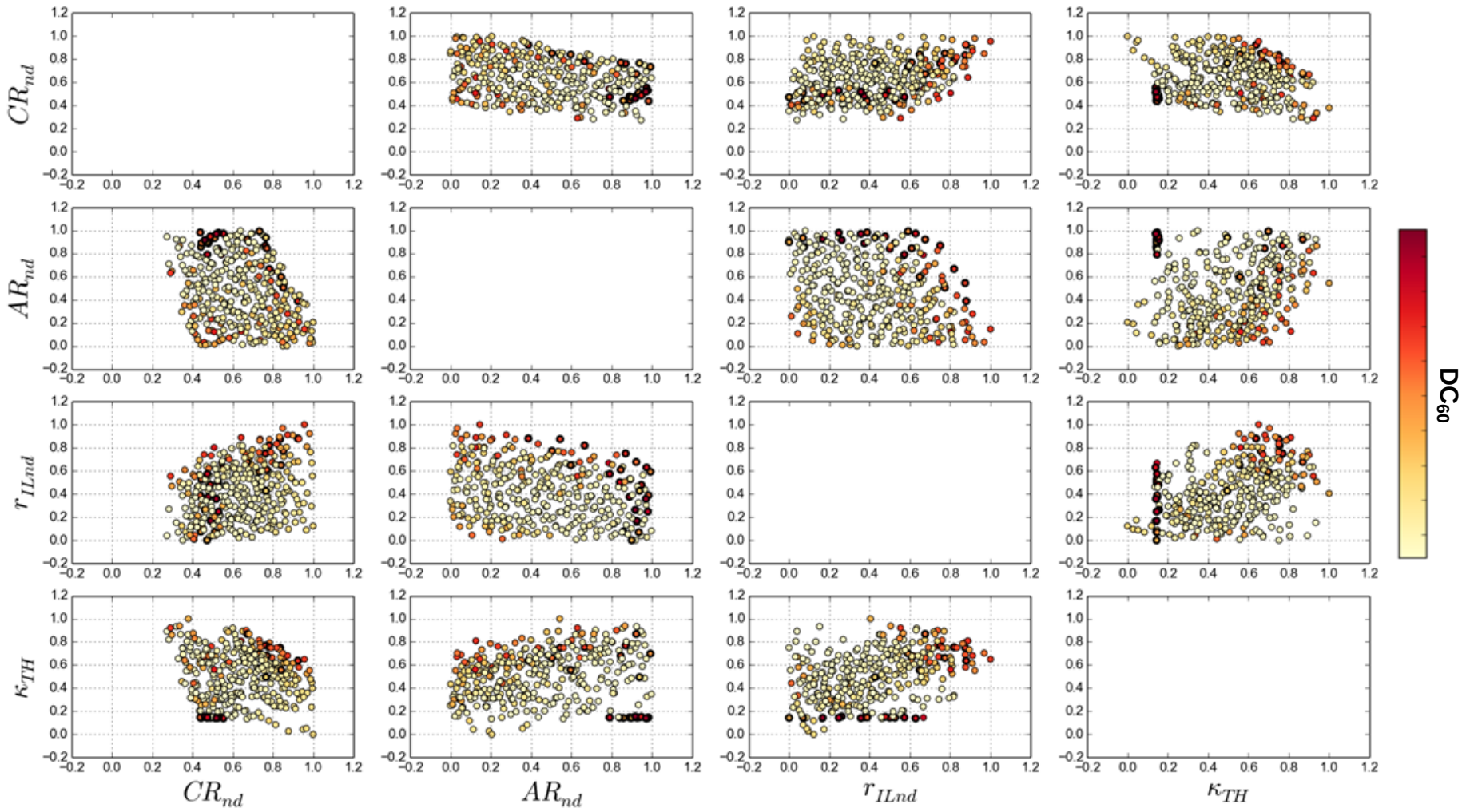
Aft-lip undesired geometries – radius of curvature distribution



Fore-lip undesired geometries - ROCOR distribution



Aft-lip undesired geometries - ROCOR distribution



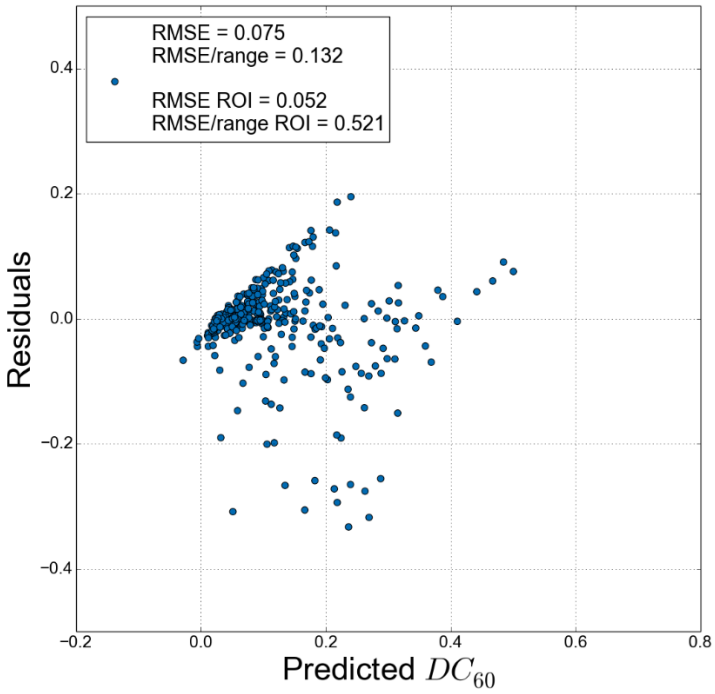
Lip undesired geometries highlighted in design space

Appendix B

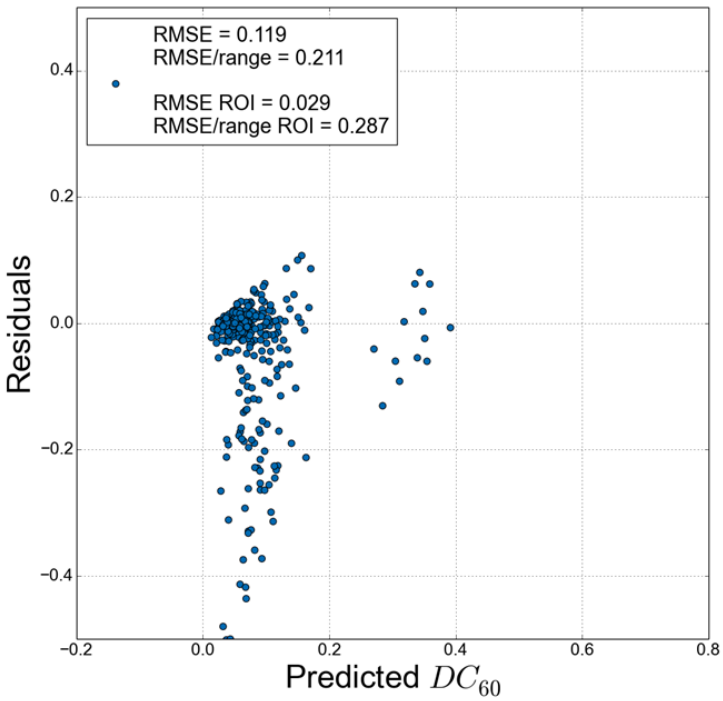
B.1 Kriging model – residuals vs predicted values

The plots representing the relationship between the residuals and the predicted values of the metric are reported. They are not included in the discussion about the Surrogate modelling in order to improve the readability, but they are can be subjected of investigation for further analysis.

B.1.1 Three variables models – non-filtered dataset

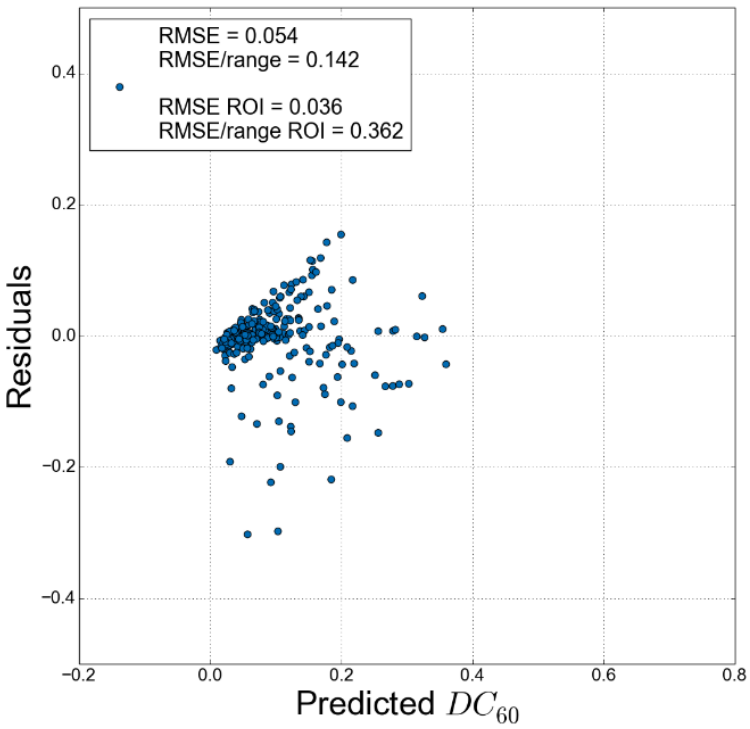


Residuals vs Predicted - 3 variables - non-filtered - UDACP

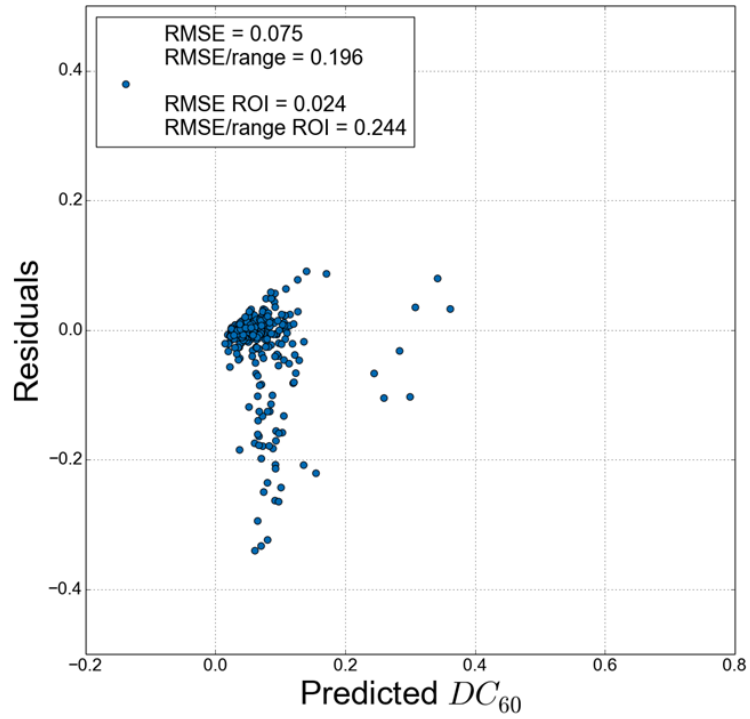


Residuals vs Predicted - 3 variables - non-filtered - MLE

B.1.2 Three variables models – filtered dataset

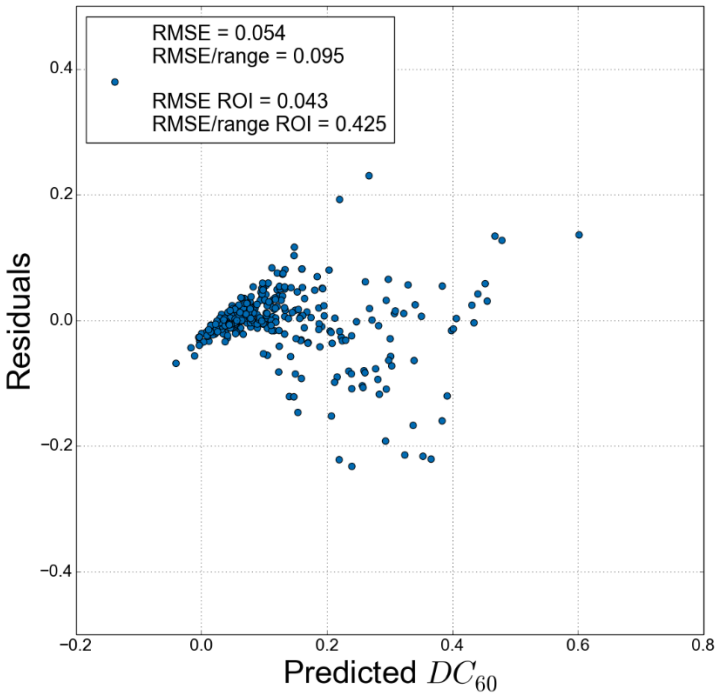


Residuals vs Predicted - 3 variables - filtered - UDACP

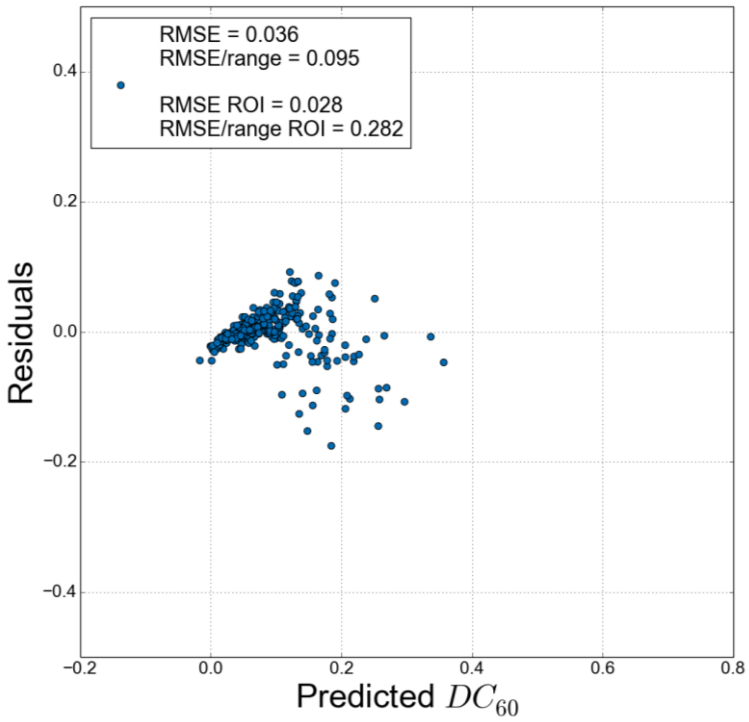


Residuals vs Predicted - 3 variables - non-filtered - MLE

B.1.3 Four variables models



Residuals vs Predicted - 4 variables - non-filtered - MLE



Residuals vs Predicted - 4 variables – filtered - ML