CHALMERS
UNIVERSITY OF TECHNOLOGY

Intake distortion synthesis using steady-state CFD data to estimate dynamic distortion

Master's thesis in Mobility Engineering

SOFIA BJÖRS

Department of Mechanics and Maritime Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Intake distortion synthesis using steady-state CFD data to estimate dynamic distortion

SOFIA BJÖRS

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mechanics and Maritime Sciences
Division of Fluid Mechanics
Chalmers University of Technology
Gothenburg, Sweden 2026

Intake distortion synthesis using steady-state CFD data to estimate dynamic distortion
SOFIA BJÖRS

Supervisor: Stefan Nilsson, Saab AB, Aeronautics
Examiner: Carlos Xisto, Department of Mechanics and Maritime Sciences

Master's Thesis 2026
Department of Mechanics and Maritime Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Intake distortion synthesis using steady-state CFD data to estimate dynamic distortion
SOFIA BJÖRS
Department of Mechanics and Maritime Sciences
Chalmers University of Technology

Abstract

A challenge when designing the engine air intake of an aircraft is the generation of non-conformities in the flow. These are often a direct consequence of the design of the intake but may also arise from other factors. The flow characteristics and performance can be modeled by using complex and resource-heavy computational fluid dynamic (CFD) modeling. However, in the early design stages, such simulations are not suitable and simpler and less demanding ways to model distortion is needed. In this thesis, five different methods are evaluated; the first method is considered a baseline, which uses only wind tunnel data to synthesize the distortion. It uses normally distributed random numbers to create pressure fluctuations based on measured turbulence. The second method replaces the measured turbulence from the previous method with a calculated turbulence based on CFD data, which is calculated using the means of least square error. For the third method, a machine learning model was trained using CFD data as inputs and measured turbulence from a wind tunnel as the output to calculate the dynamic distortion. The last two methods are variations of the same machine learning model as the third method; however, they use other inputs from the CFD simulation. The results show that all five methods worked well, with the lowest accuracy of a method being 97.9% and the highest being 98.83%. The methods are also compared by calculating three distortion indexes: CDI, RDI, and DC60. The results show that the radial distortion (RDI) has the highest accuracy but that the synthesis tends to underpredict the dynamic pressure and miss the maximum distortion. In conclusion, the best results come from using machine learning, but due to a low amount of data, it is hard to draw any definitive conclusions.

Keywords: Turbulence modeling, Intake distortion, Machine learning, Turbulent Kinetic Energy, Distortion synthesis.

Acknowledgements

I would first like to thank my supervisor, Stefan Nilsson at Saab, for his never-ending support and guidance through the whole thesis and for keeping me on track. Furthermore, I want to express my gratitude to Saab for the opportunity to perform this master's thesis and also to my colleagues for providing their knowledge and giving me new ideas. A final thanks is extended to Carlos Xisto for his work as examiner during the thesis.

Sofia Björs, Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
AIP	Aerodynamic Interface Plane
ANN	Artificial Neural Network
APD	Amplitude Probability Density
BLI	Boundary Layer Ingestion
BP ANN	Back Propagation Artificial Neural Network
CDF	Cumulative Distribution Function
CDI	Circumferential Distortion Index
CFD	Computational Fluid Dynamics
LES	Large Eddy Simulation
PIML	Physics Informed Machine Learning
PDF	Probability Density Function
PPF	Percent Point Function
RANS	Reynold's Averaged Navier-Stokes
RDI	Radial Distortion Index
ReLU	Rectified Linear Unit
RL	Reinforcement Learning
RMS	Root Mean Squared
RSM	Reynold's Stress Models
SAE	Society of Automotive Engineers
SGD	Stochastic Gradient Descent
SST	Shear Stress Transport
Tanh	Hyperbolic Tangent
UAV	Unmanned Aerial Vehicles

ν

Nomenclature

Below is the nomenclature of indices, parameters, and variables that have been used throughout this thesis.

Indices

i	Ring index for AIP
i, j, k	Index for tensor notation
T	Index for time step

Parameters

C_μ	Turbulent Viscosity Closure Coefficient
γ	Heat Capacity Ratio for air
g_c	Acceleration of Gravity [m/s^2]
R	Gas Constant for Air

Variables

a	Speed of Sound [m/s]
C	Correlation Coefficient
δ_{ij}	Kronecker delta
ε	Turbulent Dissipation Rate [J/kg $\cdot$ s]
ε	Error term
k	Turbulent kinetic Energy [J/kg]
μ	Dynamic viscosity
μ	Mean steady-state pressure for APD function
ν_t	Turbulent kinematic Viscosity [m^2/s]

ω	Specific Dissipation Rate [1/s]
p	Linearly spaced pressure axis
P_{AIP}	Average Static Pressure Over Aerodynamic Interface Plane [Pa]
$P_{DYN,i}$	Dynamic Pressure in Ring i [Pa]
$P_{T_{DYN}}$	Dynamic Total Pressure [Pa]
P_{T_f}	Fluctuating Pressure [Pa]
P_{total}	Total Pressure [Pa]
$P_{T_{SS}}$	Steady-State Pressure [Pa]
P_{synt}	Synthesized pressure [Pa]
$P_{0,AIP}$	Average Total Pressure Over Aerodynamic Interface Plane [Pa]
$P_{0,i}$	Average Total Pressure in Ring i [Pa]
$P_{0,i,LOW}$	Lowest Total Pressure in Ring i [Pa]
P_{60MAX}	Maximum Average Pressure Over a 60^o Sector of The Aerodynamic Interface Plane [Pa]
Φ	Dissipation
ρ	Density [kg/m^3]
S_{ij}	Strain-rate tensor
σ	RMS turbulence for APD function
T	Temperature [K]
T	Timestep
t	Timestep
$Turb$	RMS Turbulence
u	Axial velocity [m/s]
u	Velocity [m/s]
v_i	Velocity vector [m/s]

Contents

List of Figures

1
Introduction

In the development of new fighter jets and all other aircraft, engine intakes play a crucial role. They deliver the airflow required for both combustion and generating thrust and thus determine how the engine performs. However, due to real-world operating conditions, non-uniform flows are frequently introduced at the engine face, known as intake distortion. The intake distortion results in spatial variations in total pressure and temperature over the aerodynamic interface plane (AIP). Variations in pressure alter the local blade loading and can cause regions of the compressor to operate closer to stall than others. The resulting flow separation and instability reduce compressor performance and stall margin, while increasing aerodynamic forcing on the blades [1], [2]. Intake distortion is thus a critical factor in compressor stability and can contribute to rotating stall, surge, vibration, and reduced engine reliability. This could impair the safety of the engine and decrease the efficiency [3]. Studying the intake distortion has thus increased in relevance as the development advances and the designs of the aircraft become more complex. Challenging modern configurations such as S-duct intakes, boundary layer ingesting (BLI) propulsion systems, and closely integrated designs between the nacelle and fuselage are all examples of this, and these all introduce distortion characteristics that challenge traditional methods of designing the engine intakes [4], [5]. The increased demand for flight maneuverability and extreme conditions also plays a role in the increased complexity of the design process [6]. Furthermore, as new applications emerge, such as unmanned aerial vehicles (UAVs), blended-wing-body aircraft, and sustainable aviation concepts, a new demand has risen for a more robust understanding of how distortion is generated and how it affects the performance of varying flight conditions [7].

Synthesizing intake distortion is not a simple task, and this is a key challenge in distortion research. Doing the synthesization consistently in a controlled and repeatable manner is difficult and has always been an obstacle. However, with the improvement of computational fluid dynamics (CFD) and the usage of secondary flow generating devices (screens, wedges or flow actuators) in experimental wind tunnel testing, it is becoming easier to simulate [5]. Each method has its advantages and limitations when it comes to fidelity, cost, flexibility and relevance to actual flight conditions. Understanding which method for synthesizing distortion is the best to represent the distortion characteristics for an intake configuration is crucial for doing accurate stability assessments for the engine and for developing a reliable criteria for distortion tolerance [8], [9].

This report investigates five different methods for synthesizing distortion in the intake of a jet engine. It includes a comparative analysis of both experimental and numerical approaches and evaluates them through three different distortion descriptors which are commonly used to quantify the maximum distortion [10], [11]. An assessment is also made on how each method of synthesization captures the circumferential and radial distortion. By examining and comparing the strengths and weaknesses of each method, this thesis aims to provide insights into the best practices for distortion synthesis and to inform future studies in integration of engine and intake.

1.1 State of the art

Multiple contributions have been made to this field over the years, with new advancements in technology leading to the development of new methods. The following sections will show a timeline of how the methods for synthesizing distortion have evolved.

1.1.1 Early experimental distortion studies

In as early as the 1960's, engine researchers were starting to understand the impact of intake distortion. Although not much research was being done into predicting distortion, many studies were done which looked at the impact distortion had on turbomachines [12], [13]. These studies highlighted the need for finding reliable methods to predict the distortion in order to improve the performance of the engine. This research continued into the 1970's with researchers looking at ways to predict the distortion and laying the ground-work necessary for continued development [14]. In 1976, a study was done by King, Schuerman and Muller [15] where a method for synthesizing distortion was delivered. It worked by defining a reference sector of 180^o with the lowest steady-state total pressure and then intensified the pressure contour by reducing the probe pressures inside that sector. In the 80's a new method was presented in a report by Motycka [16] and in a report by Sedlock [8]. It worked by splitting the dynamic pressure into a steady-state and a fluctuating part, as:

$$P_{T_{DYN}} = P_{T_{SS}} + P_{T_f} \tag{1.1}$$

Where it is assumed that the fluctuating, time-variant, component has a zero mean and is normally distributed. By introducing a series of random numbers with the same zero mean and normally distributed components, the time-fluctuating component could be synthesized and thus a method for synthesizing intake distortion was born [8]. The main issue with the methods from the 80's were the complete reliance on wind tunnel data. Researchers found a way to simulate turbulence in the wind tunnels by adding, for example, wedges which obstructed the flow and created vortices [17]. However, reproducing complex three-dimensional distortion from modern intake geometries was a difficult task. The practices developed in this period of time had shortcomings, but they built a foundation on which future research could continue.

1.1.2 CFD based distortion analysis

In the 90's, the growing need for ways to predict distortion without expensive and time-consuming wind tunnel testing together with the development of more accurate turbulence modeling solvers in computational fluid dynamics enabled the development of CFD based distortion prediction. The development of comprehensive and accurate Reynolds-averaged Navier Stokes (RANS) based models, such as $k-\varepsilon$ and $k-\omega$, enabled modeling turbulence in increasingly complex geometries, such as S-ducts [18]. Norby et. al. [9] built on top of the previously mentioned synthesization method from 1989 and found a way to replace the wind tunnel with CFD data through finding a correlation between the turbulent kinetic energy from CFD and the RMS turbulence from the wind tunnel. This created a way to model the turbulence using CFD data. The advancements in turbulence modeling using CFD during this period reduced the reliance on expensive wind tunnel testing in early design stages and enabled a faster development. It came with drawbacks however, as RANS has a tendency to under-predict dynamic distortion [18].

1.1.3 Modern methods

The most notable contribution to turbulence modeling in modern times is the usage of machine learning. In 2008, a study was released in which a back propagating artificial neural network (BP ANN) was used to predict distortion by training the network on data from CFD simulations and dynamic pressure from a wind tunnel [19]. The study found that using machine learning algorithms on turbulence prediction worked superiorly to the previously proposed methods. Other applications of machine learning in turbulence models include using a machine learning model as a way to improve RANS simulations. A study by Wang et. al. [20], improved and republished by Wu et. al. [21], presented the use of a physics-informed machine learning (PIML) approach to learn the discrepancy of the functional form of the Reynold stress discrepancy present in RANS simulations. In 2022, a study by Liu et. al. [22] was presented in which machine learning was used as a tool to improve the RANS two-equation turbulence models. It combines the machine learning algorithm with the transport equations used in a conventional turbulence model. Besides machine learning, development was carried out in the turbulence modeling field as well, with the creation of the large eddy simulation (LES) and hybrid RANS-LES turbulence models for time-accurate distortion prediction. These models are good at capturing the unsteady, large-scale turbulent structures which are often missed in RANS. The downside with these models however is the increased demand in computational resources [23].

1.2 Purpose

The purpose of the thesis is to compare different methods for predicting dynamic distortion. These methods will be evaluated through a comparison of the output synthesized pressure and the true pressure as well as three different distortion indices. As a conclusion, these questions will be answered:

- Which method is the best based on accuracy, time-consumption, and complexity?
- Can steady-state CFD data be used to synthesize distortion?

1.3 Outline

This report first starts with the introduction, which outlines the importance and purpose of the report as well as defines the research questions which will be answered. In the next section, relevant theory will be presented. It will cover the governing equations for fluids, turbulence modeling as well as explaining machine learning. After theory, the methods can be found. In this chapter, five methods (three main methods and two sub-methods) are described as well as the distortion indices. The chapter starts with a description of the wind tunnel experiments and pre-processing of the data. After this, the base method is described. This method will also be a part of the next methods. The next method builds on the first method but uses CFD data instead. The last main method and the two sub-methods also builds on the first method but uses machine learning and CFD data. This chapter closes with the distortion indices. After the method chapter comes the results, which show the accuracy and distortion indices for each method. The results are also discussed and the methods are compared. In the last chapter there is a conclusion.

1.4 Limitations

Data from wind tunnel experiments will be used. However, no experiments will be performed as part of the thesis. Five main methods will be tested; other methods for predicting distortion will not be covered. Only pressure distortion will be treated and not other types of distortion as temperature or swirl. Similar to the wind tunnel data, CFD data will be used but no CFD simulations will be performed as part of this thesis.

2

Theory

In this chapter, the relevant theory will be presented. This includes governing equations of flow, turbulent flows and turbulence modeling, some statistics, machine learning and distortion indices.

2.1 Governing equations

The three fundamental governing equations used to describe fluid dynamics are the continuity, momentum and energy equations [24]. These are built from these three laws of physics:

1. Conservation of mass
2. Newton's second law
3. Conservation of energy

The subsequent equations are described and derived in the following sections.

2.1.1 Continuity equation

The continuity equation is built on the fundamental principle of conservation of mass and can be written as [25]:

$$\frac{d\rho}{dt} + \rho\frac{\partial v_i}{\partial x_i} = 0 \tag{2.1}$$

2.1.2 Momentum equations

The momentum equations, or Navier-Stokes equations, is based on Newton's second law of motion, rewritten for fluids [26]. These equations can be seen as governing the motion of fluids and are derived in all directions for three-dimensional flow. The equations can be written as:

$$\rho\frac{dv_i}{dt} = -\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j}\left(2\mu S_{ij} - \frac{2}{3}\mu\frac{\partial v_k}{\partial x_k}\delta_{ij}\right) + \rho f_i \tag{2.2}$$

This equation is also called the transport equation for momentum [27].

2.1.3 Energy equation

This equation is based on the fundamental physical principle of energy conservation, or the first law of thermodynamics. It is written as [26]:

$$\rho \frac{\partial u}{\partial t} = -P \frac{\partial v_i}{\partial x_i} + \Phi + \frac{\partial}{\partial x_i} \left(k \frac{\partial T}{\partial x_i} \right) \tag{2.3}$$

Where the dissipation, Φ, is written as:

$$\Phi = 2\mu \left(S_{ij} S_{ij} - \frac{1}{3} S_{kk} S_{ij} \right) \tag{2.4}$$

2.2 Turbulent flow

Turbulence is present almost everywhere and although there is no definition of turbulent flow, it can be characterized by a few features [27]:

- Irregularity: It may seem random, but it is not; it is irregular and chaotic, but can be described by the Navier-Stokes equations
- Diffusivity: The diffusivity increases in turbulent flow, and the exchange of momentum also increases.
- Three-dimensional: It is always three-dimensional, although if it is averaged over time, it can be treated as two-dimensional instead.
- Dissipative: An important property of turbulence is dissipation, which increases in turbulent flow. Dissipation is the process of kinetic energy transforming into thermal energy.
- Large Reynolds number: Turbulent flow is characterized by a high Reynolds number as it can only occur at high values.
- Continuum: Turbulent flow is treated as a continuum.

Turbulent flow is said to consist of a spectrum of scales, where a scale can also be called an eddy, which have no definition. It is believed that the eddies exist in a space for a period of time until they are destroyed, and that they have a characteristic length and velocity [27]. The largest eddies are the size of the flow geometry; boundary layer thickness or jet width for example. Their length scale and velocity scale are denoted by ℓ_0 and v_0 respectively. In something called an energy cascade, the largest scales passes kinetic energy to the smaller scales, until it reaches the smallest scales, called Kolmogorov scales, where that kinetic energy is dissipated. Dissipation, as previously stated, is the process of turning kinetic energy into thermal energy. The cascade process is visualized in the energy spectrum in Fig. 2.1.

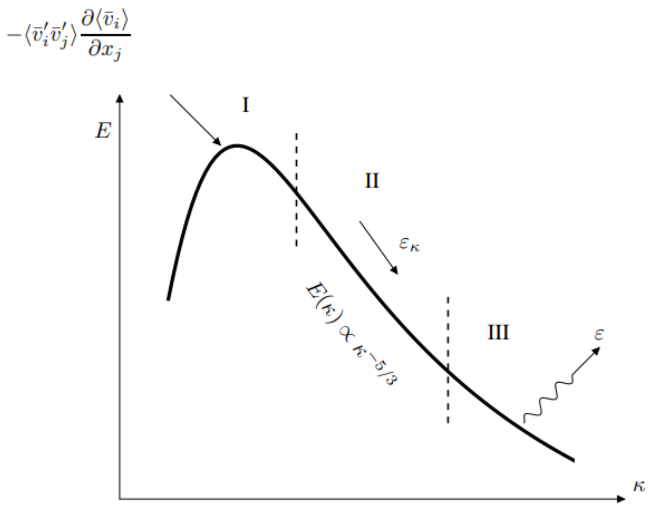

Figure 2.1: Energy spectrum [27]. The figure shows three regions where the first region, I, is where the eddies extract energy from the mean flow. The kinetic energy is then moved through the cascade process into the smaller scales of the middle layer, or the inertial sub-layer. In the third layer, the smallest scales exist where the enrgy is dissipated into thermal energy.

The kinetic energy is extracted from the mean flow by the largest scales as they have a time scale similar to that of the largest scales [27]. As can be seen, turbulent flow is complex and difficult to understand but can be described by various parameters and equations. The following sections will explain a few parameters, which are relevant to this thesis, in more detail.

2.2.1 Turbulent kinetic energy

Turbulent kinetic energy, k, is a way to quantify the intensity of the turbulence in a flow [28]. It can be written as the sum of the normal Reynolds stresses [27]:

$$k = \frac{1}{2}\left(\overline{u'^2} + \overline{v'^2} + \overline{w'^2}\right) = \frac{1}{2}\overline{v_i' v_i'} \tag{2.5}$$

The turbulent kinetic energy is thus a way to measure the amount of turbulence there is in the flow. If it is high, it is reasonable to expect a lot of turbulence and the opposite if it is low [28].

2.2.2 Turbulent dissipation

Turbulent dissipation is the process of turbulent kinetic energy turning into thermal energy and is a common occurrence in turbulent flow [27]. The transport equation for the turbulent dissipation rate, ε, is expressed as follows:

$$\varepsilon = \nu \overline{\frac{\partial v_i'}{\partial x_j} \frac{\partial v_i'}{\partial x_j}} \tag{2.6}$$

Where ν is the eddy viscosity. From the turbulent dissipation rate, the specific turbulent dissipation rate, ω, can also be expressed as the quotient between the dissipation and the turbulent kinetic energy as [27]:

$$\omega = \frac{\varepsilon}{k} \tag{2.7}$$

2.2.3 Turbulent viscosity

The turbulent kinematic viscosity, ν_t, is a parameter which is used in turbulence models to represent the momentum transfer which is enhanced by the turbulent eddies. Instead of resolving every fluctuation, the turbulent viscosity treats turbulence as an additional diffusive effect which increase the viscosity apparent in the flow. This concept is fundamental in many Reynolds-averaged Navier-Stokes (RANS) models where the turbulent viscosity is used as an extra equation to close the governing equations and predict how turbulence influences the mean flow [27]. It can be computed in different ways depending on which turbulence model is used.

2.3 Reynolds-Averaged Navier-Stokes models (RANS)

Solving the Navier-Stokes equations for turbulent flow is so complex it is practically impossible [29], however, there are ways to solve them by simplifying them. One way to go about it is decomposing the instantaneous variables, such as velocity or pressure, into two parts; a mean and a fluctuating, such as [27]:

$$v_i = \bar{v}_i + v_i' \tag{2.8}$$

The bar in the equation above denotes the mean, or time-averaged value which can be defined as:

$$\bar{v} = \frac{1}{2T}\int_{-T}^{T} v dt \tag{2.9}$$

After time averaging, eq. 2.8 becomes:

$$\bar{v}_i = \bar{\bar{v}}_i + \overline{v_i'} \rightarrow \bar{\bar{v}}_i = \bar{v}_i \rightarrow \bar{v}_i = \bar{v}_i + \overline{v_i'} \tag{2.10}$$

The equation above thus defines $\overline{v_i'} = 0$, meaning that the time-averaged fluctuations are zero [27]. Applying this process and time-averaging the Navier-Stokes equations gives the Reynolds-averaged Navier-Stokes (RANS) equations, which, together with the continuity equation, are written as below for incompressible flow [27]:

$$\frac{\partial \overline{v_i}}{\partial x_i} = 0 \tag{2.11}$$

$$\rho \frac{\partial \overline{v_i v_j}}{\partial x_j} = -\frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j}\left(\mu \frac{\partial \overline{v_i}}{\partial x_j} - \rho \overline{v_i' v_j'}\right) \tag{2.12}$$

Decomposing the variables and simplifying the equations in this way makes it possible to solve the equations for a turbulent flow. However, the equations above introduce a new term, $\overline{v_i'v_j'}$, known as the Reynolds stress tensor, which is symmetric and depicts the correlations between the fluctuating velocities [27]. The RANS equations thus describe equations which are open, and in order to solve them, new equations are needed to model the new unknown terms. These new equations are known as turbulence modeling [29]. Due to the number of unknowns, which are ten (pressure, three velocity components and six Reynolds stresses), exceeding the number of equations, which are four (three Navier-Stokes equations and the continuity equation), this system can't be solved as is. A model for the stresses is needed in order for the system to be closed [27].

There are multiple different ways to model the Reynold stresses, these are sorted into zero-equation, one-equation, two-equation and Reynolds stress models (RSM) [27]. In this thesis, only one-equation and two-equation turbulence models will be discussed.

2.3.1 One equation turbulence model - Spalart-Allmaras

Spalart-Allmaras is a one-equation turbulence model based on the RANS equations, which is designed to predict the turbulent viscosity, ν_t, with a relatively low computational cost. The model is widely used in aerodynamic applications as it is simple to implement, robust and often performs well for attached boundary layers and external flows. Due to being less costly in both time and computational resources than the other two-equation models, it is a popular choice. However, it is less accurate in more complex and highly turbulent flows. The Spalart-Allmaras model is an eddy viscosity model which contains a transport equation for the turbulent viscosity [30].

2.3.2 Two equation turbulence models

Two equation models are among the most commonly used turbulence models and include the $k-\varepsilon$ and $k-\omega$ models. These have become an industry standard and are used for most types of problems, but they are still actively being researched and the development of new and refined two-equation models is still ongoing. As the name suggests, these equations model the turbulent properties by adding two extra transport equations to the RANS equations. These extra equations enables the model to account for different historical effects of turbulent energy, such as convection and diffusion. Commonly, one of the new transported variables is the turbulent kinetic energy, k, and the second depends on which model is used. For the two examples cited above, the second variable is either the turbulent dissipation rate, ε, or the specific turbulence dissipation rate, ω. The first variable, k, determines how much energy there is in the turbulence while the second variable determines either the length-scale or time-scale of the turbulence for ε and ω, respectively [31].

The following equation is the modeled equation for k [27].

$$\frac{\partial k}{\partial t} + \bar{v}_j \frac{\partial k}{\partial x_j} = \nu_t \left(\frac{\partial \bar{v}_i}{\partial x_j} + \frac{\partial \bar{v}_j}{\partial x_i} \right) \frac{\partial \bar{v}_i}{\partial x_j} + g_i \frac{\nu_t}{\sigma_\theta} \frac{\partial \bar{\theta}}{\partial x_i} - \varepsilon + \frac{\partial}{\partial x_j} \left[\left(\nu + \frac{\nu_t}{\sigma_k} \right) \frac{\partial k}{\partial x_j} \right] \quad (2.13)$$

The equation for ε is modeled as follows [27].

$$\frac{\partial \varepsilon}{\partial t} + \bar{v}_j \frac{\partial \varepsilon}{\partial x_j} = \frac{\varepsilon}{k} c_{\varepsilon 1} \nu_t \left(\frac{\partial \bar{v}_i}{\partial x_j} + \frac{\partial \bar{v}_j}{\partial x_i} \right) \frac{\partial \bar{v}_i}{\partial x_j} + c_{\varepsilon 1} g_i \frac{\nu_t}{\sigma_\theta} \frac{\partial \bar{\theta}}{\partial x_i} - c_{\varepsilon 2} \frac{\varepsilon^2}{k} + \frac{\partial}{\partial x_j} \left[\left(\nu + \frac{\nu_t}{\sigma_\varepsilon} \right) \frac{\partial \varepsilon}{\partial x_j} \right] \tag{2.14}$$

For ω, it is modeled using eq. 2.7. Computing the turbulent viscosity in a two equation model depends on which model is used:

$$k - \varepsilon : \nu_t = C_\mu \frac{k^2}{\varepsilon} \tag{2.15}$$

$$k - \omega : \nu_t = \frac{k}{\omega} \tag{2.16}$$

Where C_μ=0.09 and is an experimentally defined constant.

2.4 Shear Stress Transport (SST)

The SST model, or Shear Stress Transport model, is an eddy-viscosity model and is a combination of the previously mentioned $k - \omega$ and $k - \varepsilon$ models. Both models have limitations in different flows and thus, by combining them, the result becomes more accurate and can be used in a broader setting. The $k - \varepsilon$ model tends to over-predict shear stress in flows with adverse pressure gradients (APG), and also requires near-wall modifications. Using a $k - \omega$ model in these regions is a better choice as they predict APG flows better. However, these models are worse in the free-stream as they are dependent on a free-stream value of ω. This means that by combining the two models, the resulting model works in both applications [27].

2.5 Large Eddy Simulations (LES)

Large Eddy Simulation (LES) is an approach in turbulence modeling which resolves the large, energy-containing eddies and modeling only the smallest scales. This process of capturing the dominant and unsteady flow structures allows LES to represent turbulent physics more accurately than steady RANS methods, whihc is especially important in flows with strong separation, unsteadiness and complex vortices. LES is thus an approach that is highly attractive for detailed flow analysis, although it requires more computational resources due to the demand for a finer mesh and smaller time steps [27].

2.6 Statistics

Statistics is the study of data, meaning the collection, summarization and the usage of samples to draw conclusions about a larger population. It aids in turning raw measurements into useful information, making it important in many different fields [32]. This thesis has used a few different statistical functions, which will now be explained.

An amplitude probability density (APD) function or probability density function (PDF) for a continuous function is defined as the probability of a variate to have the value x. The image below shows a plot of a PDF for a normal distribution [33].

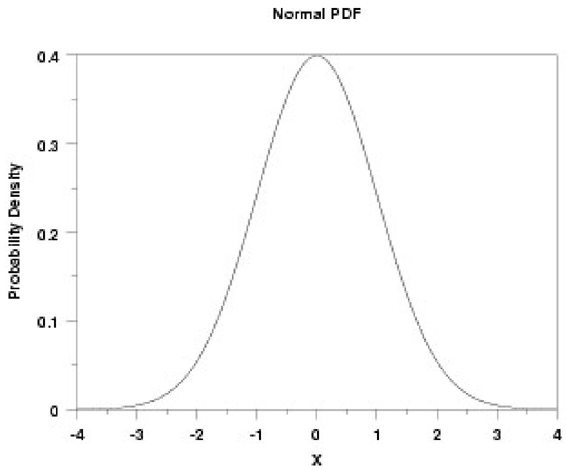

Figure 2.2: Probability density function [33]

A cumulative distribution function (CDF) is defined as the probability of a variable taking a value that is less than, or equal to x [33]. The function looks as follows for a normal distribution.

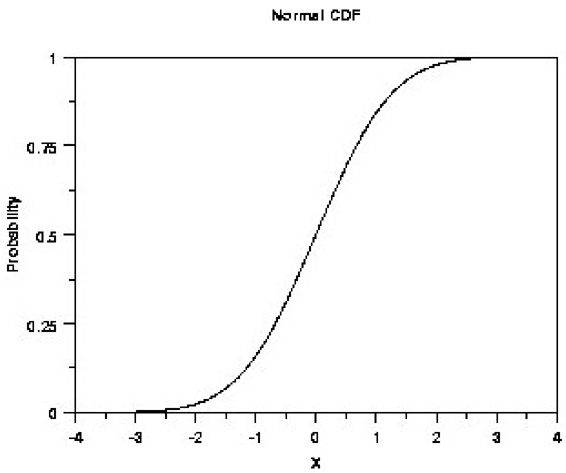

Figure 2.3: Cumulative distribution function [33]

The inverse of a CDF is called a percent point function (PPF), also known as the inverse distribution function, meaning that for a distribution function, the probability of the variable being less than or equal to x is calculated for a given x. In the case of the PPF, the probability is the starting point and then, for a cumulative distribution, the corresponding x is calculated [33]. The plot for the PPF for a normal distribution is shown below.

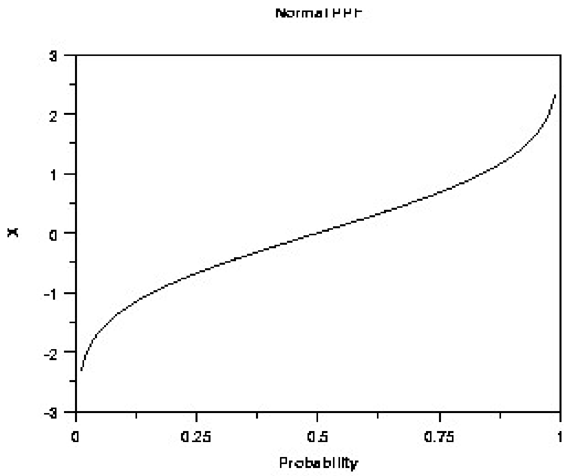

Figure 2.4: Percent point function [33]

2.7 Machine learning

Machine learning and artificial intelligence (AI) are often terms which are used interchangeably. This is not correct, instead machine learning is a subset of AI, such that "*All machine learning is AI, but not all AI is machine learning*" [34]. Machine learning algorithms uses pattern recognition on training data to accurately predict new data. It is based on mathematical logic and thus require numerical expressions of the characteristics of the data points. These data points are often presented in vector form where every element in a data point's vector embedding has a corresponding numerical value for a specific characteristic [34]. This means that for data types which are inherently numerical, like financial data, it is a straight forward process. But for data types that are not inherently numerical, like text or images for example, it is not as simple. Choosing which aspects of data are to be used in machine learning algorithms, is a process which is usually manual and is called feature selection.

There are three main types of machine learning, which all methods can be categorized in:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning (RL)

These different methods will now be explained.

1. Supervised learning

Supervised learning is an approach where a model is trained to predict an output based on an input. By learning the relationship between the inputs and outputs, the goal is for the model to make accurate predictions based on the learned patterns [34]. These algorithms are suitable for tasks which require accuracy, like regression;

predicting continuous values, or classification; predicting discrete values. In order for the model's output to be measured for accuracy, it is compared to a correct, or ideal, output for any received input. In traditional supervised learning, this correct output is supplied by labeled data. For example, training spam detection models for emails require labeling emails with "spam" or "not spam", and training an image segmentation model uses images where each pixel has been annotated by a category. The goal with supervised learning is tuning the parameters until its predictions align with the labeled targets.

In order to quantify the difference between the model's output and the correct output, a loss function is used. This is calculated over a batch of traning data and the goal with the training is to minimize this loss. After the loss has been calculated, different algorithms focused on optimization will work to identify the parameter adjustments which need to be made in order to minimize the loss. The name *Supervised learning* comes from the process needing a human to provide the correct output using annotation, and therefore, the usage of labeled data is what has defined supervised learning historically [34]. However, labeling data on large datasets can quickly become expensive and time-consuming, so other methods like self-supervised and semi-supervised learning have been created.

2. Unsupervised learning

Unsupervised learning is the opposite from supervised learning, as it doesn't require a known output. Instead, these types of algorithms notice complex patterns in data which has not been labeled. This can include patterns like similarities, potential groupings and correlations [34]. They are very useful in cases where the human eye might not notice these patterns. Most unsupervised learning methods employ one of the following functions:

- Clustering: These algorithms split unlabeled data into groups, or "clusters", according to either proximity or some similarity to each other. They are usually used for detecting fraud or market segmentation.
- Association: This type of algorithm can see correlations, such as between an action and specific conditions. These are commonly used by e-commerce companies to give recommendations based on what the customer has been looking at.
- Dimensionality reduction: As the name suggests, these algorithms can reduce the number of dimensions, or features, of a data point to reduce the complexity, while still maintaining the important characteristics of the data. This algorithm is very good when pre-processing data, as well as doing visualization or compression on data.

Consistently, unsupervised learning can be seen as 'optimizing itself' and the challenge becomes preprocesing data in an effective way and tuning hyperparameters properly as these can influence the training but are not learnable [34].

3. Reinforcement learning (RL)

Reinforcement learning (RL) works very differently from the two previous types, RL trains models universally by using trial and error. This is useful in applications like robotics and video games for example, where the possible approaches and solutions are very broad, open-ended or difficult to define. An AI system is usually referred to as an 'agent' in RL literature. In comparison to supervised and unsupervised learning, the goal with RL is not to minimize the error but instead to maximize the reward. This is done through state-action-reward data tuples which are interdependent where the 'state space' has all the information available which is relevant to any decision made by the model. In the 'action space' all decisions the model can make in a specific moment are contained. After these, there is the reward signal, which is the feedback the model receives after performing an action. The fourth component is a policy which acts as the agents thought process, it drives the behaviour [34].

One way to do machine learning is by using artificial neural networks (ANN), these are explained below. Machine learning and ANNs are thus not the same thing but instead ANNs are a subset of machine learning [35].

2.7.1 Artificial Neural Network (ANN)

Artificial Neural Networks, or ANNs, use artificial neurons to mimic how the brain processes information. It uses these neurons to do data analysis, learn patterns and consequently predict new data. The networks solve problems through using multiple layers of connected neurons that work together. There are three main components in an ANN [36]:

- Input layer: This is where the data is fed into the network. The input layer can consist of multiple different features.
- Hidden layer: In this layer, calculations are performed by every neuron on the input data and the results are passed through to the next layer. There can be multiple hidden layers with different amount of nodes, and the more hidden layers a network has, the more complex patterns the network can process. In each layer, the data is tranformed into increasingly abstrcat information [36]. The mathematical operations computed in these layers are called activation functions [34].
- Output layer: This layer comes last and is the result of the data processing done by the previous layers. The networks final prediction is presented here.

The image below shows the architecture of the ANN.

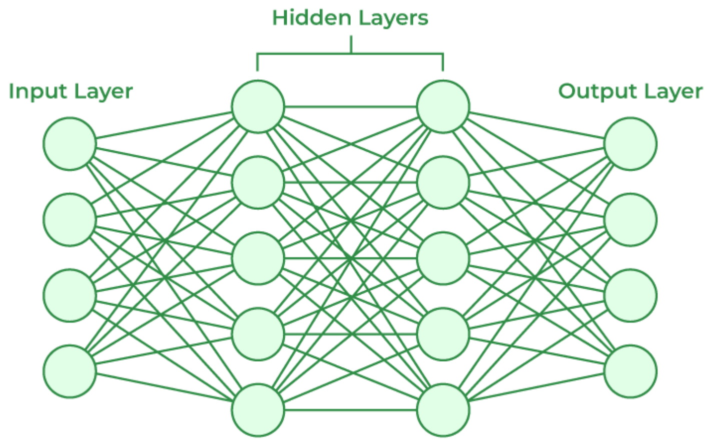

Figure 2.5: Neural network architecture [36]

In the image above, the different layers can be seen. The layers consist of nodes which, as previously discussed, perform different actions based on the layer. Between the nodes, there are connections that have weights. These weights control how much each node influences another connected node[36].

The training process works as follows; ANNs learn patterns by using a set of data to train, for example, teaching an ANN to recognize a dog is done through providing the network with thousands of pictures of dogs. The network will then train itself on these images and learn to recognize the features which define dogs. After training the network, it is tested to see if it recognizes dogs from new images. The output is compared to the label of the data and if the network is incorrect it adjusts the weights through a process called back-propagation. The process then repeats until the network accurately classifies the images [36].

In order to train a network, the set of data has to be divided into a training set, validation set and test set. This is usually done as 70%, 15% and 15% respectively [37].

2.7.1.1 Types of ANNs

There are different types of ANNs which behave in different ways and are used in different applications. This thesis will only use a back propagating artificial neural network (BP ANN).

2.7.1.1.1 Back Propagating Artificial Neural Network (BP ANN) A term that has been mentioned multiple times in the section above is backpropagation. This is a fundamental process in the training of neural networks and is commonly used in the industry [38]. The process of propagating the error calculated from the output back into the network allowing it to adjust the parameters is the foundation of training neural networks. The parameters are adjusted through calculating the gradients in respect to the loss function. This allows the optimizer to minimize the errors in an iterative manner [39]. The image below shows the process of back propagation in an artificial neural network.

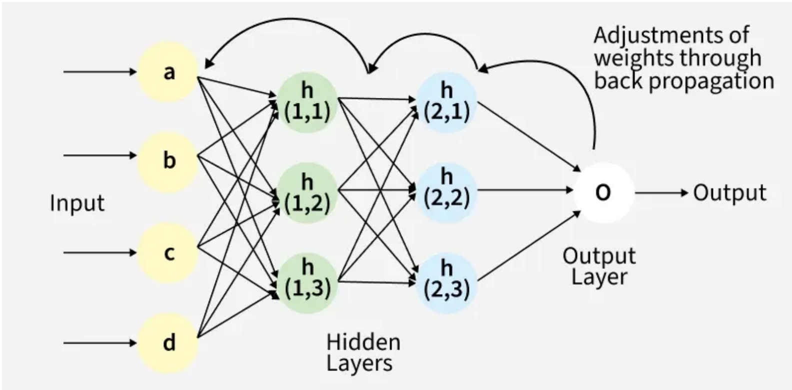

Figure 2.6: Illustration of the way back propagation works in a neural network [38]

2.7.1.2 Activation functions

As mentioned earlier, the nodes in the hidden layer uses activation functions to perform operations on the data [34]. These functions are important as they work to introduce non-linearity and help the network learn from complex patterns. There are different activation functions used in ANNs and these are a few of the common ones:

- Sigmoid Function: Used in tasks where binary classification is needed as it outputs values between 0 and 1. Binary classification tasks are tasks such as deciding if an image is of a dog or not.
- ReLU (Rectified Linear Unit): This activation function returns the input if it is positive and returns zero if it is negative. It is often used in hidden layers where it can be used to solve the issue of gradients vanishing due to becoming very small during back propagation.
- Tanh (Hyperbolic Tangent): Works similarly to the sigmoid function but instead outputs values ranging from -1 to 1. It is useful in hidden layers where a broader range is needed.
- Softmax: Used in multi-class classification tasks and converts the raw inputs into probabilities that are used in the last layer of a network.
- Leaky ReLU: Variant of ReLU which allows small negative values to "leak" through for inputs. This helps prevent so called "dead neurons" when training.

2.7.1.3 Optimization algorithms

Another property of ANNs are optimization algorithms [36]. These work during training to adjust the weights in order to minimize the errors. There are four key algorithms:

- Gradient Descent: This is the most basic algorithm for optimization. It calculates the gradient of the loss function and then updates the weights accordingly.

- Adam (Adaptive Moment Estimation): More efficient version of the gradient descent algorithm that uses deep learning to adapt the learning rate for each weight.
- RMSprop: Another variant of gradient descent that uses the average of recent gradients to adjust the learning rate. This algorithm is useful when training RNNs.
- Stochastic Gradient Descent (SGD): Is a fast but noisy optimization algorithm that updates the weights by using one sample at a time.

2.7.1.4 Challenges

There are a few challenges that come with working with ANNs [36]. The first is data dependency, which comes from the ANNs needing a lot of data that is very high quality. Gathering this data and cleaning it can be very time-consuming and expensive, which is impractical and not realizable in many industries. Another challenge is the heavy demand for computational resources which can also be expensive and in some cases is not possible. The risk for overfitting the training data is also a challenge. This happens when the model performs well on the training data but worse on the new, unseen test data. Another issue is the interpretability, as machine learning algorithms are often referred to as 'black boxes' they are hard to understand and it is not always clear how it came to a conclusion. This is an issue for industries such as healthcare where transparency is important. The last challenge is the training time, as complex ANNs with more layers often take a long time to train which can hinder development and makes them not suitable for time-sensitive applications [36].

2.8 Distortion Indices

A distortion index is a way to quantify the maximum distortion present at the aerodynamic interface plane (AIP) [10]. They are usually configured in a way which compares the maximum pressure loss to the average pressure over the AIP, although this is done in very different ways. It is very common for companies as well as government research centers to create their own engine specific indices; sometimes a new index is created for each new project. As a consequence, there are so many indices that there is confusion when comparing between different engine configurations. For this project, three distortion indices are used, DC60, CDI and RDI. These indices calculate the distortion in different ways and will give a good idea of how the synthesis performs.

2.8.1 DC60

DC60 is an index developed by Rolls Royce and is widely used in the industry. It quantifies the intake pressure distortion in an angular region, most often 60°, but for a modern compressor, an angular area of 90° can be used. The use of 60° is what gives DC60 its name [40]. In this report, the original use of 60° is employed through the following equation.

$$DC60 = \frac{\bar{P}_{AIP} - \bar{P}_{60_{MAX}}}{\bar{P}_{AIP}} \tag{2.17}$$

Where $\bar{P}_{AIP}$ is the average total pressure over the whole AIP and $\bar{P}_{60_{MAX}}$ is the maximum average total pressure in an angular area of 60° of the AIP.

2.8.2 Circumferential distortion index (CDI)

The circumferential distortion index (CDI) is also known as inlet distortion circumferential coefficient (IDC) and is one of the Society of Automotive Engineers (SAE) standard indices [11]. It looks at the circumferential distortion over the AIP and is calculated through first calculating the ratio of the lowest pressure regions in each ring with the average pressure over the whole ring. This is shown in the equation below where the index i is the ring [41].

$$CDI_i = \frac{\bar{P}_{0,i} - \bar{P}_{0,i_{LOW}}}{\bar{P}_{0,i}} \tag{2.18}$$

In the equation above, $\bar{p}_{0,i}$ is the mean total pressure of the ring and $\bar{p}_{0,i_{LOW}}$ is the lowest total pressure region in the ring. After calculating CDI for each ring, the value for CDI is calculated for the whole AIP through the following equation [41].

$$CDI = max\left(\frac{CDI_i + CDI_{i+1}}{2}\right) \tag{2.19}$$

2.8.3 Radial distortion index (RDI)

Radial distortion index (RDI) is also one of the Society of Automotive Engineers (SAE) standard indices, and is also known as inlet distortion radial coefficient (IDR) [11], [40]. It measures the radial distortion and is defined as follows [41]:

$$RDI = \frac{max(\bar{P}_{0,AIP} - \bar{P}_{0,i})}{q_{AIP}} \tag{2.20}$$

Where $\bar{P}_{0,AIP}$ is the average total pressure over the AIP, $\bar{P}_{0,i}$ is the average total pressure in each ring, the index has the same function as for CDI, and lastly q_{AIP} is calculated using this formula [41]:

$$q_{AIP} = \frac{\gamma}{\gamma - 1}\bar{P}_{AIP}\left(\left(\frac{\bar{P}_{0,AIP}}{\bar{P}_{AIP}}\right)^{\frac{\gamma-1}{\gamma}} - 1\right) \tag{2.21}$$

In the equation above, γ is the specific heat constant, which for air is 1.4 and $\bar{P}_{AIP}$ is the average static pressure over the AIP [41].

3
Methods

This chapter describes all five methods employed in this thesis to predict the dynamic distortion.

3.1 Wind tunnel experiments

Wind tunnel data is used to validate the synthesized results, and it includes multiple different configurations and flow conditions. The airflow into the engine is evaluated by defining an aerodynamic interface plane (AIP), located in the duct just upstream of the engine, which provides valuable insight into the inlet flow conditions. This way of measuring is standard in intake wind tunnel experiments, and the pressure data is collected from multiple sensors, both steady-state and dynamic. There are 40 steady-state sensors and 24 dynamic sensors which are organized in eight rakes and five or three rings respectively, as can be seen in Fig. 3.1 below. The steady-state and dynamic sensors are thus placed in different positions.

3.1.1 Pre-processing the wind tunnel data

An integral part in the following methods for synthesizing pressure is calculating the total pressure, i.e. summing the steady-state and dynamic (fluctuating) pressures. It was calculated by interpolating the steady-state pressure onto the dynamic pressure. Due to there being 40 steady-state sensors and only 24 dynamic sensors, the dynamic data needed to be interpolated into 'ghost points' onto rings 2 and 4.

The 'ghost points' were interpolated by taking the mean value of the neighboring dynamic probes in the other rings as shown in eq. 3.1 below where i is the ring number.

$$P_{DYN,i} = \frac{P_{DYN,i-1} + P_{DYN,i+1}}{2} \tag{3.1}$$

The same process was then done to interpolate the steady-state pressure onto the dynamic pressure of the same ring and these were then added together to create the total pressure. Meaning that the final configuration of sensor placement looks as follows in Fig. 3.1.

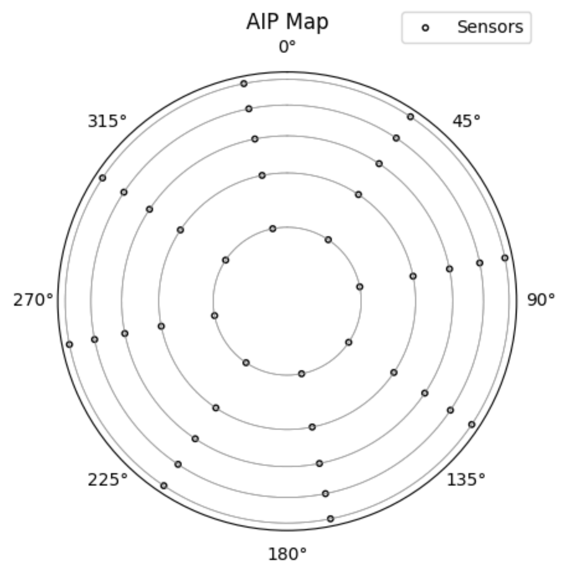

Figure 3.1: Illustration of final sensor placement after processing data

The following image shows the same configuration but with the rakes and rings named.

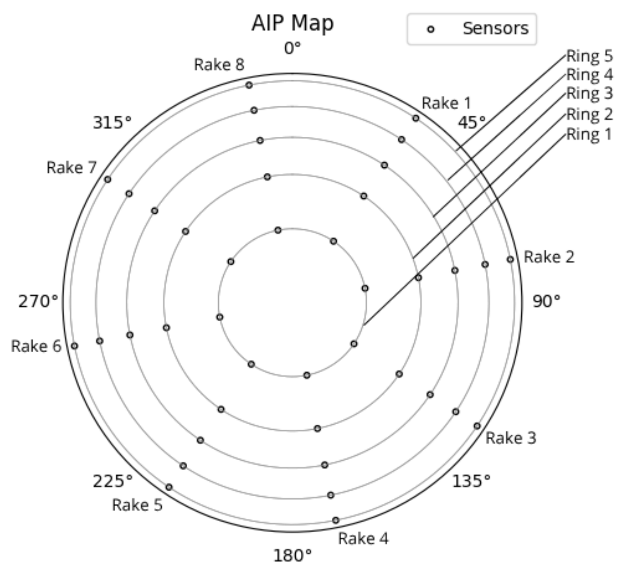

Figure 3.2: Sensor placement showing numbers of rakes and rings.

3.2 Method 1 - Predicting dynamic pressure from RMS turbulence and a random number generator

The first method was devised by D. Motycka [16] and is based entirely on wind tunnel data, as written in section 1.1 above. It requires two sets of data as inputs, the steady-state pressure as well as the filtered RMS turbulence in the intake. It also requires a list of random numbers that are used to synthesize the fluctuating component. Figure 3.3 below shows the process.

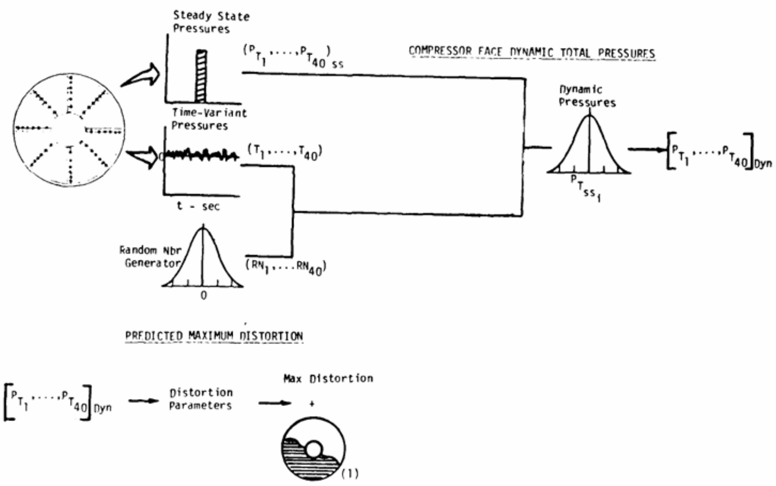

Figure 3.3: Original method by Motycka [8]

The random numbers are combined with the RMS turbulence values to create the synthesized dynamic pressure. By adding this new pressure to the steady-state pressure, the dynamic pressure can be calculated. After that, the maximum distortion can be calculated using a distortion index [8].

Figure 3.3 shows the theoretical process of synthesizing, but in practice, the process looks different. The image below shows the full process of the Motycka method [42]. The process works as follows; first, an amplitude probability density (APD) curve gets created for every probe based on the RMS turbulence readings and steady-state total pressure. By assuming normal distribution, the curve is generated and can be seen in Fig. 3.5, where σ is the RMS turbulence. The APD function is then integrated to create a cumulative probability function ($\sum$ APD) which can be seen in Fig. 3.6. In the figure, it can be seen that the random numbers are scaled to a range which spans from 0 to 1 on the y-axis. These random numbers get transformed into pressure readings, and this process repeats for every probe.
The synthesization is done in Python, by initially defining μ as the steady-state total pressure and σ as the RMS turbulence. The steady-state total pressure becomes the center of the distribution. Due to the assumption of the APD being normally distributed, it can be defined as:

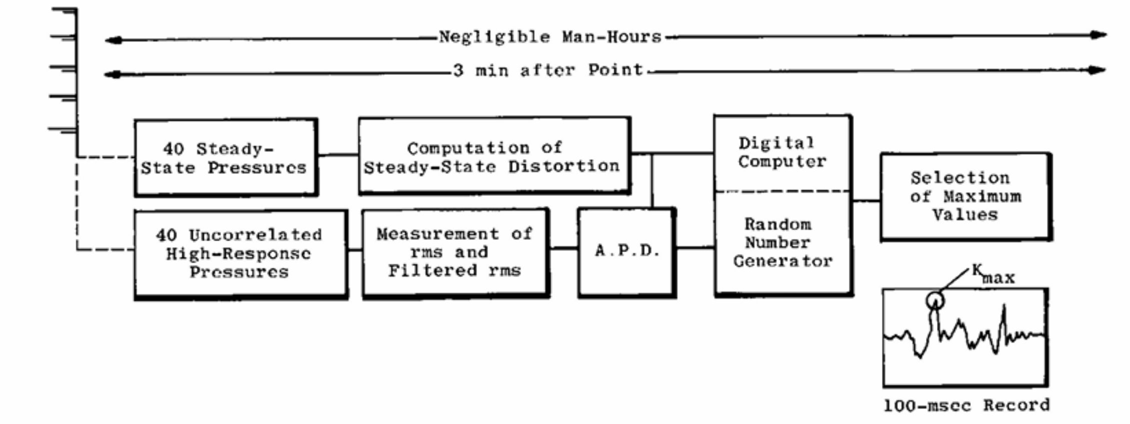

Figure 3.4: Statistical method by Motycka for synthesizing time-variant data from distortion [42]

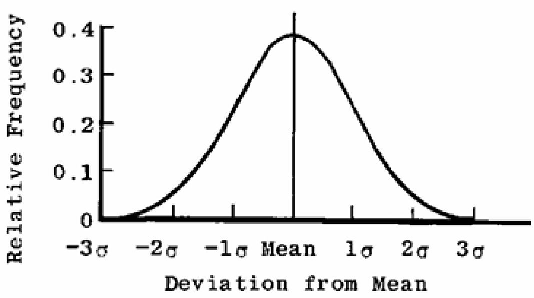

Figure 3.5: Amplitude probability density function (APD) [42]

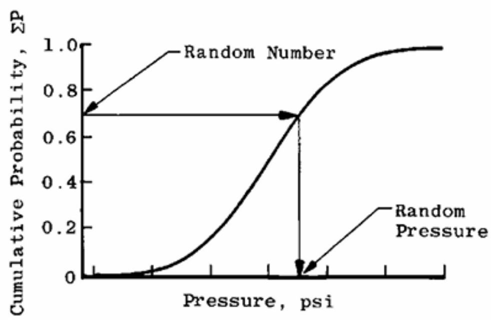

Figure 3.6: Cumulative probability with random number process [42]

$$f(p) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(p-\mu)^2}{2\sigma^2}} \tag{3.2}$$

Where £ is a linearly spaced pressure axis that ranges from $\mu - 4\sigma$ to $\mu + 4\sigma$. This range is based on standard normal distribution. In order to get the correct amount of fluctuating pressures, the random values are created based on the frequency and time of the dynamic readings from the wind tunnel data. This is done to make comparison easier. The synthesization is then done through a percent point function, which maps the random numbers to the APD. This process is repeated for every probe in every measurement and a new file is created with the same number of rows as the wind tunnel data.

3.3 Method 2 - Estimating RMS turbulence from CFD data

Method two comes from NASA, devised by W.P. Norby, J.A. Ladd and A.J. Yuhas in 1996 [9]. It builds on the first method of distortion synthesis using a random number and wind tunnel data, but due to the limitations of having to do wind tunnel experiments, as well as the development of computational fluid dynamics (CFD), this method was derived in the late 90's. It replaces the RMS turbulence by finding a correlation between it and the kinetic energy, k, from steady-state RANS simulation. The correlation is derived as follows [9]. As seen before, the turbulent kinetic energy, k, is defined according to eq. 2.5, which can be rewritten assuming isotropic turbulence:

$$k = \frac{3}{2}\overline{u'^2} \tag{3.3}$$

Turbulence in the inlet of an aircraft can be defined as the root mean square of the total pressure fluctuations at the ith probe and then normalized using the average steady-state total pressure over the plane.

$$Turb = \frac{1}{P_{t_{SS}}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{t_i} - P_{t_{SS}})^2} \tag{3.4}$$

As can be seen in eqs. 3.4 and 2.5 above, although both are defined by fluctuations, velocity fluctuations define the turbulent kinetic energy while total pressure fluctuations define the inlet turbulence. Although these are different, a fluctuation in total pressure will result in a corresponding fluctuation in velocity. This relation is what the correlation between the CFD data and the RMS turbulence will be built on.
The following equation, 3.5, is a model equation created as a starting point for the correlation [9], derived by adding the contributions to the turbulence. The equation contains four constants C_{1-4} that need to be determined.

$$\frac{Turb \cdot P_t}{\rho} = C_1 k + C_2(u+a)\sqrt{k} + C_3(u-a)\sqrt{k} + C_4 u\sqrt{k} \tag{3.5}$$

The first term on the right-hand side, $C_1 k$, describes the turbulence which results from the energy that gets extracted from the mean flow by the large scales and the interaction between the large and small scales. It represents contribution that comes from non-linear turbulence models and is expected to be smaller than from the other terms. The next two terms, $C_2(u+a)\sqrt{k}$ and $C_3(u-a)\sqrt{k}$, come from the turbulence interacting acoustically. This interaction is at a frequency which is relatively high in comparison to the engine response, the contribution is expected to be insignificant thanks to the low pass filtering which is performed on the data before calculating the RMS turbulence. The fourth term's, $C_4 u\sqrt{k}$, contribution comes from linear convected turbulence, i.e. turbulence from generators like separation at the intake and vortice ingestion.

Based on numerical experiments done in the report by Norby et. al. [9], the first three contributions can be neglected. This leads to the model equation for turbulence being

$$Turb = \frac{C}{\gamma \cdot g_c}\frac{u\sqrt{k}\rho}{P_{t_{SS}}} \tag{3.6}$$

Where C is the correlation coefficient, γ is the heat capacity for air, g_c is the acceleration of gravity, u is the axial velocity in x direction, k is the turbulent kinetic energy, ρ is the density and $P_{t_{SS}}$ is the steady-state total pressure. This equation can thus be used to calculate the predicted turbulence based on CFD values. The correlation coefficient is however unknown, meaning that it needs to be determined in order to calculate the turbulence. This is done through the least square method by first defining an error function, ε_i, based on the model equation as

$$\varepsilon_i = \frac{C}{\gamma \cdot g_c} u_i k_i^{1/2} - \frac{Turb_i \cdot P_{t_{SS_i}}}{\rho_i} \tag{3.7}$$

In the above equation, the $Turb_i$ component is the RMS measured turbulence from the wind tunnel in the ith probe. The equation contains values $\gamma = 1.4$ and $g_c = 9.81 m/s^2$ which are the heat capacity and gravitational acceleration respectively. The goal of the least square method is to determine a value of C that finds the lowest sum of squares of the error terms, as can be seen below.

$$\frac{\partial \left(\sum_{i=1}^{N} \varepsilon_i^2 \right)}{\partial C} = 0 \tag{3.8}$$

By substituting eq. 3.7 into eq. 3.8, the following equation for C can be derived.

$$C = \gamma g_c \frac{\sum_{i=1}^{N} \frac{u_i k_i^{1/2} Turb_i P_{t_i}}{\rho_i}}{\sum_{i=1}^{N} u_i^2 k_i} \tag{3.9}$$

After calculating the correlation coefficient, C, it can be used in eq. 3.5 and be used instead of the RMS turbulence in the synthesization described in section 3.2.

The CFD simulation is performed with an SST turbulence model, from which the turbulent kinetic energy is gathered. The inputs; u, k, ρ and $P_{t_{SS}}$ are all taken from the CFD data while the RMS turbulence, $Turb$, is taken from the wind tunnel data. C is calculated for every probe from the CFD data and a cumulative value for C is determined. This value is used to calculate the RMS turbulence and the synthesization method from the baseline synthesizes the fluctuating pressure. The data size for this method consisted of 9 points.

3.4 Method 3 - Using machine learning and CFD data to estimate RMS turbulence

The third method uses the same non-linear correlation found by the previous method, but instead of using the least square method, it uses machine learning [19]. Due to the non-linearity of the correlation between the measured RMS turbulence and turbulent kinetic energy from CFD a machine learning algorithm is chosen to find it.

The inputs to the neural network are:

- u: Axial velocity in each probe
- a: Sonic speed
- k: Turbulent kinetic energy
- ρ: Density

- $P_{t_{SS}}$: Steady-state total pressure

These parameters all contribute to the turbulent fluctuations such that $Turb = f(u, a, k, \rho, P_{t_{SS}})$. Parameters; u, k, ρ and $P_{t_{SS}}$, come from the CFD simulation, and a is calculated using the following formula:

$$a = \sqrt{\gamma RT} \tag{3.10}$$

Where the heat capacity for air, $\gamma = 1.4$, the gas constant for air, R = 287 J/kg·K and the temperature is taken from the CFD solution. All inputs are taken individually from every probe.

The machine learning model chosen is a back-propagating artificial neural network as this is a common algorithm used widely in the industry. It has 5 parameter inputs with 40 values in each, one for every probe on the AIP, meaning that there are 200 inputs to the network. The output has one parameter, RMS turbulence, and this value is desired in every probe leading to 40 output values. The full dataset consists of 9 data points which come from the CFD solution, and these were solved with an SST turbulence model. The 9 points were split into training, validation, and testing datasets of 5, 2, and 2, respectively, with 40 values for each variable in each data point.

Due to the inherent non-linearity of the correlation which is modeled, three hidden layers with 256, 128 and 128 neurons respectively was chosen through testing and evaluation. It uses a rectified linear unit (ReLU) as activation function and an adaptive moment estimation optimizing algorithm.

3.5 Method 4 - Testing the machine learning approach on a one-equation turbulence model

When using a one-equation turbulence model as Spalart-Allmaras [30], the turbulent kinetic energy, k, and dissipation rate, ω, are not modeled. Instead, as mentioned above, the turbulent viscosity, ν_t, is modeled and can be used as a parameter. Due to the increased simplicity between SST model and Spalart-Allmaras, it is highly beneficial to find a way to do the synthesisation by using ν_t instead of k. As can be seen in equation 2.15, in section 2.3.2, there is a correlation between turbulent kinetic energy, k, and turbulent viscosity, ν_t, as it scales with $\frac{1}{\omega}$. The non-linear correlation between them is the basis behind this method. This method is expressed $Turb = f(u, a, \nu_t, \rho, P_{t_{SS}})$ and as can be seen, there are the same amount of inputs as the previous section. The sonic speed, a, is also calculated in the same way as previously defined. This dataset consisted of 1187 points which were split into training, validation and testing as 829, 179, 179. This is still a small dataset and the same machine learning model from the third method with the same settings was used for these as well.

3.6 Method 5 - Machine learning using only steady-state pressure

Similarly to the previous section, this is a sub-method based on the machine-learning algorithm from method 3, but instead of using the 5 parameters as inputs, only one, steady-state pressure, $P_{t_{SS}}$, was used. This was done as a test to see if it was possible to model the turbulence using only steady-state pressure, this would remove the need for any turbulence model in the CFD simulations and is thus something that was worth investigating. The expression for this method becomes $Turb = f(P_{t_{SS}})$. Similarly to the two previous sections, the machine learning model stayed the same and the data size for this was 1187, with the division into training, validation and testing as 829, 179, 179. This is the same set of data as was used for the previous method.

3.7 Validation of results

In order to compare the accuracy of the methods, the difference between the true pressure and predicted pressure is calculated. This is done in all probes for every time step as well as for the total method using the following formula:

$$\text{diff} = 1 - \frac{(P_{total} - P_{synt})}{P_{total}} \tag{3.11}$$

Where the total pressure, P_{total}, comes from the correct wind tunnel data and P_{synt} is the synthesized pressure. This was calculated for each probe in every time-step of the time-accurate data and then the mean value was taken of every case and file for each method to get a subsequent mean accuracy for every method.

4
Results

In this section, the results for each method will be presented and discussed, this includes the pressure differences and accuracy for two different cases, as well as the distortion indices. The cases have different flow properties and were simply chosen due to being the only cases left after splitting the data points for the third method. The methods will also be compared and evaluated on three criteria. Possible sources of error and future work is also discussed.

4.1 Method 1 - Predicting dynamic pressure from RMS turbulence and a random number generator

The first method performed well and achieved an accuracy of 98.83%, indicating that the synthesized dynamic pressure closely matches the measured data, this suggests that using this method produces reliable results. The accuracy was the highest achieved by any method in the thesis and will be used as a baseline to compare the other results.

4.1.1 Case one

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the first case using method one.

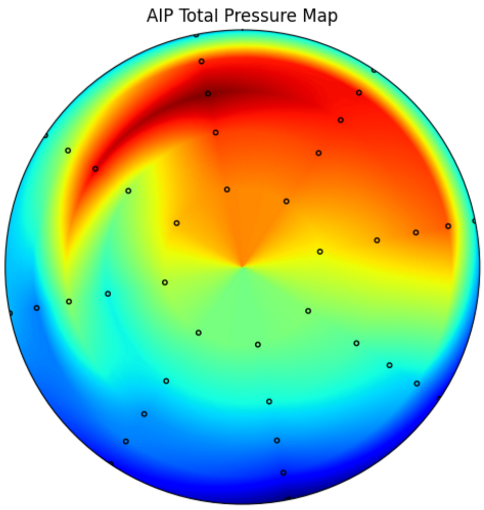

Figure 4.1: Comparison between wind tunnel data and synthesized data for case one using the first method.

The next figure shows the accuracy of the synthesized pressure using method one in relation to the total pressure from the wind tunnel for the first case.

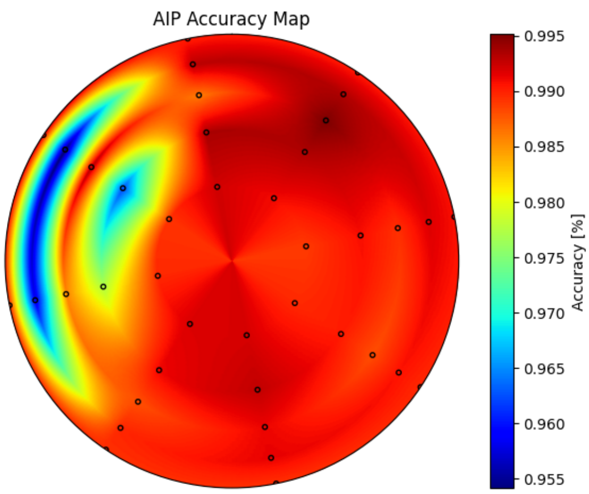

Figure 4.2: Plot of accuracy for method one case one

It can be seen that it generally works well but generalizes the high-pressure zone into a half circle and misses the low-pressure zone at rake seven. The synthesization fails to catch the small variations in pressure and appears to 'smooth over' some fluctuations. The accuracy map, Fig. 4.2, tells the same story as the pressure map, where the worst accuracy appears on the left-hand side, replacing the strong gradient with an area of relatively high pressure. The low-pressure zone looks similar to the wind tunnel data, but the high-pressure area in the top left corner gets increased in size by the synthesization to encompass all three rakes in that location.

4.1.2 Case two

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the second case using method one.

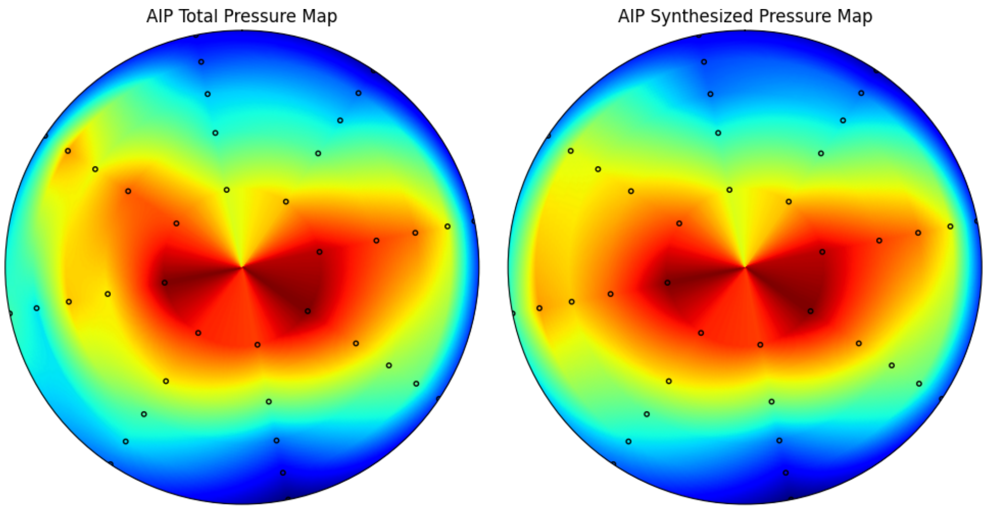

Figure 4.3: Comparison between wind tunnel data and synthesized data for case two using the first method.

This figure shows the accuracy of the synthesized pressure using method one in relation to the total pressure from the wind tunnel for the second case.

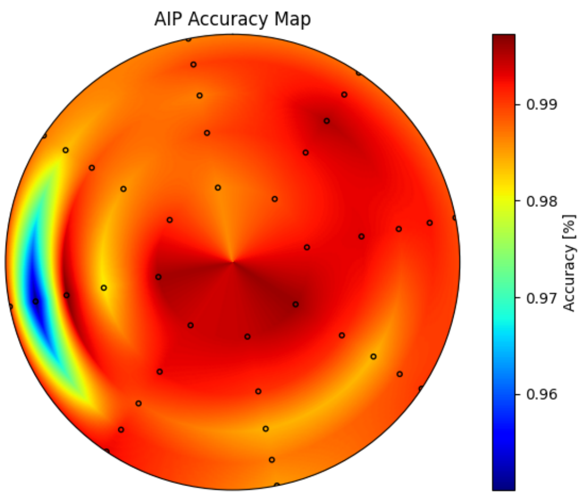

Figure 4.4: Plot of accuracy for method one case two

The second case, Fig. 4.3, shows a visible error in pressure synthesization but at rakes six and seven. In this case, however, the pressure is not generalized over a larger area, and the pressure contours don't look as smooth. The synthesization assumes a higher pressure closer to the boundary and misses the region of lower pressure just under rake six. The small, high-pressure area at rake six, probe four, is also erased in the synthesization. This is most likely due to the synthesization generalizing these areas, as has been seen and discussed previously. In the map of accuracy, Fig. 4.4, this inaccuracy can be seen where the worst accuracy occurs in the lower left corner. The accuracy in this region reaches around 95%, which is almost the same as case one and indicates that the overall synthesization can still be considered accurate. The highest accuracy region reaches almost 100%, which, due to this method being based on the RMS turbulence, could indicate that the pressure remains fairly constant during the whole sample. Just inside the high accuracy zone, there is a small zone of lower accuracy; this is also a consequence of the synthesization not picking up the small variations in pressure. By looking at the synthesized pressure AIP map, the general impression of this method of synthesization is that it generalizes and misses the small fluctuations in pressure. It has the same general areas but looks smoother, and the transitions between pressure areas are more linear.

4.1.3 Distortion indices

The following plots were created based on the distortion index equations in chapter 2. The plots all show the difference between the synthesized data from method one and true data from the wind tunnel. The blue line shows where the synthesized and actual data are equal.

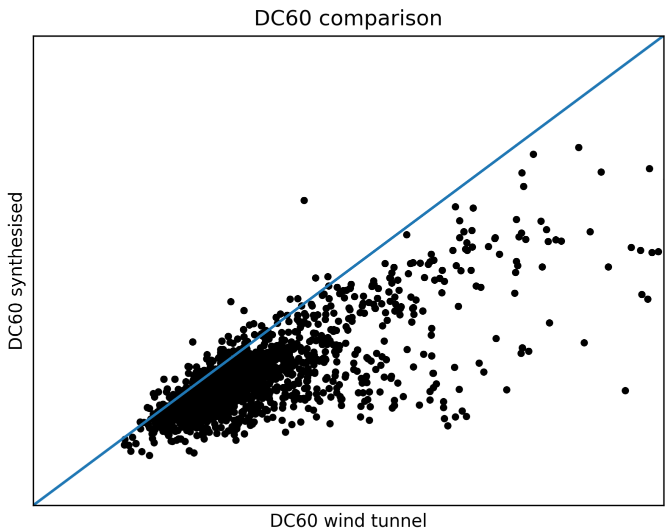

Figure 4.5: Comparison between wind tunnel data and synthesization for distortion index DC60 for the first method.

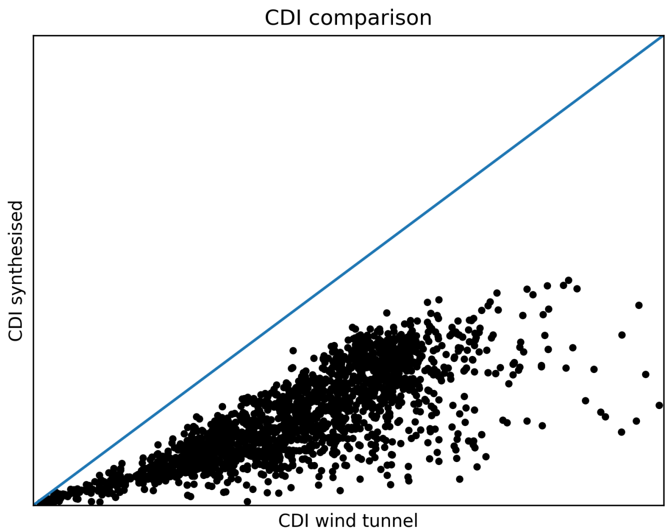

Figure 4.6: Comparison between wind tunnel data and synthesization for distortion index CDI for the first method.

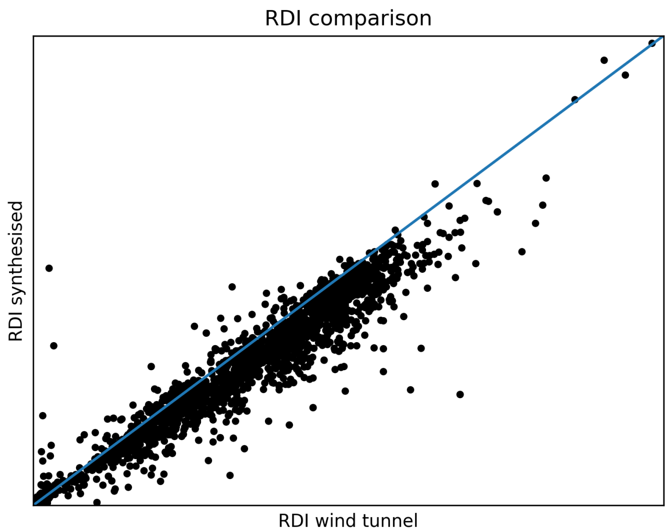

Figure 4.7: Comparison between wind tunnel data and synthesization for distortion index RDI for the first method.

Looking at the plot for DC60 for this method, Fig. 4.5, the data points are positioned slightly under the line of symmetry. Although a few points are above, most of the points are located underneath, indicating that the synthesization has a tendency to under-predict the maximum distortion. DC60, as previously mentioned, is an index that looks at the maximum distortion in a 60^o sector. The distortion over the whole AIP is thus generalized, and the results show that the synthesized data is slightly lower than the measured data. This indicates that the method performs well overall but fails to capture the extreme values, which is consistent with the calculated accuracy. The tendency to under-predict is also seen in the CDI plot, Fig. 4.6, where the data points are all located underneath the symmetry line. It appears to follow a linear curve with a lower inclination than it should. This is in stark contrast with the plot for RDI, Fig. 4.7, where a majority of the data points are on, or just below, the symmetry line. The synthesization is thus better at predicting radial than circumferential distortion.

4.1.4 Conclusion

This method is, as previously mentioned, the baseline for comparison with the other methods in this thesis. As these results are based on the wind tunnel values, they can be seen as the correct solution, and the rest of the methods should be as close to this as possible. The choice of using this as a baseline and building upon this using CFD data was made in the beginning of the thesis, and not trying other methods for the synthesization procedure was a decided delimitation. The results were generally good, with a high accuracy in pressure and good correspondence in radial distortion, but had obvious shortcomings in circumferential distortion and general (DC60) distortion. Another issue with this method is the complete reliance on wind tunnel data. This is a large disadvantage since doing wind tunnel tests for every new aircraft or intake design is costly and time-consuming.

4.2 Method 2 - Estimating RMS turbulence from CFD data

The second method achieved a similar overall accuracy to the first, at 98.82%, and is therefore also a good method for synthesizing distortion. The close agreement with the baseline indicates that both methods produce comparable results.

4.2.1 Case one

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the first case using method two.

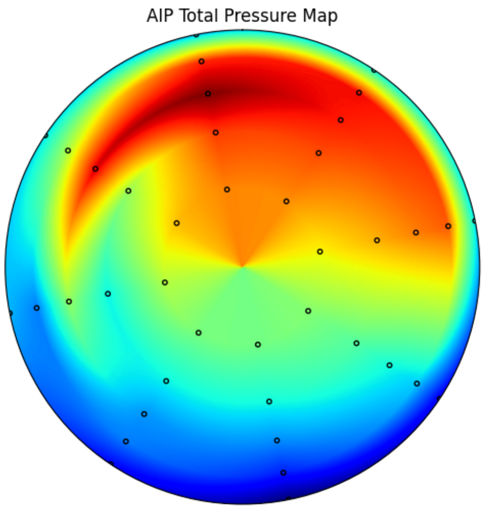

Figure 4.8: Comparison between wind tunnel data and synthesized data for case one using the second method.

This figure shows the accuracy of the synthesized pressure using method two in relation to the total pressure from the wind tunnel for the first case.

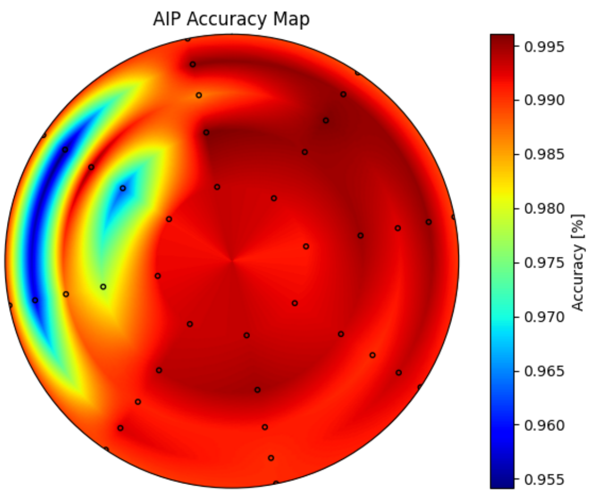

Figure 4.9: Plot of accuracy for method two case one

The pressure plots over the AIP for the first case look very similar to the baseline, which makes sense with the similar accuracy. Since the synthesization process is the same between this method and the previous, it still generalizes the high-pressure zone and misses the lower-pressure zone at the seventh rake. The accuracy, Fig. 4.9, looks very similar to method one, and it has a similar low accuracy zones on the left-hand side and a zone of lower accuracy in the top left corner. This suggests that this method of using CFD data is very accurate as a substitute for the RMS turbulence from the wind tunnel.

4.2.2 Case two

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the second case using method two.

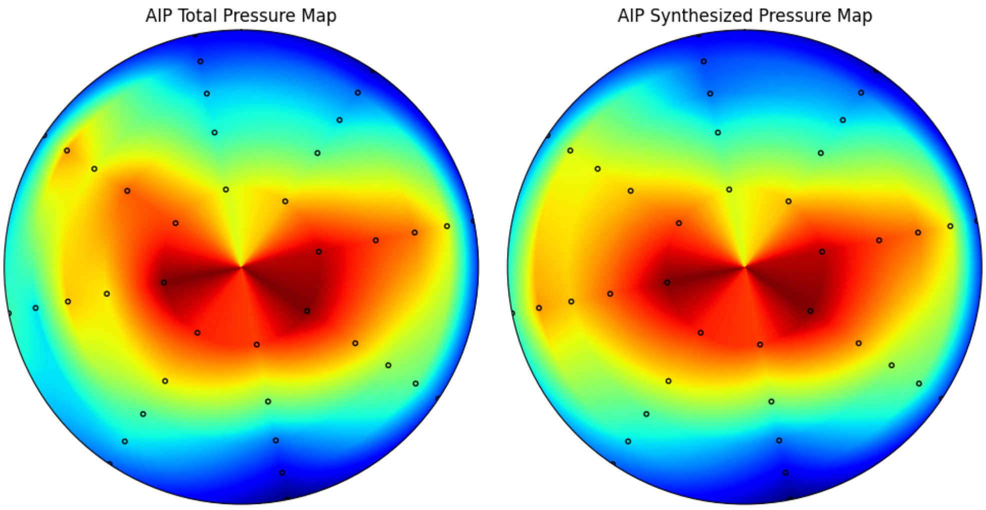

Figure 4.10: Comparison between wind tunnel data and synthesized data for case two using the second method.

This figure shows the accuracy of the synthesized pressure using method two in relation to the total pressure from the wind tunnel for the second case.

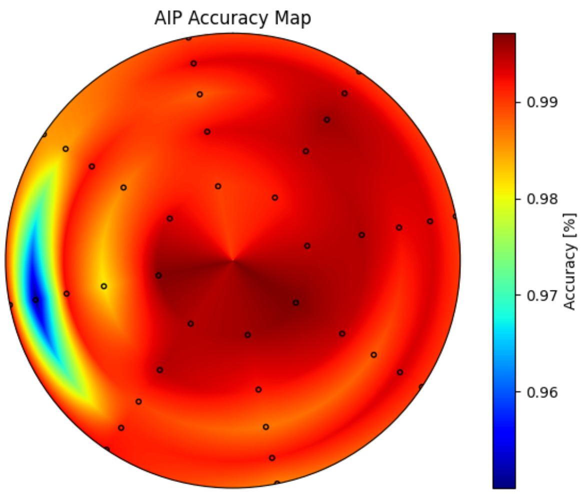

Figure 4.11: Plot of accuracy for method two case two

The second case is also very similar to the corresponding case of the baseline. Based on the pressure plot in Fig. 4.10, the synthesization wrongly predicts a higher pressure on the left-hand side and also erases the small zone of high pressure in the top left corner. This shows a good agreement between the baseline and method two, meaning that this method is a good way to predict turbulence. The accuracy plot, Fig. 4.11, also looks the same as the one for the baseline with the same scale and similar areas of similar magnitude.

4.2.3 Distortion indices

The following plots were created based on the distortion index equations in chapter 2. The plots all show the difference between the synthesized data from method two and true data from the wind tunnel. The blue line shows where the synthesized and actual data are equal.

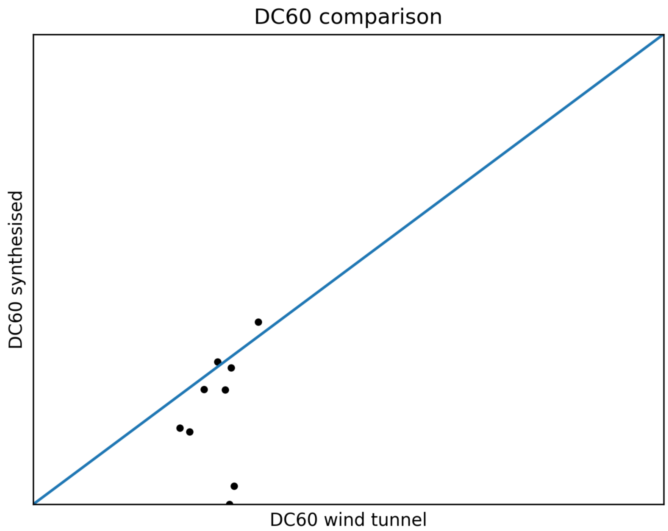

Figure 4.12: Comparison between wind tunnel data and synthesization for distortion index DC60 for the second method.

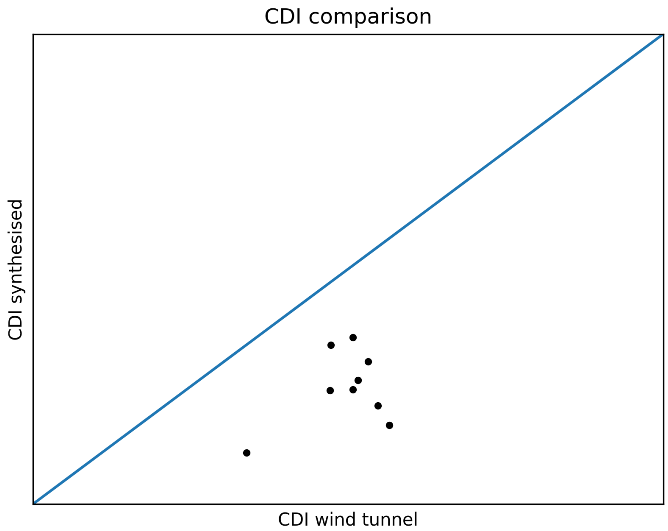

Figure 4.13: Comparison between wind tunnel data and synthesization for distortion index CDI for the second method.

The small amount of data for this batch can be seen in the comparative plot between the synthesized DC60 and true DC60 in Fig. 4.12. The cases were all fairly similar in maximum distortion, and they correspond adequately to the symmetry line. There

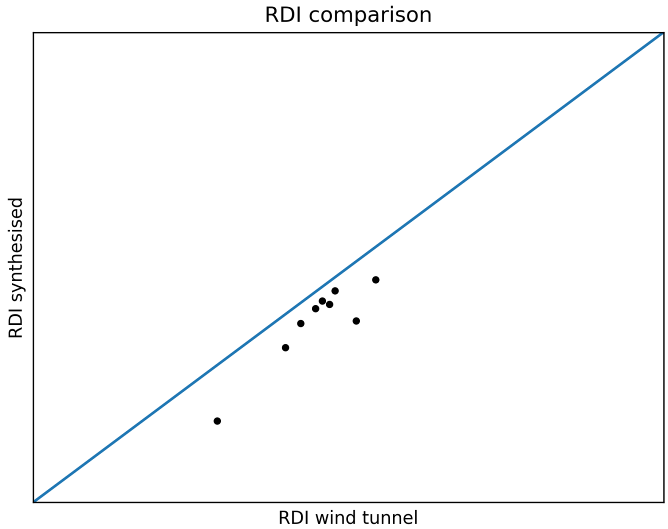

Figure 4.14: Comparison between wind tunnel data and synthesization for distortion index RDI for the second method.

are some points that are very low, and although a few points are above the line, the majority are below, which, similarly to the baseline, indicates that the synthesization under-predicts the distortion. The circumferential distortion follows the same trend, as seen in Fig. 4.13, where the points are close together. The flow cases are thus similar in maximum distortion but are far below the symmetry line, meaning they are under-predicted. This similarity continues for the radial distortion, Fig. 4.14, where the points are just below the symmetry line and also close together.

4.2.4 Conclusion

The similarities between this method's results and the baseline are many, and it appears as if this method is great at predicting turbulence and should be the best option. However, it might be the low amount of data that created a false image of how well it performs. This method is reliant on CFD data to calculate the value of the correlation coefficient, C, as in eq. 3.9 in section 3.3 above, in order to substitute the RMS turbulence. This means that for a varied dataset, this constant becomes a generalized value, which should correspond to every case, introducing other issues. One value cannot be expected to create a perfect correspondence to a set of many flow cases, and by doing that, one can get a sufficiently good result, but it shouldn't be as good as what these results suggest. For the small dataset used, the value of the constant becomes tuned for these specific and similar cases. This is the reason for the good performance of this method and could vary drastically in a larger, more varied dataset. If a more accurate result is desired for a large and varied dataset, it might be beneficial to divide the data into batches that are similar and calculate the correlation coefficient for each batch.

The method's process was very simple and straightforward, as there were only two equations to calculate. The initial equation for calculating the coefficient as a sum of every case was a bit complicated to set up. The second equation for calculating the RMS turbulence was simpler after getting the value for the coefficient. In conclusion, this method provided a high accuracy and was relatively simple but might have issues when using a large and varied dataset.

4.3 Method 3 - Using machine learning and CFD data to estimate RMS turbulence

The third method had the lowest overall accuracy of 97.88%; however, it is not a low value and shows that the method works decently in comparison to the other methods.

4.3.1 Case one

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the first case using method three.

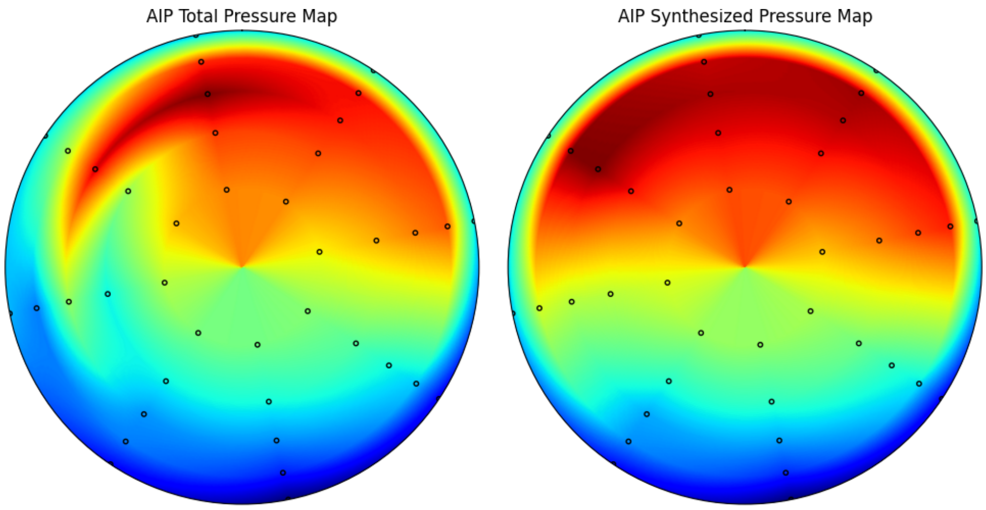

Figure 4.15: Comparison between wind tunnel data and synthesized data for case one using the third method.

The following figure shows the accuracy of the synthesized pressure using method three in relation to the total pressure from the wind tunnel for the first case.

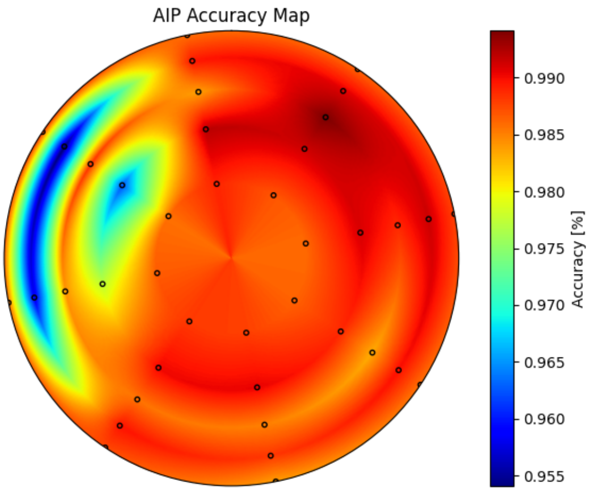

Figure 4.16: Plot of accuracy for method three case one

The pressure plots for the first case using this method are very similar to the baseline. They show the same generalized high-pressure area in the top of the AIP where the highest pressure has been extrapolated into the other nodes, as well as missing the low-pressure area on the left. The pressure map looks like the baseline and is thus an accurate way to predict the dynamic pressure for use in the synthesization procedure. The accuracy map also looks very similar, indicating a good correspondence between the baseline and the third method.

4.3.2 Case two

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the second case using method three.

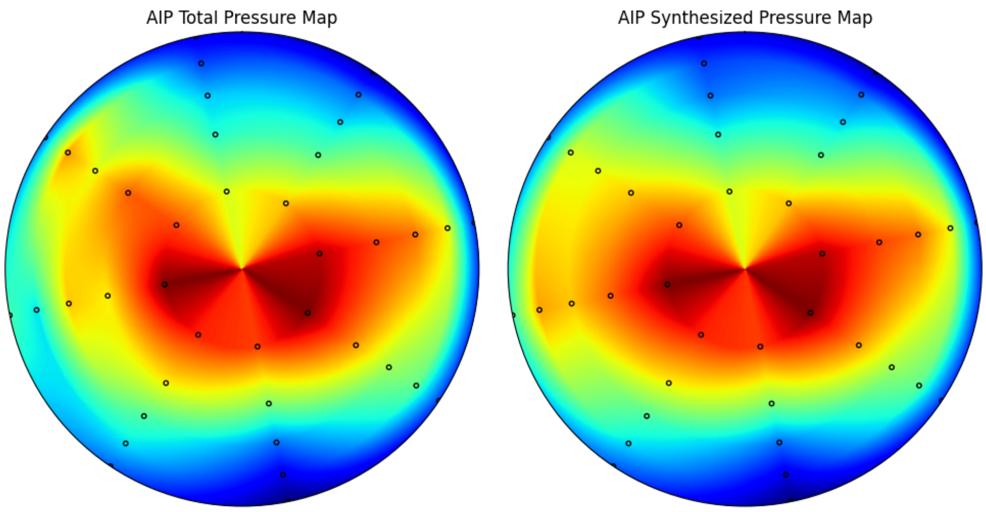

Figure 4.17: Comparison between wind tunnel data and synthesized data for case two using the third method.

This figure shows the accuracy of the synthesized pressure using method three in relation to the total pressure from the wind tunnel for the second case.

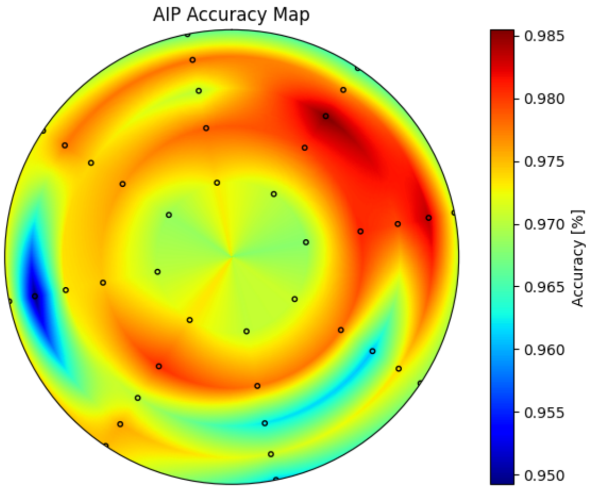

Figure 4.18: Plot of accuracy for method three case two

For the second case, the pressure maps look similar to the baseline. The differences can be seen more clearly in the map of accuracy, Fig. 4.18. It still has a lower accuracy zone on the left, where the model 'guesses' a higher value than the wind tunnel data shows and is in line with the baseline results. The difference lies in the rest of the AIP, where the highest accuracy is much lower, and instead of the average accuracy over the AIP being 99%, it varies a lot. There is a zone of higher accuracy in the top right corner and a zone of lower accuracy in the lower right corner, matching the baseline map in Fig. 4.3. They differ in the middle, where the baseline has a higher accuracy and the third method has a lower accuracy than the rest of the AIP. The accuracy map shows that this method doesn't work very well compared to the baseline, even though it looks accurate in the pressure map. However, the 'bad performance' in accuracy of this method is still only almost 1%, which is not a large difference.

4.3.3 Distortion indices

The following plots were created based on the distortion index equations in chapter 2. The plots all show the difference between the synthesized data from method three and true data from the wind tunnel. The blue line shows where the synthesized and actual data are equal.

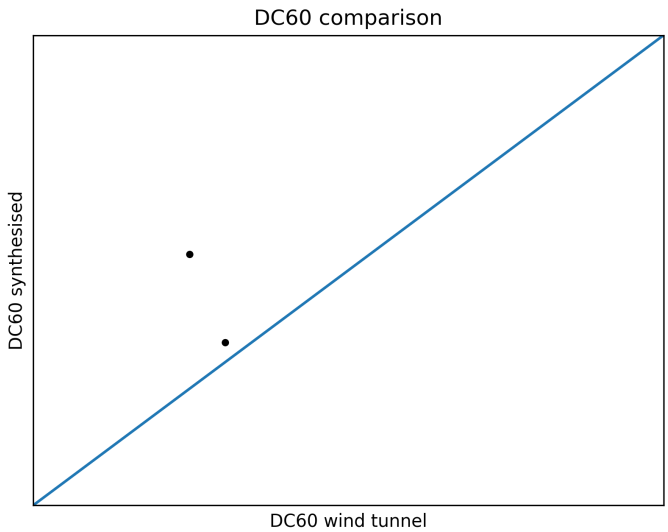

Figure 4.19: Comparison between wind tunnel data and synthesization for distortion index DC60 for the third method.

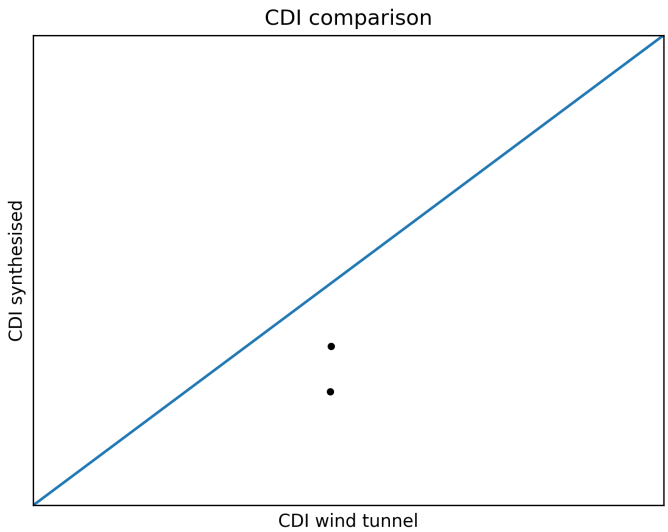

Figure 4.20: Comparison between wind tunnel data and synthesization for distortion index CDI for the third method.

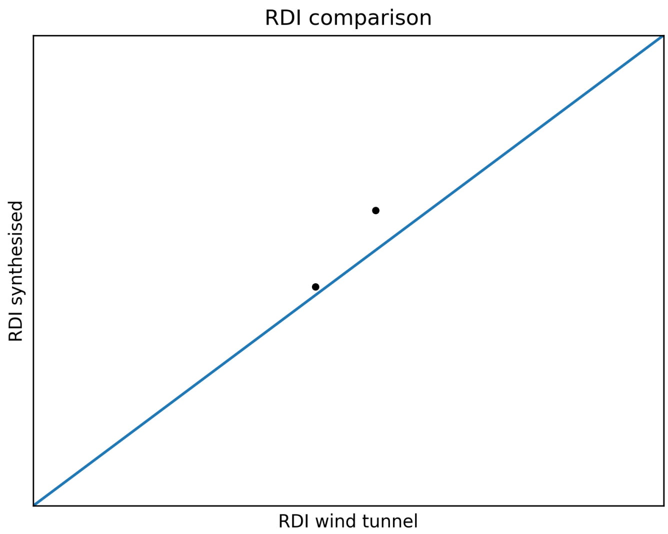

Figure 4.21: Comparison between wind tunnel data and synthesization for distortion index RDI for the third method.

As for method two, the dataset consisted of 9 points, and as written previously, the split into training, validation, and testing was 5, 2, and 2. Therefore, the plot of DC60, Fig. 4.19, only has two points in it, making it hard to draw conclusions. The two points available are located above the symmetry line, meaning that the model in this case overestimated the maximum distortion, contrary to the baseline. The CDI plot, Fig. 4.20, is also hard to draw conclusions from, but these points follow the trend of the baseline more closely, with both points being under the symmetry line. It is thus more accurate with the circumferential distortion than it is with the overall distortion based on the baseline case. The radial distortion has the same issue as DC60; in Fig. 4.21, the points that are below the line in the baseline method are now above, but still relatively close to the line, as is desired.

4.3.4 Conclusion

The results from this method do not conform to the baseline case. There are large discrepancies between the two methods, which point to this method not being a good choice. However, based on the literature study, this method should work better than the previous one. The issue here comes from the small dataset, where machine learning often requires a large amount of data to be accurately trained. As that was not available, the machine learning model never had the opportunity to be trained properly, and thus the results were not good. By using more data, the model could become more accurate in its predictions and would be a more beneficial tool for estimating dynamic distortion. As for the process, this method was more complex than the previous, and it required more time and computational power. It was also difficult to tweak the model parameters for optimization, and thus it might not perform well due to that either. Based on the literature study, this method has a lot of benefits and could work very well if used properly, but for this thesis, it performed the worst and was too time-consuming to be used in the industry.

4.4 Method 4 - Testing the machine learning approach on a one-equation turbulence model

This method had an accuracy of 98.79%, which is not far from the baseline, but not the best out of the methods tested.

4.4.1 Case one

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the first case using method four.

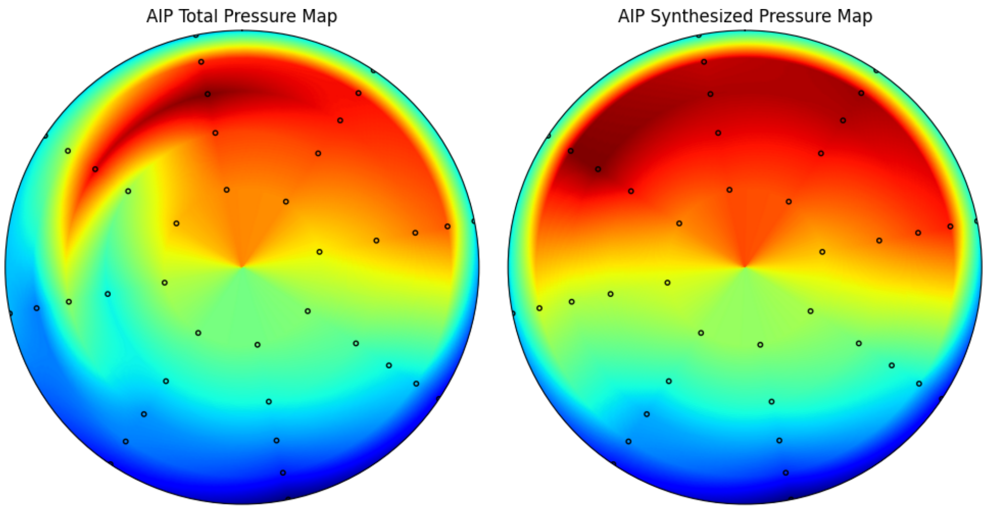

Figure 4.22: Comparison between wind tunnel data and synthesized data for case one using the fourth method.

This figure shows the accuracy of the synthesized pressure using method four in relation to the total pressure from the wind tunnel for the first case.

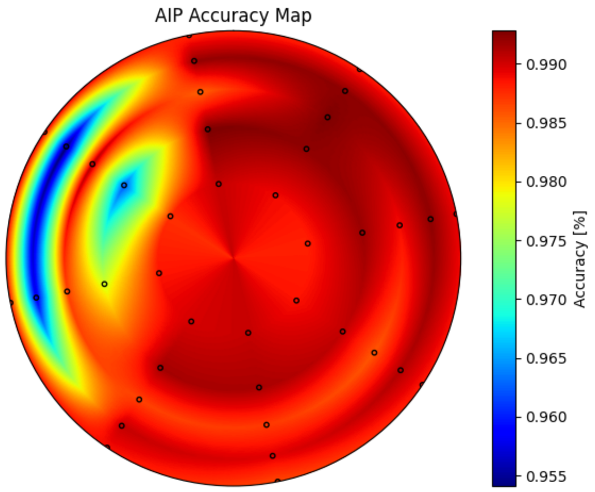

Figure 4.23: Plot of accuracy for method four case one

Due to how close the overall accuracy is, it can be assumed that the results from this method will be very close to the baseline. This can be seen in the pressure map, Fig. 4.22, for case one of this method, as it looks very similar to the baseline with the same inaccuracies as previously discussed on the left-hand side. The same can be seen in the accuracy map, where the large areas of lower accuracy on the left side, both at rings two and four, are similar to the baseline. The scale shows that this method has a lower maximum value of the accuracy, which is in line with the overall accuracy being lower.

4.4.2 Case two

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the second case using method four.

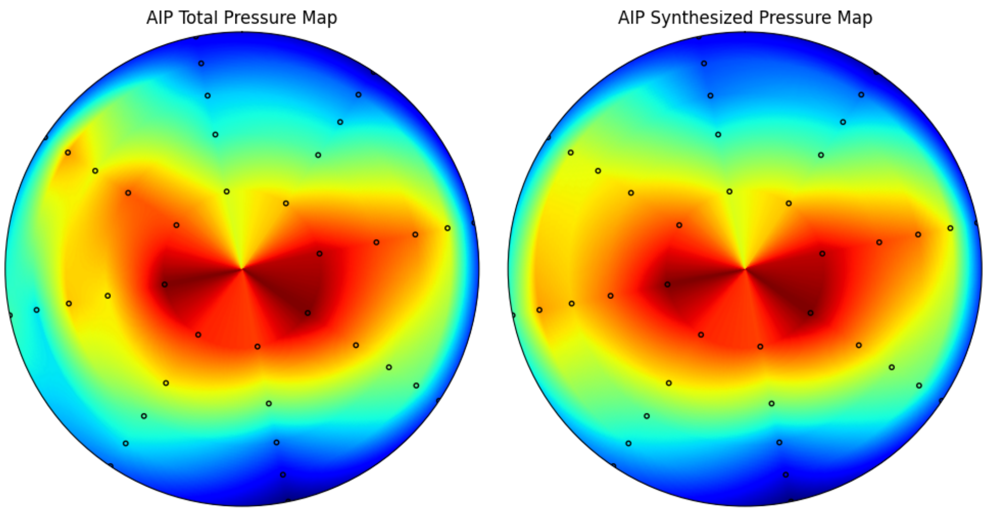

Figure 4.24: Comparison between wind tunnel data and synthesized data for case two using the fourth method.

This figure shows the accuracy of the synthesized pressure using method four in relation to the total pressure from the wind tunnel for the second case.

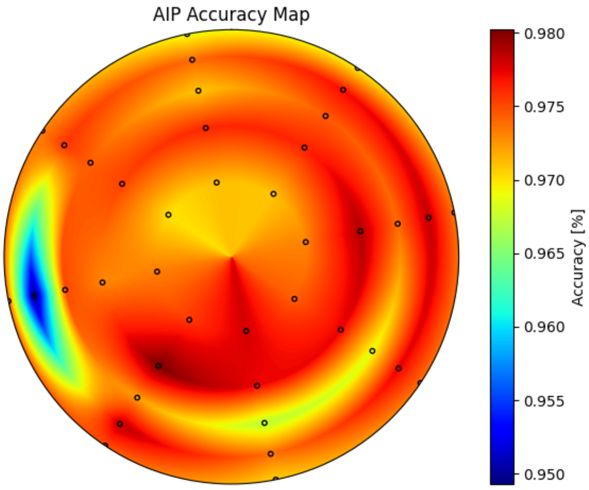

Figure 4.25: Plot of accuracy for method four case two

The plots for the second case also look very similar to the baseline, at least on the pressure side. In Fig. 4.24, the pressure comparison can be seen; it shows the same inaccuracies on the left-hand side as in the baseline. The small high-pressure area is also erased in the same way as discussed in the previous methods. Similarly to the previous methods, the accuracy map looks slightly different from the baseline case, Fig. 4.4, with the scale having a lower maximum value, meaning that this method is slightly less accurate for this case. Although there is a large zone of low accuracy on the left-hand side, which is the same in both plots, there is a 'wall' of high accuracy just inside it in the baseline, which is missing in this method. The low accuracy region in the lower right corner in Fig. 4.4 is lower for the fourth method in comparison. Fig. 4.25 also shows the same trend in the middle of the plot as the previous methods, where it has a lower accuracy than the surroundings. Generally, this accuracy map shows fewer regions of high accuracy and more regions of low accuracy than the baseline.

4.4.3 Distortion indices

The following plots were created based on the distortion index equations in chapter 2. The plots all show the difference between the synthesized data from method four and true data from the wind tunnel. The blue line shows where the synthesized and actual data are equal.

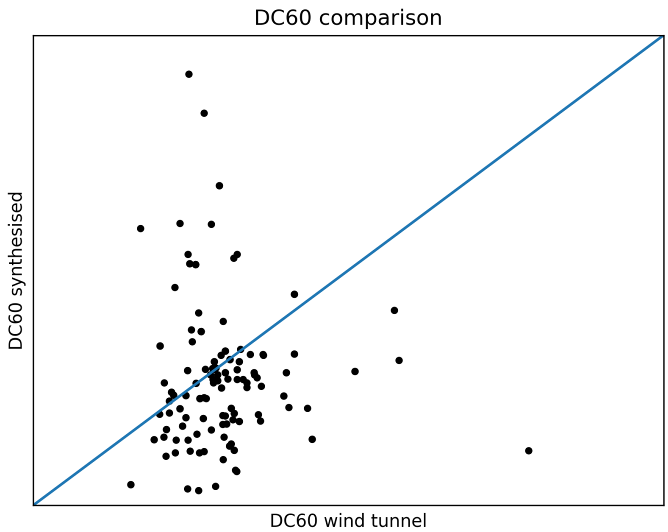

Figure 4.26: Comparison between wind tunnel data and synthesization for distortion index DC60 for the fourth method.

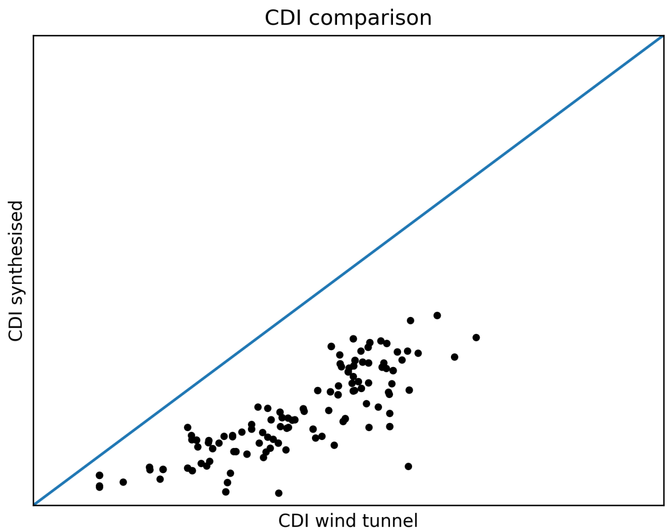

Figure 4.27: Comparison between wind tunnel data and synthesization for distortion index CDI for the fourth method.

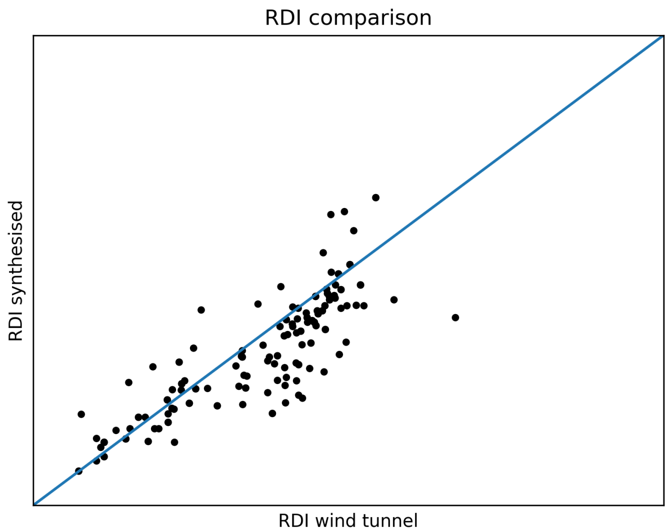

Figure 4.28: Comparison between wind tunnel data and synthesization for distortion index RDI for the fourth method.

The DC60 plot for this method, Fig. 4.26, looks very different from the baseline. A majority of the data points are in the area underneath the symmetry line, similar to the baseline, but there are a lot of points above it as well. They are not as condensed as for the baseline case either, looking less concentrated in one area and instead encompassing a much larger area. This is in line with the accuracy plots from earlier, as the low accuracy is visible in the DC60 plot as well. The same can be seen in Fig. 4.27, which shows the circumferential distortion for this method. This index follows the baseline trend but looks less condensed. Looking at the radial distortion in Fig. 4.28, the points follow a similar pattern as for the baseline but are not as condensed. This is the general impression of these plots compared to the baseline plots: the points follow a similar pattern but are less accurate and more sparse.

4.4.4 Conclusion

An issue that has been frequently mentioned above is the inaccuracy compared to the baseline. This is most likely due to the low amount of data used to train the model. Compared to the previous method, this one uses a larger dataset, but in machine learning terms it is still small. The dataset was large enough to get results that were decent, with an accuracy of 98.79%, but with more data, this accuracy could be improved significantly. This is unfortunately the issue with machine learning approaches: they work great for non-linear correlations but need a large amount of data, which is not always available. The model used, as mentioned in section 3.5, is the same as for method three without tweaking the parameters, which worked for this method as well. If more data is available, then optimizing the method for each method is something that needs to be done. For the process, creating the model and optimizing it makes this method complex and time-consuming.

4.5 Method 5 - Machine learning using only steady-state pressure

The last method had a lower accuracy than the baseline at 98.70%.

4.5.1 Case one

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the first case using method five.

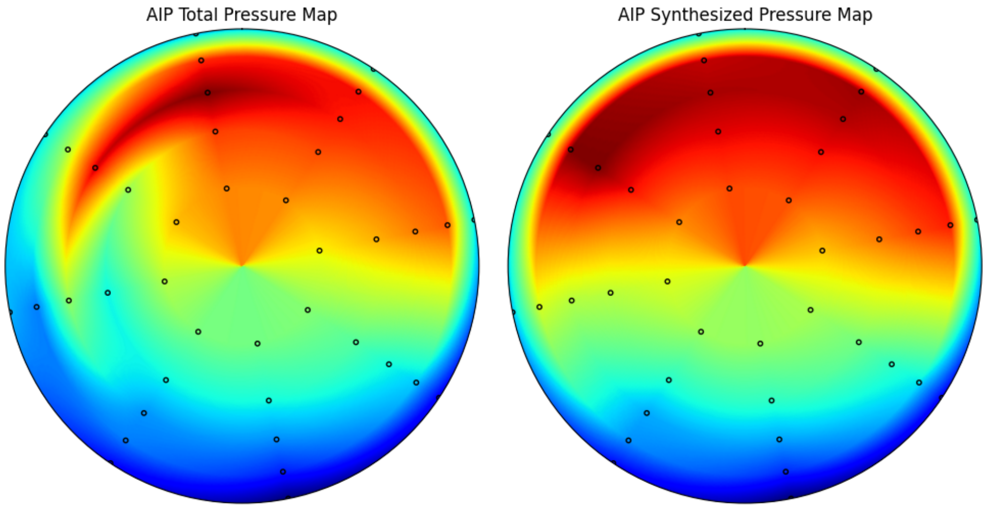

Figure 4.29: Comparison between wind tunnel data and synthesized data for case one using the fifth method.

This figure shows the accuracy of the synthesized pressure using method five in relation to the total pressure from the wind tunnel for the first case.

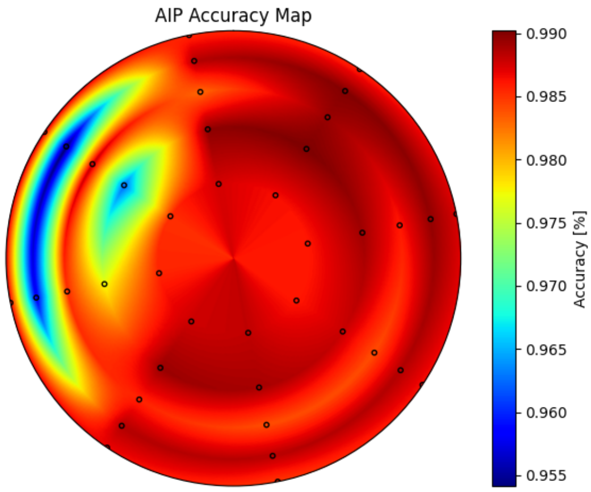

Figure 4.30: Plot of accuracy for method five case one

Result-wise, the first case of this method looks very similar to the corresponding case of the previous method. It has the same similarities and differences compared to the baseline and works decently well based on the pressure map, Fig. 4.29, and accuracy map, Fig. 4.30. Although the results are slightly lower in overall accuracy, this first case performed similarly to the previous method.

4.5.2 Case two

The following plots show the total pressure from the wind tunnel data and the synthesized pressure for the second case using method five.

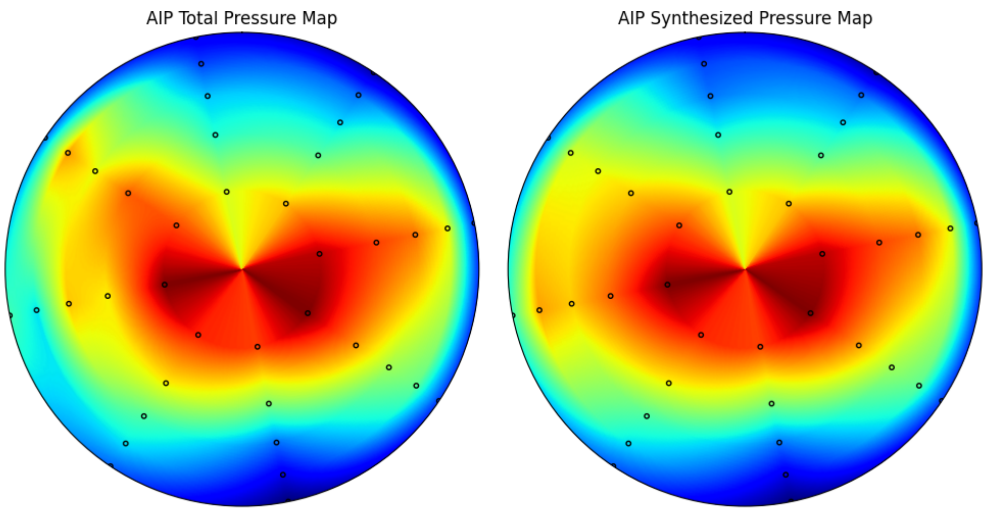

Figure 4.31: Comparison between wind tunnel data and synthesized data for case two using the fifth method.

This figure shows the accuracy of the synthesized pressure using method five in relation to the total pressure from the wind tunnel for the second case.

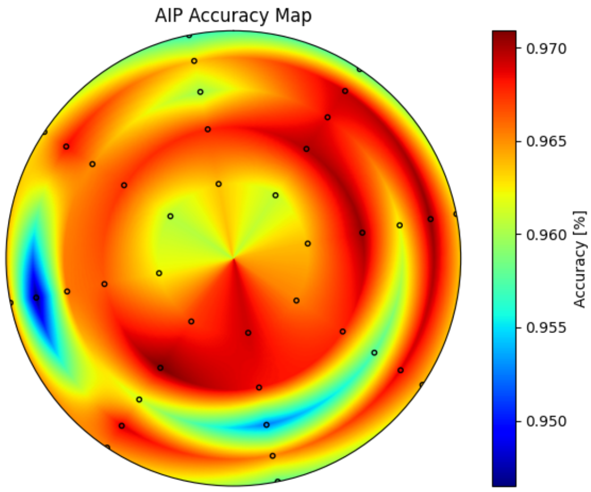

Figure 4.32: Plot of accuracy for method five case two

For the second case, the pressure map also looks very similar to the baseline. The accuracy map differs from the baseline, however, and looks more similar to the fourth method. Figure 4.32 shows a less homogeneous map than the baseline, with larger inaccuracies and more variation. The region with very low accuracy on the left is still there, but a new low-accuracy area is also introduced in the bottom right corner, the same as for the fourth method. At the boundary, specifically on the top and bottom, the accuracy is also lower, which is a difference compared to the baseline.

4.5.3 Distortion indices

The following plots were created based on the distortion index equations in chapter 2. The plots all show the difference between the synthesized data from method five and true data from the wind tunnel. The blue line shows where the synthesized and actual data are equal.

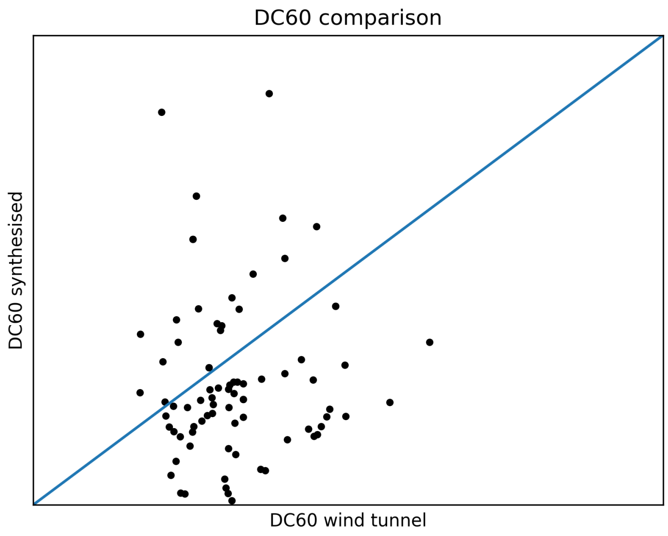

Figure 4.33: Comparison between wind tunnel data and synthesization for distortion index DC60 for the fifth method.

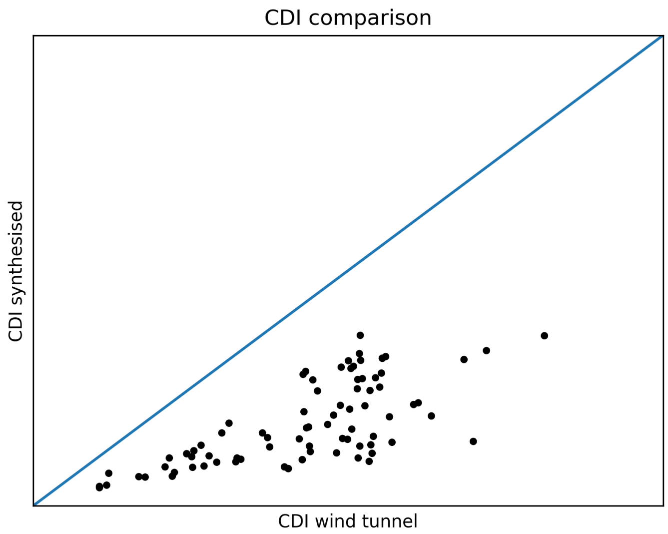

Figure 4.34: Comparison between wind tunnel data and synthesization for distortion index CDI for the fifth method.

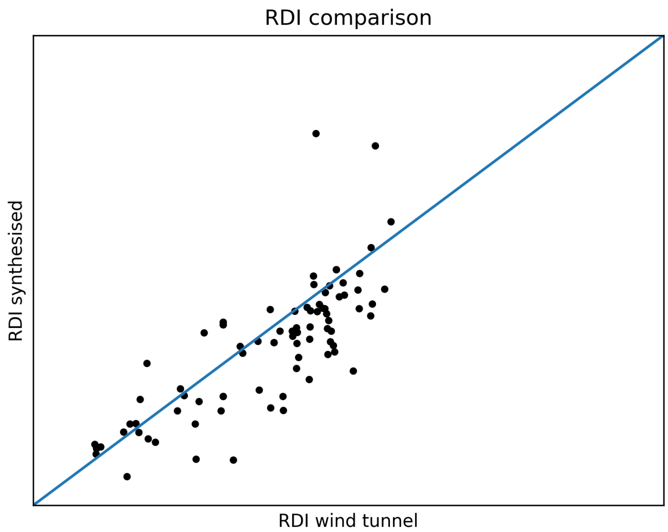

Figure 4.35: Comparison between wind tunnel data and synthesization for distortion index RDI for the fifth method.

Moving to the distortion indices, the DC60 plot, Fig. 4.33, looks vaguely similar to the fourth method but more spread out. There are remnants of the trend that can be seen in the baseline, but it looks like the model had difficulties calculating an accurate maximum distortion. The vague resemblance, however, indicates that the model still works, although not well. The circumferential and radial distortion, Fig. 4.34 and Fig. 4.35, respectively, have the same issue. The points are placed in similar areas to the baseline but are slightly less accurate and condensed.

4.5.4 Conclusion

This method was very similar to the fourth method, using the same model and dataset, but changed the amount of inputs. This resulted in a lower accuracy and indicates that using only the steady-state pressure requires a larger dataset to get a good model. It worked decently well, with the accuracy being close to the baseline and the pressure maps looking largely the same. But, in comparison with the other methods, it performed at a lower level and is thus not a great method for this application. However, the idea of using only steady-state CFD pressure data to estimate dynamic distortion is a major advantage, and although the method did not perform particularly well in these experiments, it remains a promising approach. The relation between the steady-state pressure and dynamic pressure is unfortunately highly non-linear, and the same steady-state pressure could result in slightly different dynamic pressures. So even though the model gets a large dataset, there is a possibility that it will never be 100% accurate. The process of making the machine learning model was complex and time-consuming, as for the previous two methods.

4.6 Method comparison

In this section, the methods will be compared based on three criteria; accuracy, time and complexity. The best method will be chosen based on these criteria.

4.6.1 Accuracy

All methods worked very well, with the accuracies being over 97%. Therefore, it is reasonable to state that any method could be used to do the synthesization and get a reliable result. The best method was the baseline at 98.83%, but due to the usage of only wind tunnel data in this method, it is not optimal to use it in the early design stages. The very purpose of this thesis is using steady-state CFD data to estimate the dynamic distortion, and thus the first method is unsuitable. The second-best method accuracy-wise is the second method in the thesis, with an accuracy of 98.82%. This is almost as good as the best accuracy and should point to the fact that this method works great for predicting distortion. But as was discussed above, the amount of data was homogeneous and small, leading to the value of the correlation coefficient C becoming tuned for the cases tested. This is not inherently bad, but for a larger and more varied dataset, the value of C could instead become too general to the point of being bad for almost every case, which is important to remember. The method with the third-best accuracy was method 4, at 98.79%, which is not far behind the previous two. This method worked well but fell victim to the low amount of data used to train the model, which is a reoccurring argument in this thesis. With a larger amount of data, this method could only work better, and the model could be optimized. The second-to-worst method in terms of accuracy was the fifth method, at 98.70% overall accuracy. This method had the same issues as the fourth method, and even though it used the same dataset, it performed worse. This indicates that using only the steady-state pressure requires more data for training. The worst method was the third, with an accuracy of 97.88%, where conclusions are very hard to draw due to the low amount of data.

It is hard to decide the 'best' method out of the five mentioned above due to the differences in size of data, but from a pure accuracy standpoint, the second method is the best. However, due to the arguments discussed above, it might not be the best in a real-world scenario where there are thousands of data points. Instead, it can be argued that a machine learning method works best in the real world as it is better at adapting to different conditions and learning non-linear correlations instead of generalizing them with a correlation coefficient. Thus, from the accuracy perspective, based on how much data is available, a machine learning approach is probably the most accurate.

4.6.2 Time

In terms of time, the first method was the quickest, with only having to do the synthesization based on data received from the wind tunnel. But as discussed previously, this method is not suitable due to using only wind tunnel data. Thus, disregarding the synthesization process in the next methods as it is the same for every one, the next fastest method is the second method. It was the fastest in researching and setting up the script as well as running the calculations, as it was only math. However, data collection took some time for this, as it required at least a two-equation turbulence model in the CFD simulations to obtain the turbulent kinetic energy. The third least time-consuming method was the fifth, followed closely by the fourth. These methods both needed a machine learning model, which took some time to set up as well as research beforehand. The data collection also required a one-equation turbulence model, which is easier than the two-equation but still takes some time. The difference between them comes from the fifth method only using one input, whereas the fourth uses five, and thus they are ranked in that order. In last place comes the third method; as it is a combination of machine learning and a two-equation turbulence model, this method requires the most time both for research and for collecting data and running the calculations.

Based on the arguments above, the quickest method was the second. It had the most straightforward process based on simple principles, which don't require a lot of time to understand. The process of calculating the dynamic distortion also went fast and required no resource-heavy computational simulations. The issue with this method was the usage of a two-equation turbulence model. Deriving a new model equation for the turbulence and correlation coefficient based on the turbulent viscosity could be highly beneficial for a quick process in the future.

4.6.3 Complexity

The last criteria evaluated is complexity, where the easiest, least complex method, once again is the first but it is disqualified for the same reasons. As time consumption and complexity goes hand in hand, the method ranking for this criteria looks the same as for time above. The easiest method is the second and the jump to the next three is quite large as the complexity increases drastically for a machine learning approach. Once again, the usage of a two-equation turbulence model increases the complexity as well compared to using a one-equation.

Choosing the easiest method is as simple as for the least time consuming one. The large increase in complexity with the machine learning model makes the second method stand out as the least complex.

4.6.4 Conclusion

By looking at which method was the best purely based on the previously mentioned criteria, the second method stands out as it was the best on two out of three criteria. Thus, if the purpose of predicting the distortion is fast and easy but not necessarily the most accurate results, then this method is preferable. However, as previously discussed, for a varied dataset, it can become less accurate. For a small or homogeneous set of data, this method works the best for the studied methods. There is also the added difficulty of having to use a two-equation turbulence model in the CFD simulations. If accuracy is the goal with the distortion synthesis, then a method using machine learning is the best. There is also more variation in the application, as it can learn from a data set with largely varied data without sacrificing accuracy. The advantage of using only steady-state pressure from the CFD solution also makes it highly preferable. Spending some time setting up a model that can then be used on just the steady-state pressure from different configurations without running a complex turbulence model can save more time than is required to set up and tune the machine learning model.

4.7 Sources of error

A potential source of error in all methods (except the first one) is the usage of RANS simulations for the CFD data. The figures below show how much three different simulation models differ in results for the same flow field. Figure 4.36 shows the resolved AIP using a time-accurate hybrid RANS-LES CFD simulation. The figure next to it, Fig. 4.37, shows the same solution but averaged over time. The third image, Fig. 4.38, shows the same solution but using a RANS model to solve it.

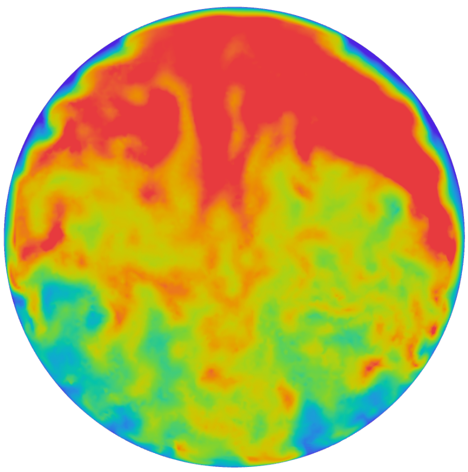

Figure 4.36: Time-accurate hybrid RANS-LES CFD simulation

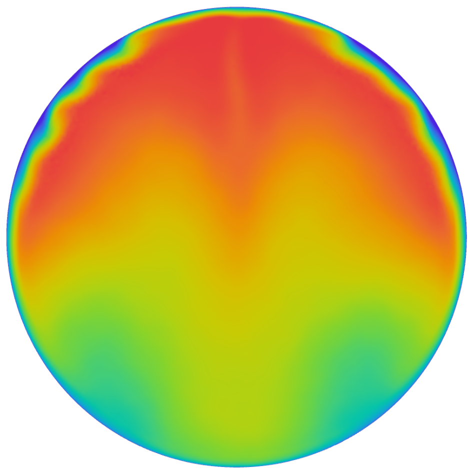

Figure 4.37: Average of time-accurate simulation

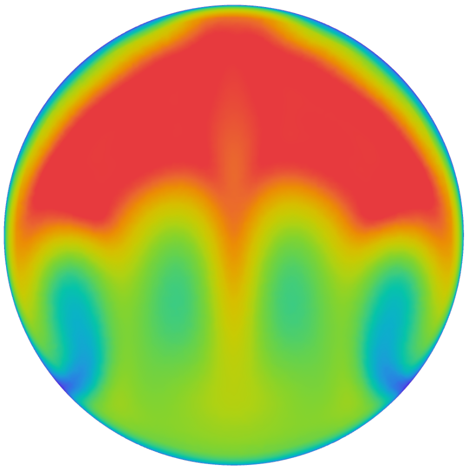

Figure 4.38: Plot of RANS simulation

As can be seen in the images, the solutions look similar, but there are huge differences. The time-accurate solution, Fig. 4.36, is not symmetrical. The turbulent eddies, or vortices, can be seen clearly, as well as the large variations in pressure. A large part of this resolution is lost in the time-averaged simulation, and the turbulence looks more symmetric than in the hybrid solution. Even more resolution is lost in the RANS solution, where everything becomes more exaggerated and the different zones become larger and more average.

As discussed previously in chapter 2, RANS simulations are used widely in the industry due to them being computationally efficient. However, as they average the unsteady turbulence and use a closure model to replace the Reynolds stresses, they introduce model-form error [27]. This means that the predicted turbulence field isn't a direct or accurate measurement of the turbulence. It is instead a reconstruction dependent on a model that can deviate substantially from how the flow actually behaves. This is especially true when the flow deviates from the conditions used to calibrate the model.

4.8 Future work

This thesis focused on the same baseline method of synthesization and then varied which inputs were put in the synthesization to test the possibility of using steady-state CFD to estimate dynamic distortion. No other methods for synthesization were tested, which presented a limitation to the accuracy of the results. This was an active choice made in the beginning of the thesis to limit the scope, but in future work, experimenting with the synthesization method and finding the best way to do it would be interesting. It can be seen that a consequent issue with all methods is predicting circumferential and overall distortion. Especially circumferential distortion is always very low and follows a linear curve; however, not the one where the indices are equal between the synthesized and true data. This error comes from the synthesization being bad at predicting the maximum pressure distortion and the statistical method being conservative by nature, as the normal distribution favors values close to the mean.

Another way to go about future work is working more with the machine learning model, as discussed earlier. Due to the low amount of data, the work in this thesis should be considered as a pre-study rather than a complete assessment of using machine learning with turbulence modeling. The results show promise, but no absolute conclusions can be drawn other than that it could work. Thus, spending time getting more data and optimizing the model is a good choice for future work.

4.9 Societal and ethical aspects

A significant ethical aspect of the development of aircraft engines is the emissions generated from flying these aircraft. During 2019, airplanes represented 2.5% of global CO_2 emissions and is said to increase with each year [43]. However, this data is for subsonic commercial airplanes, not military or other types of aircraft. Military aircraft, as well as other applications, are major contributors to global greenhouse gas emissions every year, but because reporting the amount of emissions is optional in many states, it is difficult to know exactly how large the contribution is [44]. Moving towards more sustainable solutions in the future is something that needs to be done to try and mitigate the damage done to the environment.

5

Conclusion

In this thesis, five different methods were investigated and compared that synthesize the dynamic distortion in an engine intake. The evaluation of these methods was done in respect to their ability to replicate experimental data from a wind tunnel. Their performance for the distortion indices DC60, CDI, and RDI, and the practical usefulness of the processes through time-consumption and complexity were also explored. Overall, all methods produced reasonably good results, with all accuracies above 97%, showing that dynamic distortion can be estimated with high accuracy using the proposed method.

The first method, referenced as the baseline, relied entirely on wind-tunnel data and resulted in the highest accuracy. This method served as a useful reference for the remaining four methods; however, it is not a suitable method for early design stages due to the dependence on experimental data, which are costly and time-consuming to obtain. The second method replaced the wind-tunnel RMS turbulence with a value estimated from CFD simulations. It achieved a nearly identical accuracy and showed a similar behavior to the baseline. This makes it a promising alternative, especially since it replaces the need for wind tunnel experiments with CFD simulations, even though a two-equation turbulence model is needed. The machine learning methods had the greatest potential for future development. And although the best result using machine learning did not outperform the baseline, it demonstrated that data from CFD simulations can be used to train models capable of estimating the dynamic distortion with a high accuracy. Concurrently, the small amount of data available for training made it difficult to draw definitive conclusions, and the performance of the model was constrained by the limited dataset. Using a larger and more varied dataset, the performance of these methods could be improved significantly.

A general trend noticed across the results in the thesis is the accuracy of the radial distortion in comparison to the circumferential distortion. The maximum distortion had a tendency of being underpredicted. This trend indicates that the methods of synthesis are better at reproducing the overall flow pattern than the maximum distortion. Another parameter that may have contributed to the uncertainty of the methods is the choice of turbulence model in the CFD simulation. As RANS-based turbulence models have a predisposition for smoothing out unsteady flow structures, important distortion features may therefore be missed.

In conclusion, the study shows that CFD data from the steady-state RANS simulations can be used to estimate dynamic distortion, which supports the main purpose of the thesis. Comparing the methods investigated, the approach using machine learning has the best overall potential. Simpler mathematical method also performed well and may be more practical in situations where the data is limited. Future work should focus on optimizing the machine learning model using more data as well as testing additional synthesis approaches to improve the prediction of circumferential and maximum distortion.

Bibliography

[1] F.-q. Chen, X.-d. Ren, Y.-h. Li, and C.-w. Gu, "Experimental and numerical study on aerodynamic stability of compressor l-inlet duct," *Aerospace Science and Technology*, vol. 139, p. 108 368, 2023, ISSN: 1270-9638. DOI: https://doi.org/10.1016/j.ast.2023.108368. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1270963823002651.

[2] C. C. Ciepluch, "Effect of inlet air distortion on the steady-state and surge characteristics of an axial-flow turbojet compressor," *Langley Research Center*, 1948.

[3] T. J. Biesiadny, W. M. Braithwaite, R. H. Soeder, and M. Abdelwahab, "Summary of investigations of engine response to distorted inlet conditions," *NASA Lewis Research Center*, 1986.

[4] M. Migliorini, P. K. Zachos, D. G. MacManus, and P. Haladuda, "S-duct flow distortion with non-uniform inlet conditions," *Institution of Mechanical Engineers*, 2022.

[5] G. Tanguy, D. G. Macmanus, and E. Garnier, "Numerical investigation of the unsteady distortion for an sduct intake with mechanical vortex generators," *International Journal of Heat and Flow*, 2022.

[6] T. Triantafyllou, T. Nikolaidia, M. Diakostefanis, and P. Pilidis, "Total pressure distortion levels at the aerodynamic interface plane of a military aircraft," *The Aeronautical Journal*, 2015.

[7] J. Ren, Z. Wu, H. Tan, D. Li, and Y. Zhou, "Experimental investigation of the aerodynamics of a uav inlet with double 90° bends," *Aerospace Science and Technology*, vol. 154, p. 109 502, 2024, ISSN: 1270-9638. DOI: https://doi.org/10.1016/j.ast.2024.109502. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1270963824006333.

[8] D. Sedlock, "Improved statistical analysis method for prediction of maximum inlet distortion," Air Force Wright Aeronautical Laboratories, Tech. Rep., 1985.

[9] W. P. Norby, J. A. Ladd, and A. J. Yuhas, "Dynamic inlet distortion prediction with a combined computational fluid dynamics and distortion synthesis approach," *NASA Contract*, 1996.

[10] A. F. Campbell, "An investigation of distortion indices for prediction of stalling behaviour in aircraft gas turbine engines," M.S. thesis, Virginia Polytechnic Institute and State University, 1981.

[11] "Gas turbine engine inlet distortion guidelines," SAE Aerospace, Tech. Rep., 1978.

[12] B. S. Siedel, “Asymmetric inlet flow in axial turbomachines,” *Massachusetts Institute of Technology*, 1959.

[13] H. Yeh, “An actuator disc analysis of inlet distortion and rotating stall in axial flow turbomachines,” *University of Penssylvania*, 1959.

[14] H. C. Melick, “Analysis of inlet flow distortion and turbulence effects on compressor stability,” NASA Ames Research Center, Tech. Rep., 1973.

[15] R. W. King, J. A. Schuerman, and R. G. Muller, “Analysis of distortion data from tf30-p-3 mixed compression inlet test,” *NASA Lewis Research Center*, 1976.

[16] D. L. Motycka, “Determination of maximum expected instantaneous distortion from inlet pressure rms measurements,” AIAA Paper No.-78-970, Tech. Rep., July 1978.

[17] J. F. Runckel, “An investigation of the effects of a vortex-generator configuration on the aerodynamic characteristics of a 1/4-scale model of the x-1 airplane (10-percent-thick wing),” Langley Research Center, Tech. Rep., 1959.

[18] G. J. Harloff, B. A. Reichert, and S. R. Wellborn, “Navier-stokes analysis and experimental data comparison of compressible flow in a diffusing s-duct,” AIAA, Tech. Rep., 1992.

[19] Z. Lifen, L. Zhenxia, Q. Jiyun, and W. Xiaofeng, “An improved cfd-synthesis method for predicting inlet dynamic distortion,” *Chinese Journal of Aeronautics*, 2008.

[20] J.-X. Wang, J. Wu, J. Ling, G. Iaccarino, and H. Xiaoa, “A comprehensive physics-informed machine learning framework for predictive turbulence modeling,” Physical Review Fluids, Tech. Rep., 2018.

[21] J.-L. Wu, H. Xiao, and E. Paterson, “Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework,” *Physical Review Fluids*, vol. 3, no. 7, Jul. 2018, ISSN: 2469-990X. DOI: `10.1103/physrevfluids.3.074602`. [Online]. Available: `http://dx.doi.org/10.1103/PhysRevFluids.3.074602`.

[22] W. Liu, J. Fang, S. Rolfo, C. Moulinec, and D. R. Emerson, “An iterative machine-learning framework for rans turbulence modeling,” *International Journal of Heat and Fluid Flow*, 2021.

[23] R. Schiestel and B. Chaouat, “Turbulence modeling and simulation advances in cfd during the past 50 years,” *Comptes Rendus Mecanique*, 2022.

[24] J. D. A. Jr, *Governing Equations of Fluid Dynamics*. National Air and Space Museum, 2009.

[25] M. Ekh, “Mechanics of solids fluids – introduction to continuum mechanics,” Chalmers University of Technology, Tech. Rep., 2017.

[26] COMSOL, “Navier-stokes equations,” *Multiphysics Cyclopedia*, 2017, Available from: `https://www.comsol.com/multiphysics/navier-stokes-equations` (fetched May 13th 2026).

[27] L. Davidson, *Fluid mechanics, turbulent flow and turbulence modeling*. Chalmers University of Technology, 2025, CC Public Domain.

[28] C. CFD, “A guide to understanding turbulent kinetic energy,” *Cadence*, 2013, Available from `https://resources.system-analysis.cadence.com/`

blog/msa2022-a-guide-to-understanding-turbulent-kinetic-energy (fetched 12th April 2026).
[29] T. Saad, "Turbulence modeling for beginners," *University of Tennessee Space Institute*,
[30] P. V. Raje, "Spalart-allmaras turbulence model for compressible flows," Department of Aerospace Engineering, Indian Institute of Technology, Bombay, Tech. Rep., 2015.
[31] "Two equation turbulence models," *CFD Online*, Available from: https://www.cfd-online.com/Wiki/Two_equation_turbulence_models (fetched 27th May 2026).
[32] "What is statistics?" *Michigan Tech*, Available from: https://www.mtu.edu/math/undergraduate/statistics/what/ (fetched February 12th 2026).
[33] "Related distributions, sec. 1.3.6.2," *National Institute of Standards and Technology*, Available from: https://www.itl.nist.gov/div898/handbook/eda/section3/eda362.htm (fetched February 12th 2026), CC Public Domain.
[34] D. Bergmann, "What is machine learning?" *IBM Think*, Available from: https://www.ibm.com/think/topics/machine-learning (fetched February 20th 2026).
[35] S. J. Bigelow, "Machine learning vs. neural networks: What's the difference?" *TechTarget*, 2024, Available from: https://www.techtarget.com/searchenterpriseai/answer/Machine-learning-vs-neural-networks-Whats-the-difference (fetched March 28th 2026).
[36] "Artificial neural networks and its applications," *Geeks for geeks*, 2025, Available from: https://www.geeksforgeeks.org/deep-learning/artificial-neural-networks-and-its-applications/ (fetched February 20th 2026), CC Public Domain.
[37] "Splitting data for machine learning models," *Geeks for geeks*, 2023, Available from: https://www.geeksforgeeks.org/machine-learning/splitting-data-for-machine-learning-models/ (fetched 26th May 2026).
[38] "Backpropagation in neural network," *Geeks for Geeks*, 2026, Available from: https://www.geeksforgeeks.org/machine-learning/backpropagation-in-neural-network/ (fetched March 16th 2026), CC Public Domain.
[39] N. Benji, "Breaking down backpropagation in pytorch," *Medium*, 2024, Available from: https://medium.com/@noel.benji/breaking-down-backpropagation-in-pytorch-3762ea107d3a (fetched March 25th 2026).
[40] J. Pecinka, G. T. Bugajski, P. Kmoch, and A. Jilek, "Jet engine inlet distortion screen and descriptor evaluation," Czech Technical University in Prague, Tech. Rep., 2017.
[41] D. E. Funes-Sebastian and L. P. Ruiz-Calavera, "Numerical simulations of wind tunnel effects on intake flow of a uav configuration," American Institute of Aeronautics and Astronautics, Tech. Rep.
[42] M. E. Sanders and R. J. Christenson, "An evaluation of statistical methods for the prediction of maximum time-variant inlet total pressure distortion," Arnold Engineering Development Center, Tech. Rep., 1980, CC Public Domain.

[43] H. Ritchie, “What share of global CO_2 emissions come from aviation?” *Available from: `https://ourworldindata.org/global-aviation-emissions` (fetched 30th jan 2026)*, 2024.
[44] CEOBS, “The military emissions gap,” *Available from: `https://militaryemissions.org/` (fetched 30th jan 2026)*, 2021.

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY