"Development of a validation method for multivariate arbitrarily distributed results"

Von der Fakultät für Georessourcen und Materialtechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

genehmigte Dissertation

vorgelegt von **M.Sc.**

Marinus Johannus van Enkhuizen

aus Heusden, Niederlande

Berichter: Univ.-Prof. Dr.-Ing. Stefan Reh
Univ.-Prof. Dr.-Ing. Gerhard Hirt
Univ.-Prof. i.R. Dr.-Ing. R. Marissen

Tag der mündlichen Prüfung: 16. Mai 2017

Acknowledgements

When I started my MSc at the TU Delft at the department Design and Production of Composite Structures, I never expected it would lead to a PhD research abroad at the German Aerospace Center, institute of material research. For this great opportunity, I would like to thank my supervisor and promotor Prof. Dr.-Ing. Reh and Dr.-Ing. Marissen for their encouragement, support and advises. Furthermore, I would like to thank Prof. Dr.-Ing. Hirt and Dr.-Ing Marissen for their interest and evaluation of this thesis as co-referee.
Although the name of the institute Material research was not one of the reasons to take this opportunity, certainly the people who were to become my colleagues were. To two very special colleagues, Michael and Christian, I would like to thank both of you for your support, the many scientific discussions and the chit chatting. Henning, my first office mate, thank you for your support, the German lessons and bringing me up to speed. To Matthias, thank you for your nearly endless support with the Workstations. I would not know how I could have done it without you keeping the systems running. Furthermore, I would like to thank Ulrike for her all-round support and making my start at DLR so much easier. To Janine, thank you for being my mentor.
I would also like to thank Mr Fuchs and Mr Nodeh-Farahani for their support regarding the testing rigs and I would like to thank Mr Sick for calibrating the measurement equipment and the discussions about measurement uncertainties. To Philipp and Alexander, thank you for your assistance with all the metallographic preparations of the many specimens. Rebekka, thank you for your expertise and patience during the Laser Scanning Microscope sessions to create high quality images. To all other colleagues, thank you for the pleasant working environment.
Outside the institute I have met some very special people at the German Aerospace Center, who happen to share the same passion, scuba diving. Stefan, Klaus, Dirk and Valeska thank you for your moral support, great dives and the opportunity to become a PADI Divemaster. To the lot of divers, thank you for all the dinners, diving trips and so much more that kept me busy and very happy.
Desiree, my girlfriend, I would like to thank you for your patience, understanding and moral support.
Finally, this really comes down to two people. Mom and dad, thanks for the understanding and moral support.

Contents

List of Figures

List of Tables

List of symbols

Symbol	Description
α	Significance level of a hypothesis test, type I error
β	Type II error
$\beta_{D_M{}^2}$	Type II error of the Mahalanobis distance statistic
β_m	Type II error of the Box-M statistic
ΔL	Specimen elongation
ε_{eng}	Engineering strain
$\varepsilon_{\text{true}}$	True strain
ε_{pl}	Plastic true strain
γ	Saturation stress
γ_{start}	The ratio of the sub-domain based on the global domain
λ	Coefficient of the power transformation
λ	Non-centrality parameter
λ_1	Parameter of lognormal distribution 1
λ_2	Parameter of lognormal distribution 2
∇^2	Laplacian operator
μ_1	Mean value of variable 1
μ_2	Mean value of variable 2
ν	Poisons ratio of the material being indented
ν_i	Poisons ratio of the indenter
ϕ	Probability that the difference between the switches samples is greater than or equal to the difference between the original samples
ϕ_0	Probability that the difference between the Pseudo Samples is greater than or equal to the difference between the original samples
ϕ_j	Probability value of the j-th hypothesis test
ϕ_l	Limit of the probability ϕ
ρ	Correlation coefficient
ρ_0	Linear correlation coefficient between images
ρ_1	Correlation coefficient between the variables of multivariate sample 1
ρ_2	Correlation coefficient between the variables of multivariate sample 2
ρ_d	Density of the material of the plate
ρ_m	Correlation coefficient according to Manders
ρ_r	Rank correlation coefficient between images
σ	Standard deviation
σ_1	Standard deviation of variable 1
$\sigma_{1,\text{exp}}$	Standard deviation of the first experimental results
σ_2	Standard deviation of variable 2
σ_{data}	Standard deviation of the data

Continued on next page

Symbol	Description
σ_{eng}	Engineering stress
σ_i	Initial yield stress
σ_j	Standard deviation in the j-th direction of the sample
σ_{opt}	Theoretical optimal kernel width of a gaussian kernel estimator
σ_s	The difference between saturation stress and initial yield stress
σ_{true}	True stress
σ_v	Von misses stress for 2-dimensional problems
σ_y	Flow stress of the material
θ	Angle of rotation
${\zeta_1}^2$	Parameter of lognormal distribution 1
${\zeta_2}^2$	Parameter of lognormal distribution 2
$\boldsymbol{\mu}_{\text{exp}}$	Mean value vector of the experiments
$\boldsymbol{\mu}_{\text{sim}}$	Mean value vector of the simulations
$\boldsymbol{\sigma}_{\text{exp}}$	Covariance matrix of the experiments
$\boldsymbol{\sigma}_{\text{sim}}$	Covariance matrix of the simulation
$\boldsymbol{\sigma}_{\text{kernel}}$	Standard deviation of the Gaussian kernel
a	Length of the plate
$A(y)$	Cross sectional area at position y of a tensile specimen
A^*	The Anderson Darling statistic corrected for the size of the samples
A^2	The Anderson Darling Statistic
A_{SIE}	Slip line area of a sphere indentation experiment
A_t	Slip line area of the tensile experiments
A_0	Initial cross section
A_c	Contact area
A_i	Size of the i-th local area
$A_s(y)$	Slip line area at position y of a tensile specimen
A_t	Total area
b	Width of the plate
c	Constant
C_1	Kernel dependent constant 1
C_2	Kernel dependent constant 2
c_d	The n-dimensional volume of a unit sphere
d	Dimensionality of the data
d_1	Mean value distance between
d_2	Ratio of standard deviations
d_c	Distance between two univariate cumulative distribution functions
$d_{\text{flow stress}}$	Difference between the experimentally measured flow stress and the fitted flow stress
d_i	Diameter of the indentation
d_p	Distance between two multivariate probability density functions
D_i	Diameter of the Indentor
D_{j0}	Original distance between the j-th experimental sample and the j-th simulated sample
D_{ji}	Distance between the i-th randomisation based on the j-th experimental sample and j-th simulated sample

Continued on next page

Symbol	Description
D_m	Normalised mean value distance
$D_{\mathrm{M}}{}^2$	Mahalanobis distance
D_p	Distance between two discrete probability distribution functions
D_v	Ratio between two variances
e_i	The observed probability that a data point falls in interval i
E	Youngs modulus
E_{g}	Global stiffness
E_{IT}	Indentation modulus
$E_{l,i}$	Local stiffness of i-th area
E_l	Local stiffness
E_i	Youngs modulus of the indenter
E_r	Reduced youngs modulus
E_s	Slope of saturation
$f(\phi)$	Probability density function of ϕ
$f(\mathbf{x})$	Multivariate estimated probability density function
$f(x)$	Probability density function of x
$f_{\frac{1}{2}d(d+1),\lambda}(m)$	Non-central Chi-square distribution function of m
$f_{\mathrm{exp}}(\boldsymbol{x})$	Multivariate probability density function of the experiments
$f_{\mathrm{sim}}(\boldsymbol{x})$	Multivariate probability density function of the simulations
$f_{d,v-d+1,\lambda}(x)$	Non-central F-distribution of x
$f_1(\mathbf{x})$	Multivariate distribution function 1
$f_2(\mathbf{x})$	Multivariate distribution function 2
f_j	The j-th edge distribution of the probability density function
F	Force
$F(\phi)$	Cumulative distribution function of ϕ
F_{max}	Maximum applied force
$F_{d,v-d+1}$	F-Distributed variable with d and $v-d-1$ degrees of freedom
F_1	Force applied on the length of the plate
$F_1(x)$	Cumulative distribution function 1 of x
F_2	Force applied on the width of the plate
$F_2(x)$	Cumulative distribution function 2 of x
F_x	The cumulative distribution function of the population of x
G	Desired quantity
h	kernel width
h_{max}	Maximum indentation depth
h_c	The indentation depth
H_0	Null hypothesis
H_1	Alternative hypothesis
H_{IT}	Indentation hardness
$k_i(\mathbf{x})$	Indicator function that determines in what bin the point i could lay
k_{th}	Gaussian kernel with theoretical optimal kernel width
k_{unc}	Gaussian kernel with uncertainty based kernel width
L_0	Initial gauge length
m	Box-M statistic
m_{crit}	Critical value of the Box-M statistic

Continued on next page

Symbol	Description
M	Distance between two covariance matrices
n	Sample size
n_1	Number of data points in sample 1
n_2	Number of data points in sample 2
nge	Number Greater than or Equal to
N	Number of data points per sample
NS	Number of Pseudo Samples
PS	Total number of Pseudo Samples
r	Balloon radius
r^*	Normalised balloon radius
r_i	Balloon radius of the data point i
r_p	Radius of the plastic zone
R	Radius of the sphere indenter
s_{pl}	Pooled standard deviation
s_1	Standard deviation of sample 1
s_2	Standard deviation of sample 2
S	Contact stiffness
$\overline{S}_1$	Averaged pixel intensity of image 1
$\overline{S}_2$	Averaged pixel intensity of image 2
$S_{1\mathrm{i}}$	Pixel intensity of i-th pixel of image 1
$S_{2\mathrm{i}}$	Pixel intensity of i-th pixel of image 2
t	Student-t distributed variable
t	Thickness of the plate
t_{crit}	Critical value of the null hypothesis
t_{H_1}	The distribution of the distance if the alternative hypothesis is true
t_i	The theoretical probability that a data point falls in interval i
T^2	Hotelling-T^2 distributed variable
$T^2_{d,v}$	Hotelling-T^2 distributed variable with d and v degrees of freedom
$\mathrm{u}_{\mathrm{exp},1}$	1st component of the uncertainty of the experiments
$\mathrm{u}_{\mathrm{exp},2}$	2nd component of the uncertainty of the experiments
$\mathrm{u}_{\mathrm{sim},1}$	1st component of the uncertainty of the simulations
$\mathrm{u}_{\mathrm{sim},2}$	2nd component of the uncertainty of the simulations
u_1	Uncertainty component 1
u_2	Uncertainty component 2
u_i	Component i of the uncertainty
W	Power transformed sample
$\bar{x}_1$	Mean value of sample x_2
$\bar{x}_1$	Mean value of sample 1
$\bar{x}_2$	Mean value of sample x_1
$\bar{x}_2$	Mean value of sample 2
$x_{i,j}$	The j-th component of the i-th data point of the sample
x_j	The j-th component of the variable x
$\mathbf{1}$	Vector with only ones
$\mathbf{a}$	Vector of the variances of the sample
$\mathbf{A}$	The correlation matrix

Continued on next page

Symbol	Description
$\boldsymbol{c}$	Vector of constants
$\mathbf{L}$	Lower triangular matrix
$\mathbf{r}_0$	Global domain of parameter
$\mathbf{R_c}$	The correlated standard normal distributed sample
$\mathbf{R}$	The uncorrelated standard normal distributed sample
$\mathbf{S}_1$	Covariance matrix of sample 1
$\mathbf{S}_2$	Covariance matrix of sample 2
$\mathbf{S}_{\text{pl}}$	Pooled covariance matrix
$\mathbf{S}$	Covariance matrix
$\mathbf{w}$	Width vector of the bins
$\mathbf{x}$	Multivariate normal distributed sample
$\mathbf{x}_0$	Origin of the histogram
$\mathbf{x}_0$	Start point of response surface
$\mathbf{x}_0^l$	Lower boundary of x_0
$\mathbf{x}_0^u$	Upper boundary of x_0
$\mathbf{x}_\text{i}$	Coordinates of the data point i
$\mathbf{x}_{\text{max}}$	Upper boundary of sample x
$\mathbf{x}_{\text{min}}$	Lower boundary of sample x
$\mathbf{z}$	Normal distributed sample with unit variance and a covariance of zero
$\bar{\mathbf{z}}_1$	Normalised mean value vectors of sample 2
$\bar{\mathbf{z}}_2$	Normalised mean value vectors of sample 1

List of abbreviations

Abbreviation	Full name
A-D	Anderson -Darling
AIAA	American Institute of Aeronautics and Astronautics
ASME	American Society of Mechanical Engineers
DoD	Department of Defense
EPDF	Estimated Probability Density Function
FORM	First Order Reliability Method
IBDM	Image Based Distance Measure
IBVM	Image Based Validation Method
K-S	Kolmogorov-Smirnov
LHS	Latin Hypercube Sampling
MCS	Monte Carlo Simulation
SORM	Second Order Reliability Method
SRS	Simple Random Sampling
V &V	Verification and Validation

1 Introduction

1.1 Background

Engineers and scientists use a variety of simulations during product development to determine how sophisticated the product behaves in operation. In this context, a product is an assembly, a component or a material. To describe how these products behave, performance relevant quantities are used. Some examples of performance quantities are: weight, stiffness, lifetime, efficiency and energy consumption. Typically, there are two ways to determine these performance quantities as presented in figure 1.1. One way is to simulate the behaviour of the product, the other way is to experimentally measure the behaviour of the product. Often it is not possible to determine all performance quantities using one way or the other. Therefore, it is necessary to combine the experimental and numerical approaches to obtain the necessary information about the real behaviour. To obtain consistent results, it is essential that both methods represent reality accurately.

1.1.1 The experimental branch

During the development of an experimental setup to measure multiple performance quantities, it is often discovered that the performance quantities cannot be measured directly (e.g. the gas tem-

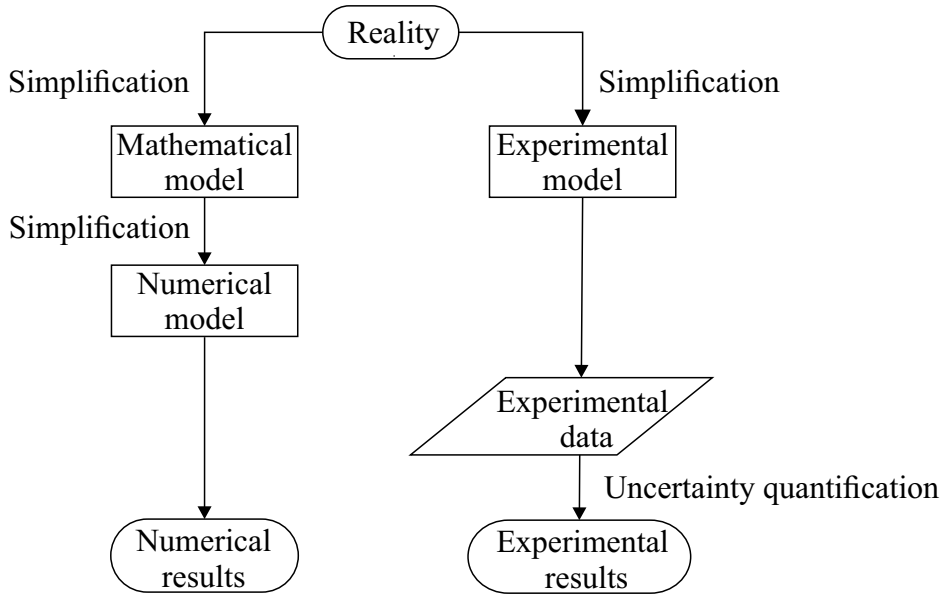


Figure 1.1: Different ways to make a statement about the behaviour of a product. (figure based on [1])

perature in a combustion chamber) or that measuring the performance quantities is too expensive (e.g. crash behaviour of an aeroplane). Therefore, it is necessary to resort to a simplified representation of reality, e.g. an experiment under laboratory conditions. This experiment has to capture the physical process of interest. Each assumption could cause the measured performance quantities to divert from reality. Therefore, it is necessary to check if each assumption is valid to ensure the experimental results represent reality. For such checks it could be required to perform some preliminary calculations. If it is asserted that the influence of the simplifications on the performance quantities are insignificant, it is concluded that the experiment results represent reality. In addition to the simplification of reality by means of laboratory conditions, it is often observed that the measurement equipment itself has an influence on the performance quantities.
After the experimental setup is defined, it is necessary to investigate the influence of various sources of uncertainty on the performance quantities to determine if the measurements are performed accurately with respect to the requirements of the product. These uncertainties include manufacturing variations and measurement errors, but are not limited to these sources. The combination of the results of the uncertainty analysis and experimentally recorded data form the experimental results [2], as presented in figure 1.1.
To avoid confusion about possible experimental simplifications, two examples are presented below. To calculate the thermal expansion coefficient of a material, one could measure the change of length of a product due to a known temperature increase. For this method it is assumed that the thermal expansion coefficient is constant over the measurement length. Another example of an experimental simplification is the use of scaled models to measure the performance quantities. This is a common practice in aerodynamics, because it is often not feasible to measure the aerodynamic behaviour at full scale due to equipment limitations and budget. However, performing aerodynamic measurements on scaled aerodynamic models gives results at a different Mach or Reynolds number compared to the full scale model. Thereby, the engineer assumes that the deviation in Mach or Reynolds number is acceptable.

1.1.2 The modeling branch

To obtain performance quantities by means of a simulation, three steps are required as presented in figure 1.1. In the first step, a mathematical model is created. This mathematical model is a set of equations and modelling data that describes reality, which includes geometric data, governing equations, constitutive equations, initial conditions and boundary conditions [2].
In the next step a numerical model is developed. This is a numerical implementation of the mathematical model to calculate the performance quantities [2]. Although a numerical model can be an analytical solver implementation if the problem is simple enough, it is common for mechanical engineers to use approximation methods like the boundary element method and the finite element method to solve complex problems. Such a numerical model includes the type and degree of discretisation of the geometry, the temporal discretisation of the governing equations and the iterative convergence criteria. It is beyond the scope of this work to explain the theory of such methods like the boundary element method and the finite element method. The reader is referred to literature for further details e.g. [3, 4].
In the third step, the numerical results are obtained from the numerical model and the accuracy of these numerical results is determined by means of a convergence study. Furthermore, it is determined which assumptions are unverified and which are verified. These unverified assumptions could cause the numerical results to differ from the experimental results.

1.1.3 Qualification, Verification and Validation

To determine if a model is an accurate representation of reality, it is mandatory to perform the following three steps: qualification, verification and validation. [2, 5, 6, 7, 8, 9, 10, 11] These steps are schematically presented in figure 1.2. To avoid confusion about the definition of these terms, the following definitions are used throughout this document, which are consistent with the definitions published by the DoD [6, 9], AIAA [10, 12] and ASME [2].

Qualification:	The process of determining the adequacy of a mathematical model to provide an acceptable level of agreement for the domain of intended application. [8]
Verification:	The process of determining that a computerised model accurately represents the underlying mathematical model and its solution. [2, 6]
Validation:	The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended use of the model. [2, 6]

The first step of the process to validate a computerised model is analysing the problem and defining the intended use of the model. Based on the problem analysis, a high quality mathematical model is generated by following guidelines and documented rules. These guidelines and documented rules are used to perform qualification of the mathematical model as shown in figure 1.2. Performing qualification ensures that an adequate mathematical model is created.
Based on the mathematical model, the numerical model is generated as presented in figure 1.1. During programming of the numerical model, some additional simplifications or assumptions have to be made as explained in subsection 1.1.2. In the second step, verification, it is investigated which assumptions and simplifications of the numerical model are sound. The more assumptions and simplifications are proven valid, the less model variables remain possibly incorrect. Verification includes determining the adequacy of the mesh resolution and code implementation checks, but it is not limited to these. For example, solving a similar problem with a known solution to determine if the computerised model is implemented correctly is also called verification.
Subsequently, validation of the numerical model is performed. During validation it is determined if the results of the numerical model differ from reality where reality is represented by the experimental results. This implies that the difference between reality and experimental results is

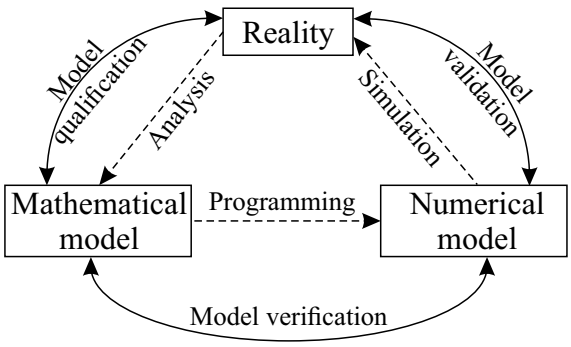


Figure 1.2: Phases of Verification & Validation during model development. (figure redrawn from [8])

neglectable (see figure 1.1). Although validation by definition is comparing the numerical results to the experimental results, an engineer is interested in a valid model. Therefore, it is necessary to decrease the difference between the numerical results and the experimental results until a valid numerical model is obtained by changing unverified assumptions of the numerical model. To minimise this difference an optimisation routine is used. The optimisation routine changes the possible unverified model variables until a valid numerical model is obtained. The less unverified variables remain after verification, the less degrees of freedom remain for the optimisation problem (minimising the difference between simulation and reality) which has to be solved to obtain a valid numerical model.
In the context of validation, it is essential to point out that the performance quantities vary in reality due to variation in the production process and the operational conditions to which the product is subjected. [13] Therefore, it would be appropriate to compare stochastic experimental results to stochastic numerical results [14]. The closer we operate products at their limits, the more important it is to correctly predict the performance of these products to avoid malfunctioning products. To correctly predict the performance of these products, it is required to use a valid numerical model. Consequently, it becomes more important to validate numerical models.

1.2 Scope

As explained in the previous sections, the performance of a product is commonly described with multiple quantities that are stochastic by nature. Since the performance quantities are generally arbitrarily distributed [15, 16], it is required to use a validation method suitable for multivariate results which are arbitrarily distributed. However, currently no validation method exists to compare two multivariate arbitrarily distributed data sets. Therefore, this work is focused on validating computerised models using experimental data and numerical results, which are multivariate arbitrarily distributed. Furthermore, an optimisation routine is used to decrease the difference between numerical and experimental results if the difference is too large.

1.3 Outline of the thesis

In chapter 2 of this document, it is identified which type of validation problems can be solved and which are currently unsolved. These unsolved issues are used to define the research goal as well as the research approach. The essential theoretical background related to validation is presented in chapter 3. This chapter is divided into the following four sections: Errors and uncertainties, experimental, simulation and validation theory. The new developed validation method is presented in chapter 4. To investigate the performance of the developed validation method, it is compared to some of the currently used validation methods by means of benchmarks which are presented in chapter 5. Moreover, a multivariate engineering validation problem is used to demonstrate the developed validation method under engineering conditions, which is presented in chapter 6. By means of this engineering problem, the problems of validation are addressed and the methods used to solve these issues with the developed validation method are presented. In the final chapter, the overall conclusions are drawn from the benchmark results and the results of the engineering relevant validation example with respect to the developed validation method.

2 Thesis justification

This chapter is divided into two sections, namely a brief description of the state of the art validation methods and an explanation of the research approach. These descriptions of the state of the art validation methods are necessary to identify currently unsolved issues. In the second section, the research approach is presented, which is used to develop a validation method for arbitrary multivariate results. Furthermore, it is explained why these steps are necessary to answer the research question: How to validate a numerical model using multivariate arbitrary distributed results?

2.1 State of the art

As explained in subsection 1.1.3, the goal of a validation process is to quantify the difference between a numerical model and an experimental model as well as the minimisation of the difference between them if the difference is found to be too large [8, 10, 11, 17]. This process to obtain a valid numerical model is presented in figure 2.1. To quantify the difference between numerical results and experimental results, it is mandatory to define a distance measure. The next step is to determine if the distance is too large by means of a validation criteria. If the distance is too large, it is necessary to change unverified assumptions of the numerical model by means of an optimisation loop to obtain a valid numerical model. The optimisation process is an iterative process and requires to solve the numerical model every time an assumption is changed.

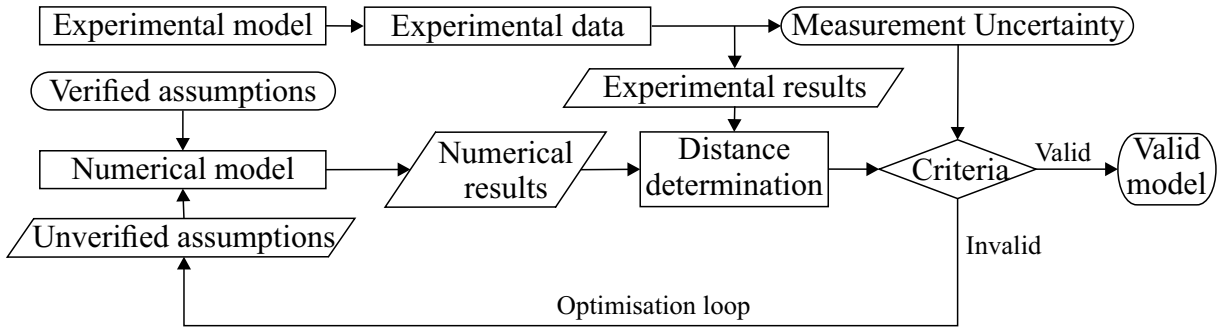


Figure 2.1: Schematic overview of the validation process.

2.1.1 Difference quantification and validation criteria

The first step of validation is quantifying the difference between numerical results and experimental results. To quantify the difference, it is mandatory to use a distance measure that describes the difference adequately. However, the currently used distance measures describe the difference only adequately if certain assumptions about the results are true. Based on these assumptions the following cases are defined in ascending order of complexity: Deterministic performance quantities, univariate normally distributed performance quantity, univariate arbitrarily distributed performance quantity, multivariate normally distributed quantities and multivariate arbitrarily distributed quantities. Each case has its own specific distance measures and validation criteria as schematically presented in figure 2.2. Further details about these cases are presented in the following subsections.

Deterministic performance quantities

The distance between two deterministic multivariate results is determined with the euclidean distance as shown in figure 2.2. To determine if the numerical model is valid, it is determined if the euclidean distance is larger or smaller than the validation criteria. These criteria consist of a criterion based on the measurement uncertainty and a criterion based on the required model accuracy. If the absolute difference is less than the measurement uncertainty and less than the acceptable model inaccuracy, the numerical model is valid for this response quantity. Moreover, it is possible to validate each response quantity separately, which could lead to a partially validated model.
More common in practice is comparing the results of a deterministic numerical model to the experimental results from repeated experiments. To determine the difference between arbitrarily distributed experimental results and the results of a single numerical simulation, it is common practice to reduce the arbitrarily distributed results to a single characteristic value as shown in figure 2.2. This characteristic value represents the nominal experimental result which is commonly defined as median or mean value. Since the experimental results are reduced to a single value for each performance quantity, the problem is solved as if the problem was deterministic.

Stochastic peformance quantities

As mentioned in chapter 1, it would be appropriate to compare stochastic experimental results to stochastic numerical results. At this point, the details regarding creating such a numerical model are irrelevant and will be introduced in chapter 3. Assuming one could create a numerical model that represents the stochastic nature of the experimental results, it is necessary to quantify the distance between the numerical and experimental arbitrarily distributed results for validation. Since the experimental results are stochastic by nature, it is likely that the results of a second series of experiments are slightly different from the first series. This will lead to variation of the measured distance. Consequently, the distance is not deterministic but is also has a certain distribution. Since the quantified distance is also stochastic, it is necessary to determine whether the distance between the numerical results and the experimental results is stochastically significant. At this point it is sufficient to know that the quantified distances are stochastic. The details concerning the theoretical background will be presented in chapter 3.

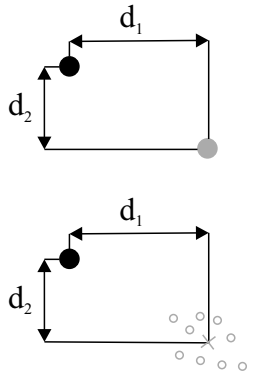
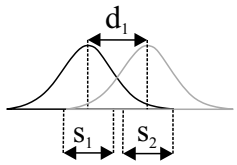
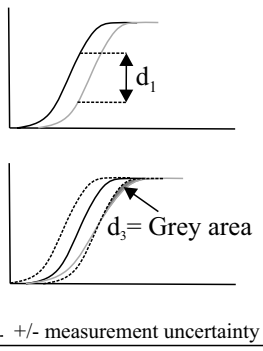
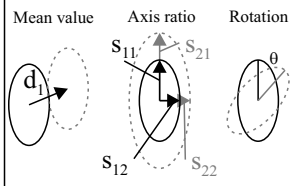
Problem type	Problem example **Experimental results** Numerical results	Distance	Criteria
Deterministic, N-D		$d_e^2 = d_1^2 + d_2^2$ d_e^2 = euclidean	d_e < Measurement uncertainty
Stochastic, Gaussian distributed, 1-D		d_1 = euclidean $d_2 = s_1/s_2$	d_1 < Measurement uncertainty d_2 = Insignificant
Stochastic, arbitrary distributed, 1-D		e.g. d_1 = K-S statistic d_2 = A-D statistic d_3 = P-box	d_1 = Insignificant d_2 = Insignificant d_3 = 0
Stochastic, Gaussian distributed, N-D		d_1 = euclidean $d_2 = s_{11}/s_{21}$ $d_3 = s_{12}/s_{22}$ $d_4 = \theta$	d_1 < Measurement uncertainty d_2 = Insignificant d_3 = Insignificant d_4 = Insignificant
Stochastic, arbitrary distributed, N-D		to be developed	to be developed

Figure 2.2: Overview of the types of validation problems, their distance measures and validation criteria.

Univariate normally distributed performance quantities

If the univariate arbitrary distributed numerical and experimental results are assumed to be normally distributed, the distance between them is equal to the difference between the characteristic parameters of the normal distributions as illustrated in figure 2.2. These characteristic parameters are the mean value and the standard deviation of the distribution. The mean value difference and the difference in standard deviation are calculated with equation 2.1 and 2.2, respectively [18, 19, 20].

$$d_1 = \mu_1 - \mu_2 \tag{2.1}$$

$$d_2 = \frac{\sigma_1}{\sigma_2} \tag{2.2}$$

Where μ_1 is the mean value of normal distribution 1 and σ_1 is its standard deviation. Correspondent variables are used for normal distribution 2. To determine if the numerical model is valid, it is essential to determine whether the distances are stochastically significant and if the distance is larger than the measurement uncertainty and accuracy of the numerical model.

Univariate arbitrarily distributed performance quantities

Removing the assumption that the results are normal distribution increases the complexity to measure the distance between two samples, because the distribution is unknown. To quantify the distance between two unknown distributions, goodness of fit statistics are used. These measure the difference between the two distribution functions on the probability axis. A few examples of such statistics are the Kolmogorov-Smirnov statistic [19, 21], the Anderson-Darling statistic [19, 22] and the Chi squared statistic [19]. Furthermore, it is possible to calculate the integral difference between the two distribution functions by means of equation 2.3 [23].

$$d_c = \int_{-\infty}^{+\infty} |F_1(x) - F_2(x)|\, dx \tag{2.3}$$

Where F_1 and F_2 are the cumulative distribution functions of distribution 1 and 2, respectively. An example of a method based on this principle is the P-box method [8, 17]. To conclude if the numerical model is valid, it is necessary to determine if the distance is larger than the required model accuracy and the measurement uncertainty as illustrated in figure 2.3.

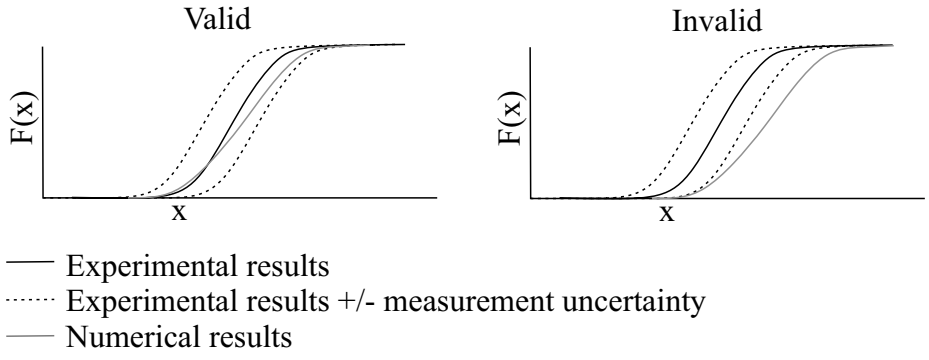


Figure 2.3: Illustration of a validation criterion for univariate arbitrarily distributed result quantities based on the measurement uncertainty.

Multivariate normally distributed quantities

The multivariate normal distribution can be plotted as isoline plot, where every isoline is a n-dimensional ellipsoid. The fact that every isoline is an ellipsoid leads to a difference quantification based on the shape of the isolines using three difference measures as presented in figure 2.4 for a bivariate normal distribution [18].

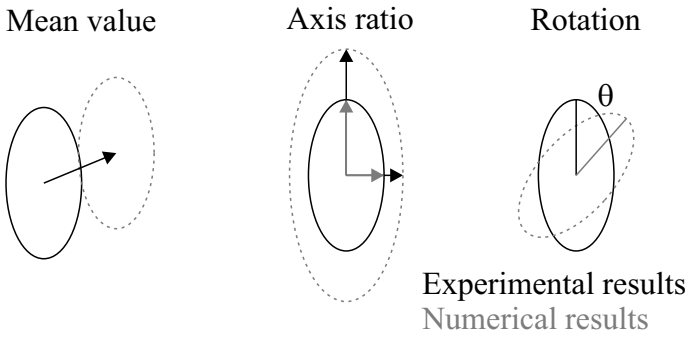


Figure 2.4: Illustration of the differences between multivariate normally distributed performance quantities.

These three distance measures describe the mean value difference, the difference between the standard deviations and the rotational difference between the distributions. To determine if the multivariate normally distributed result quantities are different, it is essential to determine if the quantified differences are stochastically significant. Furthermore, it is necessary to determine if the distance is larger than the measurement uncertainty and the required model accuracy.

Multivariate arbitrary distributed quantities

Currently, no mathematical distance exists to quantify the difference between numerical and experimental arbitrarily distributed results, because the distribution is unknown. If the distribution function would be known, it would be possible to quantify the difference between the two distributions by means of equation 2.4.

$$d_p = \int f_1(\mathbf{x}) - f_2(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{2.4}$$

Where f_1 and f_2 are the multivariate probability density functions of the numerical and the experimental results, respectively.

2.2 Objective and research approach

As explained in chapter 1, the goal of validation is to obtain a valid numerical model that represents the experimental results. To determine if a model is valid, it is essential to quantify the difference between the numerical results and the experimental results. Since these results are mostly multivariate and arbitrarily distributed [15, 16], it is essential to use a validation method which is suitable for multivariate results that are arbitrarily distributed. However, currently no method exists to validate a stochastic numerical model that produces arbitrarily distributed multivariate results [14]. Consequently, the goal of this work is to develop a suitable validation method that is capable to de-

termine if there is a significant difference between numerical and experimental multivariate results with an arbitrary distribution.
To achieve this goal, the research approach presented in figure 2.5 is developed. The first step towards developing a validation method for arbitrarily distributed results is estimating the distribution function of the arbitrarily distributed results. Retrieving the distribution function is vital to identify shape differences between the arbitrarily distributed numerical and experimental results. Based on the estimated distribution functions, a distance measure has to be defined that quantifies the difference between two distribution functions. Subsequently, it is essential to determine if the measured difference is statistically significant. Next, it is essential to define validation criteria, which are used to determine if the numerical model is valid. One of the criteria, the most essential one, is based on the experimental measurement uncertainty, because validating a numerical model beyond the accuracy of the experimental results is meaningless.
After development of the new validation method, it is essential to perform a benchmark test to quantify the effectiveness of the developed validation method. By means of a benchmark test, it is possible to compare different validation methods to show the improvements of the developed validation method. Moreover, a multivariate numerical model is validated to demonstrate the developed validation method under engineering conditions.

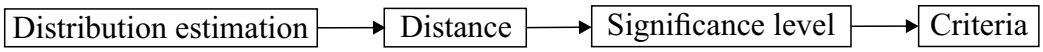


Figure 2.5: Research approach to develop a validation method for multivariate arbitrarily distributed results.

3 Fundamentals

In the previous chapters it has been explained that quantification, verification and validation are required to determine if a numerical model is valid. To perform any of these steps, it is essential to have experimental results, a numerical model and a method to determine if the numerical model represents reality. These experimental results scatter for two reasons. First of all, the experimental results scatter due to the natural variation of the specimens. Second, every experimental measurement has measurement uncertainties. The overall variation of the experimental results is a superposition of the natural variation and the measurement uncertainties.
The measurement uncertainties are used to define some of the validation criteria. These criteria are used to decide if a significant difference is present between the numerical and the experimental results. Without a criterion based on the measurement uncertainty, there is no quantitive means of determining the validity of a model [24]. If it is decided on basis of the validation criteria that there is a significant difference, it is essential to minimise this difference to obtain a valid numerical model.
In the following subsections the different types of uncertainties are presented as well as the theoretical background of the experiments, simulations and theory required to perform validation and minimisation of the distance between simulation and experiment.

3.1 Errors and uncertainties

In context of validation, errors have to be eliminated while uncertainties have to be taken into account. The following definitions for error and uncertainty are used throughout this work, which are consistent with the definitions published by ASME [2].

Error:	A recognisable deficiency in any activity of modelling or experimentation that is not due to lack of knowledge.[2]
Uncertainty:	A potential deficiency in any activity of modelling or experimentation that is due to inherent variability (aleatory uncertainty) or lack of knowledge (epistemic uncertainty). [2]

Since errors should be eliminated before performing validation, only uncertainties are explained here. Uncertainties can be divided into aleatory and epistemic uncertainties [2, 8, 17, 25]. An aleatory uncertainty represents the physical variability present in a system. It cannot be eliminated, but it might be characterised more precisely by using additional measurements. Typically, a probabilistic approach is used to characterise aleatory uncertainty [2, 8]. Epistemic uncertainties are potential deficiencies due to lack of knowledge [2, 8, 17, 25]. This can either be recognised lack of knowledge or unnoticed lack of knowledge. For example, assumptions are a form of lack of knowledge that can be quantified. Unnoticed lack of knowledge is mostly caused by human error

and it is difficult to estimate the magnitude of such uncertainties. Typically, the unnoticed lack of knowledge is minimised using procedures and protocols, different testing facilities, equipment calibration, various testing procedures, and multiple expert opinions to detect unnoticed lack of knowledge [8].
Furthermore, in many cases it is impossible to measure the desired variable directly (e.g. Young's modulus). Therefore, it is calculated using a combination of several directly measurable variables. Consequently, the measurement uncertainty of the variable of interest has to be determined indirectly using either interval arithmetic [26] or the uncertainty propagation theory of Gauss [26, 27]. Assuming the uncertainty of the directly measured variables is much smaller then the measured value, it is possible to determine the uncertainty ΔG of the desired quantity G using interval arithmetic which is a Taylor series expansion of the first order as shown in equation 3.1 [26, 27].

$$\Delta G = \left|\frac{\partial G}{\partial x}\right| \Delta x + \left|\frac{\partial G}{\partial y}\right| \Delta y + \ldots \tag{3.1}$$

The function of which one wants to determine the uncertainty is G, which depends on multiple variables (e.g. x and y). Furthermore, Δx and Δy represent the maximum uncertainty of the variable x and y respectively. If the measurements are independent of each other, it is more probable that some uncertainties counter other uncertainties. To calculate the uncertainty of a variable including this effect, equation 3.2 is used, which is the uncertainty propagation theory of Gauss [26].

$$\Delta G = \sqrt{\left(\frac{\partial G}{\partial x}\Delta x\right)^2 + \left(\frac{\partial G}{\partial y}\Delta y\right)^2 + \ldots} \tag{3.2}$$

Where G is the quantity of which one wants to calculate its uncertainty ΔG. The variables x and y are two of the variables on which the variable G depends. The constants Δx and Δy represent the maximum uncertainty of the variable x and y respectively.

3.2 Experimental fundamentals

For this work, it was chosen to validate a numerical sphere indentation model to demonstrate the developed validation method (see chapter 6). Besides the experimental indentation experiments, tensile experiments have been performed to accurately determine the stress-strain behaviour of the material. This stress-strain behaviour of the material is needed as input for the numerical model. In this section the theory used to calculate the indirectly measured variables is presented.

3.2.1 Tensile experiments

The tensile experiments used to determine the stress-strain tensile behaviour of materials can be performed using several standardised geometries [28]. During the experiments, the specimen elongation ΔL and applied force F are measured. Using these two measurements, the engineering stress σ_{eng} and strain ε_{eng} are calculated using respectively equation 3.3 [29, 30] and equation 3.4 [29, 30].

$$\sigma_{\text{eng}} = \frac{F}{A_0} \tag{3.3}$$

$$\varepsilon_{\text{eng}} = \frac{\Delta L}{L_0} \tag{3.4}$$

Here is A_0 the initial cross-sectional area and L_0 the initial gauge length of the specimen. The relation between engineering stress and strain is assumed to be linear for small strains. This linear elastic behaviour is expressed by means of the Youngs modulus, which is calculated using equation 3.5 [29],

$$E = \frac{d\sigma_{\text{eng}}}{d\varepsilon_{\text{eng}}} \tag{3.5}$$

where E, σ_{eng} and ε_{eng} are respectively the Youngs modulus, the engineering stress and engineering strain. Since the stress-strain behaviour of the material has to be used as input for the numerical indentation model, it is mandatory to transform the engineering stress-strain behaviour to the true stress-strain behaviour. The true strain is based on the rate of instantaneous increase of instantaneous gauge length. The relation between the engineering strain and the true strain is presented in equation 3.6 [31, 30],

$$\varepsilon_{\text{true}} = \ln(1+\varepsilon_{\text{eng}}) \tag{3.6}$$

where $\varepsilon_{\text{true}}$ and ε_{eng} are respectively the true strain and the engineering strain. Assuming the material volume remains constant and deforms homogeneously, the relation between engineering stress and true stress is given by equation 3.7 [31, 30],

$$\sigma_{\text{true}} = \sigma_{\text{eng}}(1+\varepsilon_{\text{eng}}) \tag{3.7}$$

where σ_{true} is the true stress and σ_{eng} the engineering stress.

3.2.2 Instrumented indentation experiments

Instrumented indentation experiments can be used to identify local material properties [32, 33]. Depending on the type of indenter (e.g. spherical or pyramidal) being used, the obtainable material properties vary. For this work a sphere indenter is used, which is a non-self-similar indenter. The strain field under the indent created by a spherical indenter depends on the ratio of indentation diameter d_i and indenter diameter D_i. This effect can be used to obtain the hardening behaviour of the material.

Evaluation of classical instrumented indentation experiments via the method of Oliver & Pharr results in indentation hardness H_{IT} and indentation modulus E_{IT}. The indentation hardness is calculated with equation 3.8 [32, 33],

$$H_{IT} = \frac{F_{\max}}{A_c(h_c)}, \qquad h_c = h_{\max} - \frac{3}{4}\frac{F_{\max}}{S} \tag{3.8}$$

where $F_{\max}$, $h_{\max}$, A_c, h_c and S are respectively the maximum applied force, the maximum indentation depth, the contact area, the indentation depth and the contact stiffness. These variables are illustrated in figure 3.1. To calculate the indentation modulus, equation 3.9 is used [32, 33].

$$E_{IT} = \frac{1-\nu^2}{\frac{1}{E_r} - \frac{1-\nu_i^2}{E_i}}, \qquad E_r = \frac{S\sqrt{\pi}}{2\sqrt{A_c}} \tag{3.9}$$

Here E_i, ν_i and ν are respectively the Youngs modulus of the indenter, the Poisson ratio of the indenter and the Poisson ratio of the material being indented. The Young's modulus and Poisson ratio of the indenter (E_i, ν_i) aswell as the Poisson ratio of the material being indented (ν) must be known to solve equation 3.9. Furthermore, the relation between contact area and indentation depth

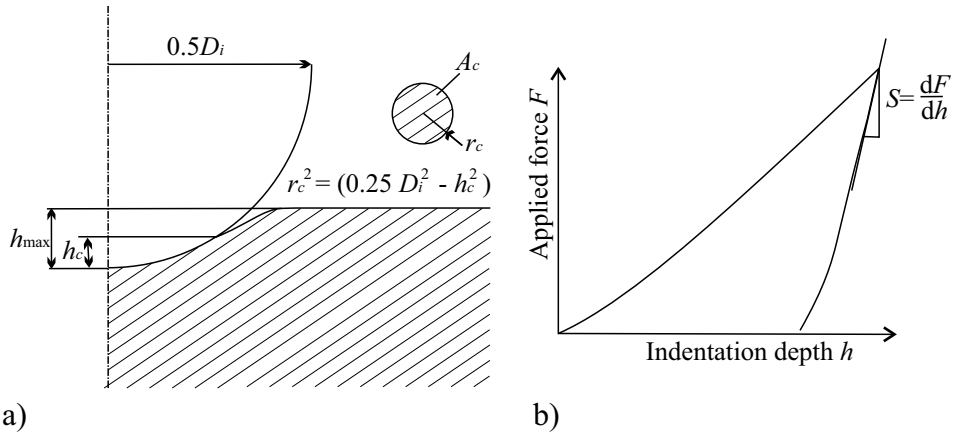


Figure 3.1: a) Sketch of the contact area during a spherical indentation experiment, b) Sketch of the force indentation depth curve of a sphere indentation experiment.

is required, which can deviate from the one of an ideal indenter. Besides the classical material properties E_{IT} and H_{IT}, it is possible to determine the hardening behaviour of the indented material by means of a semi-empirical model [34] or inverse parameter identification using finite element simulations [35, 36, 37, 38].

3.3 Simulation fundamentals

To validate a numerical model, it is mandatory to compare the experimental and numerical results with each other as explained in chapter 2. Since the experimental results are stochastic by nature and the numerical model should represent the experimental results (reality), it is reasonable to assume that the numerical results are stochastic as well. To obtain these stochastic numerical results, it is mandatory to extend the deterministic model (e.g. a finite element model) with a probabilistic method to obtain stochastic numerical results. In order to implement the probabilistic part of the model, it is essential to quantify the random variation of the input parameters of the numerical model which are presented exemplarily on the left side of figure 3.2. Subsequently, the variation of the input variables is propagated to the output variables which is illustrated with the two arrows in figure 3.2. In order to explain further details of the probabilistic model, it is necessary to introduce the terms population, sample and distribution. These terms are explained in the next subsection. Details concerning the probabilistic model are presented in subsequential subsections.

3.3.1 Population, sample and distribution

A population in statistics is referred to the total set of observations [20]. This set of observations is often either impossible to obtain or it is too large to be practical for further stochastic analysis. Therefore, it is common practice to use samples. A sample is a set of observations which is randomly drawn from a population [20]. To determine whether a sample is an accurate representation of the population, the distribution of the sample is compared to the distribution of the population. A distribution assigns a probability to each possible value taken by a random variable. It can be defined as probability density function and cumulative distribution function. The cumulative dis-

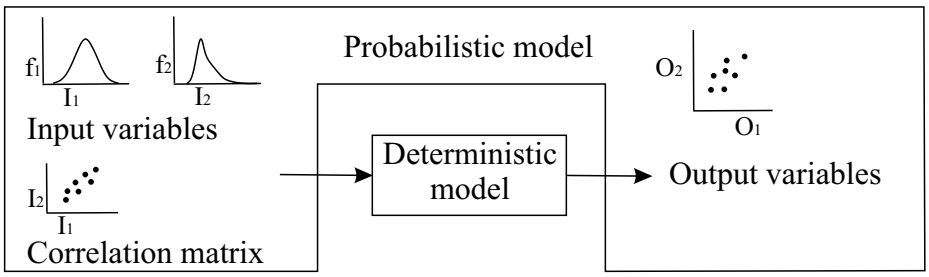


Figure 3.2: Structure of stochastic models. (figure based on [39]).

tribution function is the integrated form of the probability density function. If the integral of the probability density function is evaluated using the domain $[-\infty, \infty]$, it is equal to one as shown in equation 3.10 [19, 20, 40],

$$\int_{-\infty}^{\infty} f(x)\partial x = 1 \tag{3.10}$$

where $f(x)$ is the probability density function of the variable x.

3.3.2 Distribution identification

For a stochastic numerical model it is mandatory to randomly generate the values of the input variables. These random values are generated based on the distribution of the input variables. Therefore, it is essential to know the distribution of the input variables. However, only a sample of the population of the input variables is measured experimentally. Thus, it is essential to identify the distribution belonging to the sample. Typically, a distribution is assumed for the population which is then compared to the discrete cumulative function of the experimentally measured sample. To determine the goodness-of-fit of the assumed distribution of the population, the difference to the discrete distribution of the sample is typically determined with the Kolmogorov-Smirnov (K-S) [19, 21] or Anderson-Darling (A-D) statistic [19, 22]. The K-S statistic determines the largest difference on the probability axis between two cumulative distribution functions. Therefore, it is focussed on the central part of the cumulative distribution function. Since the maximum difference is never detected in the tails of a distribution with the K-S statistic, the A-D statistic is developed. It puts more weight at the tails of the distributions, but it is assumed that the samples are infinitely large. The A-D statistic is calculated by means of equation 3.11 [19, 22],

$$A^2 = \sum_{i=1}^{n}((2i-1)(\ln[F_x(x_i)] + \ln[1 - F_x(x_{n+1-i})])/n) - n \tag{3.11}$$

where F_x, x_i and n are respectively the proposed cumulative distribution function of the population, the i^{th} data point of the ascending sorted sample and the size of the sample. To correct the distance measure with the A-D statistic for sample size effects an correction factor is added which is distribution dependent. When the population is assumed to be normally distributed, than the correction factor is calculated with equation 3.12 [19, 22],

$$A^* = A^2\left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right) \tag{3.12}$$

where A^*, A and n are respectively the corrected A-D statistic, the uncorrected A-D statistic and the sample size. Correction factors for other distributions can be found in [19]. Based on hypothesis tests, it is determined if the differences measured with the K-S or A-D statistic are stochastically significant. Hypotheses and hypothesis tests are explained in section 3.4.4.

3.3.3 Probability distribution of the response quantities

In a stochastic model the probability distribution function of the output variables is calculated using one of the following methods: First Order Reliability Method (FORM) [41, 42], Second Order Reliability Method (SORM) [41, 42], response surface method [39] or Monte Carlo simulation [39, 43].
FORM and SORM are methods to determine the probability that a product fails. These methods approximate the border between failure and safe design using a linear or second order function, respectively. The error of the approximation functions is related to the probability of failure. Consequently, these methods are not suitable to determine the probability that two samples are different. In contrast to the reliability methods (FORM/SORM), the response surface method can be used to determine the entire distribution of the output variables. But, the accuracy of the method depends on the accuracy of the approximated response surface.
The Monte Carlo simulation method provides the probability distribution of the responses without additional approximations. Thereby, it guarantees that the probability distribution calculated for the response quantities is accurate in the domain where the probability density is large. Consequently, this method is the preferred method to create numerical models for validation. Further details about a Monte Carlo simulation are presented in the next section.

3.3.4 Monte Carlo simulation

A Monte Carlo simulation uses the deterministic model as a "black box" to create a probabilistic model. For a Monte Carlo simulation the input and output variables of the deterministic model are considered as random variables. The values of these random input variables are taken randomly from the population of the input variables. [43, 44] The values of the random output variables corresponding to these input values are obtained by solving the deterministic model multiple times. These values for the random input variables are generated using the following steps.
First, a randomly distributed multivariate sample is generated with an uniform distribution between zero and one for each random input variable using the Simple Random Sampling (SRS) method [40, 43]. When it is required to generate a small sample, which represents the uniform distribution accurately, the Latin Hypercube Sampling (LHS) method is used [45]. For this method the range [0, 1] is divided into N equal sub-ranges, where in each sub-range one random data point is created SRS method. This results in greater numerical efficiency.
The uniformly distributed multivariate sample is transformed into a standard normally distributed multivariate sample. Commonly the variables of the multivariate population are not independent of each other. Therefore, it is essential to generate multivariate samples, which poses the correlation present in the population. The multivariate variables of the standard normally distributed sample are correlated with each other using the Cholesky decomposition shown in equation 3.13 [46, 47],

$$\mathbf{A} = \mathbf{L^T L}, \qquad \mathbf{R_c} = \mathbf{R\,L} \tag{3.13}$$

where $\mathbf{R}$, $\mathbf{R_c}$, $\mathbf{A}$, $\mathbf{L}$, $\mathbf{L^T}$ are respectively the uncorrelated standard normally distributed sample, the correlated standard normally distributed sample, the correlation matrix, a lower triangular matrix and its transpose. The correlation matrix shows the correlation coefficients between all variables

in matrix form. The correlation coefficient between two variables is calculated using equation 3.14 [19, 20, 40],

$$\rho_{xy} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}} \tag{3.14}$$

where x and y are two random variables. For successful Cholesky decomposition, the correlation matrix must be a hermitian positive definite matrix. Although the standard normally distributed samples must be uncorrelated, some correlation by chance is always present when working with random numbers. When the correlation is significant, it is required to uncorrelated the standard normally distributed samples using the Cholesky decomposition. [46] To determine if the samples are significantly correlated, a hypothesis test is performed based on the statistic presented in equation 3.15 [20, 48, 46],

$$t = r\sqrt{\frac{N-2}{1-r^2}} \tag{3.15}$$

where N is the number of data points in the sample and r is the correlation coefficient. The quantity t is student-t distributed with N-2 degrees of freedom under the null hypothesis $r_1 = r_2$ [20, 48, 46]. When the null hypothesis is rejected, it is concluded that the samples are still correlated. This is an iterative process that is repeated till no significant correlation remains. These uncorrelated standard normally distributed variables are correlated using the correlation matrix of the normally distributed population. To obtain other distributions as the standard normal distribution, other models as the Nataf model can be applied to transform the correlated standard normally distributed samples [44].

3.4 Validation fundamentals

In this section the theory is presented, which is essential to develop a new validation method to determine whether a difference exists between stochastic numerical results and stochastic experimental results. To develop a validation method for arbitrary multivariate results, it is essential to follow the steps presented in chapter 2 (on page 10). The essential theory required to implement these steps is presented in this section and consists of the following topics: probability density estimation functions, distance measures, distributions of distance measures and hypothesis tests. Furthermore, some additional theoretical background of currently used validation methods is presented which is required to compare the newly developed validation method to the currently used validation methods.

3.4.1 Estimated density functions

In this subsection several methods are presented for constructing estimated density functions based on a sample (dataset). These methods are suitable to estimate the distribution of a sample without assuming the distribution of the population of which the sample is drawn, but each method has its own characteristic that could be undesired with respect to validation. To visualise the differences between these distribution estimation methods, only one example dataset is used because the differences are a direct consequence of the assumptions made for the methods. This example dataset is presented in table 3.1 and figure 3.3. It contains twenty bivariate results, which are generated

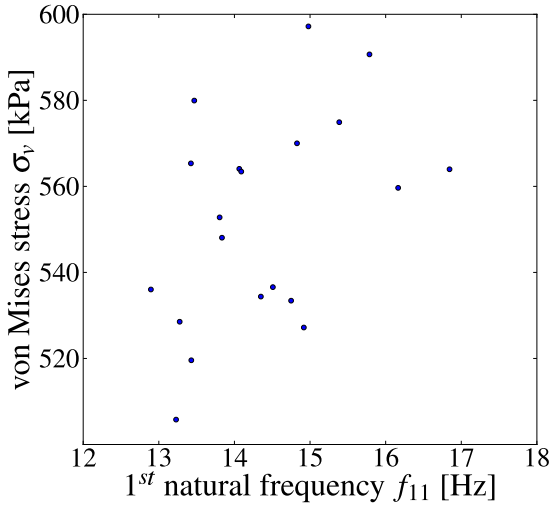


Figure 3.3: Scatter plot of the example data set containing 20 paired results.

Table 3.1: Values of the example data set.

Data point	1^{st} natural frequency [Hz]	von Mises stress [kPa]
1	16.84	563.97
2	14.35	534.38
3	15.39	574.92
4	13.42	565.35
5	14.51	536.58
6	14.98	597.18
7	13.80	552.79
8	14.06	564.07
9	13.47	579.94
10	14.09	563.44
11	12.90	536.03
12	14.75	533.43
13	13.84	548.07
14	15.79	590.68
15	14.92	527.19
16	16.17	559.65
17	13.23	505.79
18	13.43	519.58
19	14.83	570.01
20	13.28	528.54

with a Monte Carlo simulation that calculates the first natural frequency of a simply supported plate and the von Mises stress in a thin preloaded aluminium plate based on stochastic input variables. The first natural frequency f_{11} is calculated using equation 3.16 [49],

$$f_{11} = \sqrt{\left(\frac{\pi^2(1+\frac{a^2}{b^2})^2}{2\pi a^2}\sqrt{\frac{Et^2}{12\rho_d(1-\nu^2)}}\right)^2 + \frac{F_1}{4\rho_d t a^2 b} + \frac{F_2}{4\rho_d t a b^2}} \tag{3.16}$$

where a, b, t, E, ρ, ν, F_1, F_2 are respectively the length of the plate, the width of the plate, the thickness of the plate, the Youngs modulus, the density of the material, the Poisson ratio, the applied normal force on the length of the plate and the applied force on the width of the plate. The von Mises stress σ_v is calculated with equation 3.17 [31, 50],

$$\sigma_v = \sqrt{\frac{{F_1}^2}{ta} + \frac{{F_2}^2}{tb} - \frac{F_1}{ta}\frac{F_2}{tb}} \tag{3.17}$$

The input variables are chosen as normally distributed which produces an unknown distribution for the first natural frequency and the von Mises stress due to the nonlinear relation between input variables and the response quantities as presented in equation 3.16 and equation 3.17.

The histogram method

Histograms are the oldest and the most used density estimators that are created using so called bins [15, 40]. The estimated probability density function is calculated using equation 3.18 [15],

$$f(\mathbf{x}) = \frac{1}{n\mathbf{1}\mathbf{w}^{\mathrm{t}}}\sum_{i=1}^{n} k_i(\mathbf{x}) \tag{3.18}$$

where $f(\mathbf{x})$, n, $\mathbf{w}^{\mathrm{t}}$ and $\mathbf{1}$ are respectively the estimated probability density function, the amount of data points, the transposed multivariate width vector of the bins and a vector with only entries equal to 1. Furthermore, the variable k_i is the indicator function that determines in which bin the data points lie. This indicator function is calculated using equation 3.19 [15],

$$k_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x}_0 + \left(\left\lceil\frac{\mathbf{x}-\mathbf{x}_0}{\mathbf{w}}\right\rceil - \mathbf{1}\right)\mathbf{w} < \mathbf{x}_{\mathrm{i}} \leq \mathbf{x}_0 + \left\lceil\frac{\mathbf{x}-\mathbf{x}_0}{\mathbf{w}}\right\rceil \mathbf{w} \\ 0 & \text{if } \mathbf{x}_0 + \left(\left\lceil\frac{\mathbf{x}-\mathbf{x}_0}{\mathbf{w}}\right\rceil - \mathbf{1}\right)\mathbf{w} \geq \mathbf{x}_{\mathrm{i}} > \mathbf{x}_0 + \left\lceil\frac{\mathbf{x}-\mathbf{x}_0}{\mathbf{w}}\right\rceil \mathbf{w} \end{cases} \tag{3.19}$$

where $k_i(\mathbf{x})$, $\mathbf{w}$, $\mathbf{x}_{\mathrm{i}}$, $\mathbf{x}_0$ are respectively the indicator function, the vector of the bin width, the coordinates of data point i and the origin of the histogram. Since the bins (intervals) are open on the left side and closed on the right side, it is essential to use the ceiling function to calculate the value of the indicator function. Although the histogram method is a common used estimator, it has the following two drawbacks: the estimated probability density function is not continuous and it depends on the chosen origin ($\mathbf{x}_0$) and bin width ($\mathbf{w}$). [15] The discontinuous behaviour of the estimated probability density function is shown in figure 3.4, which is generated using the example data presented at the beginning of this section.

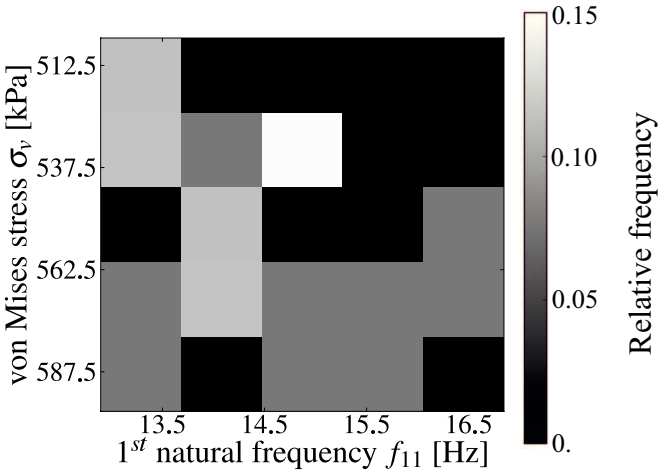


Figure 3.4: Discretised estimated probability density function of the example dataset based on the histogram method.

Balloon density estimator

To estimate the probability density function of a sample without a user selected start point (origin), it is essential to use an estimation method which is based on the sample points. The simplest possible estimator is a rotation symmetric estimator with a uniform density, which is called balloon estimator [16] or naive estimator [15]. In this document, it is chosen to use the name balloon estimator. The estimated probability density function based on the balloon density estimator in d-dimensions is calculated with equation 3.20 [15],

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{k_i(\mathbf{x})}{c_d\, r^d} \tag{3.20}$$

where n, r, d and c_d are respectively the amount of sample data points, the radius of the balloon, the dimensionality of the balloon and the n-dimensional volume of a unit sphere. If the sphere is 2-dimensional, c_d is equal to π. Furthermore, $k_i(\mathbf{x})$ is the indicator function which indicates if a random value $\mathbf{x}$ is within the balloon around sample data point $\mathbf{x}_\mathrm{i}$. This indicator function is calculated using equation 3.21 [15],

$$k_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \sqrt{(\mathbf{x}-\mathbf{x}_\mathrm{i})^T(\mathbf{x}-\mathbf{x}_\mathrm{i})} \leq r \\ 0 & \text{if } \sqrt{(\mathbf{x}-\mathbf{x}_\mathrm{i})^T(\mathbf{x}-\mathbf{x}_\mathrm{i})} > r \end{cases} \tag{3.21}$$

where $\mathbf{x}_\mathrm{i}$ is a vector with the coordinates of sample data point i and r is the chosen balloon radius. By using sphere shaped balloons the method is limited to samples with an equal variance for all components of the sample. To overcome this issue, it is mandatory to transform the spherical balloons into an ellipsoidal balloons where the axis ratio is equal to the ratio of the variances of the multivariate sample. The probability density function based on ellipsoidal balloons is calculated by substituting equation 3.22 in equation 3.20 [15],

$$k_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \sqrt{(\mathbf{x}-\mathbf{x}_\mathrm{i})^T\left(\frac{1}{\mathbf{a}}(\mathbf{x}-\mathbf{x}_\mathrm{i})\right)} \leq r^* \\ 0 & \text{if } \sqrt{(\mathbf{x}-\mathbf{x}_\mathrm{i})^T\left(\frac{1}{\mathbf{a}}(\mathbf{x}-\mathbf{x}_\mathrm{i})\right)} > r^* \end{cases} \tag{3.22}$$

where **a** and r^* are the vector of variances of the sample and the selected balloon radius r normalised by the smallest standard deviation in vector **a**. [15]. To visualise the characteristics of the balloon estimator, the estimated probability density function based of the example dataset (table 3.1) is shown in figure 3.5a. The circular shape of the visualised balloons in figure 3.5a is caused by the axis ratio of the diagram. Any other axis ratio would result in elliptical balloons. Although the balloon estimator does not require a user selected origin, the estimated probability density function still shows the discontinuous character of the histogram method.
Alternatively to using a constant balloon radius, it is also possible to use a variable balloon radius based on the nearest neighbour distance [15]. Instead of using a constant balloon radius, the balloon radius is different for each data point. Hence, the estimated probability density function is calculated with equation 3.23 [15],

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{k_i(\mathbf{x})}{c_d\, {r_i}^d} \tag{3.23}$$

where r_i is the data point dependent balloon radius. The other variables in this equation remain identical to their definition for a constant balloon radius except the indicator function which is calculated using equation 3.24 [15],

$$k_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \sqrt{(\mathbf{x}-\mathbf{x_i})^T\left(\frac{1}{\mathbf{a}}(\mathbf{x}-\mathbf{x_i})\right)} \leq r_i \\ 0 & \text{if } \sqrt{(\mathbf{x}-\mathbf{x_i})^T\left(\frac{1}{\mathbf{a}}(\mathbf{x}-\mathbf{x_i})\right)} > r_i \end{cases} \tag{3.24}$$

where r_i is the data point dependent radius. The other variables are previously defined for equation 3.21. Applying the variable balloon density estimator to the example data presented in table 3.1 gives the estimated probability density function that is shown in figure 3.5b. The main difference between the balloon density estimator and the variable balloon density estimator is the amount of detail preserved in different parts of the estimated probability density function.

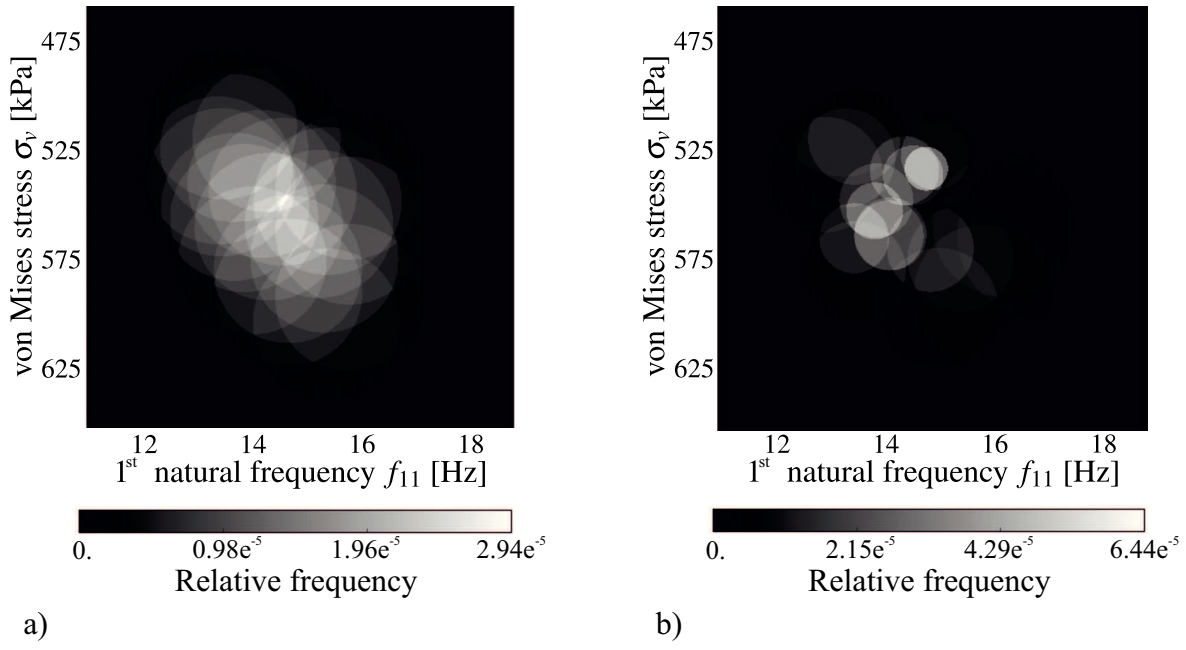


Figure 3.5: Discretised estimated probability function of the example dataset using 500 x 500 pixels based on: a) The balloon estimator, b) The variable balloon estimator.

Kernel density estimator

The kernel density estimator is a generalisation of the balloon density estimator that is continuous if and only if the kernel itself is continuous [15]. Although, any probability density function can be used as kernel, it is common to use a symmetric kernel. A typical kernel is the multivariate gaussian probability density function as shown in equation 3.25 [15],

$$k_i(\mathbf{x}) = \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-0.5\frac{(x_j - x_{i,j})^2}{\sigma_j{}^2}} \tag{3.25}$$

where σ_j, d and $x_{i,j}$ are the j^{th} component of standard deviation of the multivariate gaussian kernel, the dimensionality of the kernel and the j^{th} component of the i^{th} sample data point respectively. Hence, the estimated probability density function using a gaussian kernel is calculated with equation 3.26 [15, 51],

$$f(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} k_i(\mathbf{x}) \tag{3.26}$$

where n and $k_i(\mathbf{x})$ are the size of the sample and the kernel density function positioned at data point i respectively. For an optimal result, the kernel width should be chosen such that difference between the estimated distribution function and the probability density function of the population is minimal. Using the approximate integrated mean error, the optimal kernel width can be calculated using equation 3.27 [15],

$$\sigma_j^{d+4} = \frac{d\,C_2}{C_1{}^2 n \int (\nabla^2 f_j)^2} \tag{3.27}$$

where d, ∇^2, C_1 and C_2 are the dimensionality of the problem, the laplacian operator and two kernel dependent constants respectively. Furthermore, the function f_j is the j^{th} edge distribution of the multivariate probability density function of the population from which the sample is drawn. The definition of the constants C_1 and C_2 are derived for a gaussian kernel in appendix B. Since equation 3.27 depends on the assumed distribution of the population of which the sample is

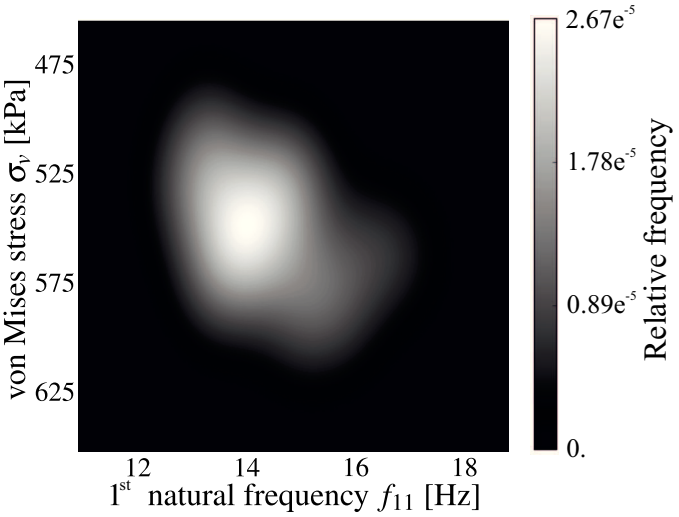


Figure 3.6: Discretised estimated probability function of the example dataset using 500 x 500 pixels based on the Gaussian kernel estimator.

drawn, it is mandatory to assume a distribution type for the population. Assuming the sample is drawn from a multivariate normal distribution, the equation to calculate the optimal kernel width simplifies to equation 3.28 [15, 51],

$$\sigma_{\text{opt}} = \sigma_{\text{data}}\, n^{-1/(d+4)} \tag{3.28}$$

where d is the problem dimensionality, σ_{data} the standard deviation of the data and n the amount of data points in the sample. Although this solution is derived for assumed normally distributed data, it works well for distributions which moderately deviate from normality. To investigate the error made when using equation 3.28 instead of 3.27 when the sample is non-normally distributed, it is necessary to evaluate both equations for several non-normally distributed samples. In literature [15], it is published that the optimal kernel width is insensitive for the kurtosis which is investigated by using a sample that is student-t distributed. Even if the samples are drawn from a multimodal normal distribution, the error made by using equation 3.28 is small [15]. Furthermore, it is published in literature [15] that the error is still less than 18% when the optimal kernel width is calculated with equation 3.28 for lognormally distributed samples with a skewness up to 1.0. Of course, for heavily skewed data or data with multiple significantly separated peaks in the probability density function, equation 3.28 will produce a significant error. The estimated probability density function of the example dataset (see table 3.1) is calculated with equation 3.26 and it is visualised in figure 3.6. This estimated probability density function is generated using equation 3.28 to define the kernel width.

3.4.2 Distance measurement methods

The next step required in the validation process of a numerical model is quantification of the difference between the numerical results and the experimental results as presented in figure 2.5. To quantify the difference, it is essential to mathematically describe the difference between the numerical and experimental results. This difference can be measured using many distance measures. The distance measures relevant for this work are the mean value difference, the covariance matrix difference, the univariate distributional difference and the multivariate distributional difference. Although the goal of this work is to develop a validation method for arbitrarily distributed multivariate samples, also some univariate distance measures are presented. These univariate distance measures are used to explain the different type of distance measures before generalising these distance measures to the multivariate case.

Mean value and covariance matrix sample difference

The distribution of samples is commonly assumed to be normally distributed. Consequently, the mean value difference is the mostly used distance to quantify the difference between two samples. Whether the quantified distance is significant depends on the variation present in the samples. To eliminate this dependancy, it is mandatory to normalise the mean value difference by means of the sample variation. This normalisation is possible for multivariate as well as univariate samples.
For a univariate sample, the mean value difference is normalised using the sample standard deviation as presented in equation 3.29 [18, 19, 20, 40],

$$D_m = \frac{|\bar{x}_2 - \bar{x}_1|}{s_{\text{pl}}} \tag{3.29}$$

where $\bar{x}_1$, $\bar{x}_1$ and s_{pl} represent the mean value of sample one, the mean value of sample two and the pooled standard deviation. The pooled variance is calculated with equation 3.30 [18, 19, 20, 40],

$$s_{\text{pl}}{}^2 = \frac{(n_1-1)s_1{}^2 + (n_2-1)s_2{}^2}{n_1+n_2-2} \tag{3.30}$$

where n_1, n_2, s_1 and s_2 are respectively the sample size of sample one and two and the standard deviation of sample one and two. Equation 3.30 is only valid if the standard deviation of the populations of which the samples are drawn are identical [18, 19, 20, 40, 52]. To verify this assumption, it is necessary to calculate the difference between the standard deviation of the samples with equation 3.31 [18, 20, 40],

$$D_v = \frac{s_1{}^2}{s_2{}^2} \tag{3.31}$$

where s_1 and s_2 are the standard deviation of sample one and two respectively. The distance D_v is used to measure the difference between the sample variances and is used in combination with the mean value difference to determine the difference between two sample drawn form normal distributions.
Similar to the univariate case is it essential to normalise the euclidean mean value distance to determine if the samples are significantly different. Such a normalised sample has a multivariate normal distribution with unit variance and a covariance equal to zero. This transformation is expressed in equation 3.32 [18],

$$\mathbf{z} = \mathbf{S}^{-1/2}\,\mathbf{x} \tag{3.32}$$

where $\mathbf{x}$ is the multivariate normally distributed sample and $\mathbf{S}$ is the covariance matrix of this sample. The variable $\mathbf{z}$ is the normalised sample with unit variance and a covariance of zero. Concerning this transformation one remark has to be made. The transformation is only possible if the covariance matrix is a positive definite matrix. Combining the euclidean distance between the normalised samples with the transformation gives the Mahalanobis distance which is calculated with equation 3.33 [18],

$$D_{\text{M}}{}^2 = (\bar{\mathbf{z}}_2 - \bar{\mathbf{z}}_1)^{\text{t}}(\bar{\mathbf{z}}_2 - \bar{\mathbf{z}}_1) = (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^{\text{t}}\mathbf{S}_{\text{pl}}^{-1}(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1) \tag{3.33}$$

where $\bar{\mathbf{z}}_1$, $\bar{\mathbf{z}}_2$, $\bar{\mathbf{x}}_1$, $\bar{\mathbf{x}}_2$ and $\mathbf{S}_{\text{pl}}$ are respectively the normalised mean value vectors of sample one and two, the mean value vectors of samples one and two and the pooled covariance matrix. The pooled covariance matrix is calculated using equation 3.34 [18],

$$\mathbf{S}_{\text{pl}} = \frac{(n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2}{n_1+n_2-2} \tag{3.34}$$

where the variable $\mathbf{S}_1$, $\mathbf{S}_2$, n_1 and n_2 are the covariance matrix of sample one and two and the sample size of the sample one and two. Moreover, it is assumed that the covariance matrices of the populations are identical [18]. In addition, it is necessary that the condition $n_1+n_2-2 \geq d$ is true in order to have a non-singular pooled covariance matrix, where d is the dimensionality of the problem. These singularities are caused by dependent variables, which are no random variables and should be excluded before determining the distance. Exclusion of the dependent variables leads to a positive definite covariance matrix. Calculating the Mahalanobis distance based on such a reduced dataset is common practice [18]. To verify whether the covariance matrices are identical of the populations of which the samples are drawn, it is necessary to determine the difference between

the covariance matrices of the samples. This difference is calculated with equation 3.35 [18],

$$M^2 = \left(\frac{|\mathbf{S}_1|}{|\mathbf{S}_{\text{pl}}|}\right)^{n_1-1} \left(\frac{|\mathbf{S}_2|}{|\mathbf{S}_{\text{pl}}|}\right)^{n_2-1} \tag{3.35}$$

where $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_{\text{pl}}$ are the covariance matrix of sample one and two and their pooled covariance matrix. Furthermore, the variables n_1 and n_2 are respectively the size of sample one and two.

Difference between probability density and cumulative distribution functions

If the samples (experimental and numerical results) are not normally distributed, it is essential to compare the cumulative distributions functions or the probability density functions with each other. Classically, the difference between a measured cumulative distribution function and an assumed cumulative distribution functions is determined with the Kolmogorov-Smirnov (K-S) statistic [19, 21] or Anderson-Darling (A-D) statistic [19, 22] which are methods to determine the goodness-of-fit as presented in subsection 3.3.2. Another method to quantify the difference is to integrate the difference between the two cumulative distribution functions as expressed in equation 2.3 and illustrated in figure 3.7a. Furthermore, it is possible to incorporate the measurement uncertainties to determine the difference that cannot be explained by measurement uncertainties, as shown in figure 3.7b. This method is called the P-box method [8, 17]. To determine the difference between two probability density functions, it is necessary to divide the probability density function in small intervals which are used to calculate the probability that a value falls within that interval. This method produces a histogram of the probability density function. The difference between two histograms with equal bins is calculated with equation 3.36 [20, 19, 40],

$$D_p = \sum_{i=1}^{n} \frac{(e_i - t_i)^2}{t_i} \tag{3.36}$$

where e_i, t_i and n are respectively the observed probability that a data point falls in interval i, the theoretical probability that a data point falls in interval i and the amount of intervals.

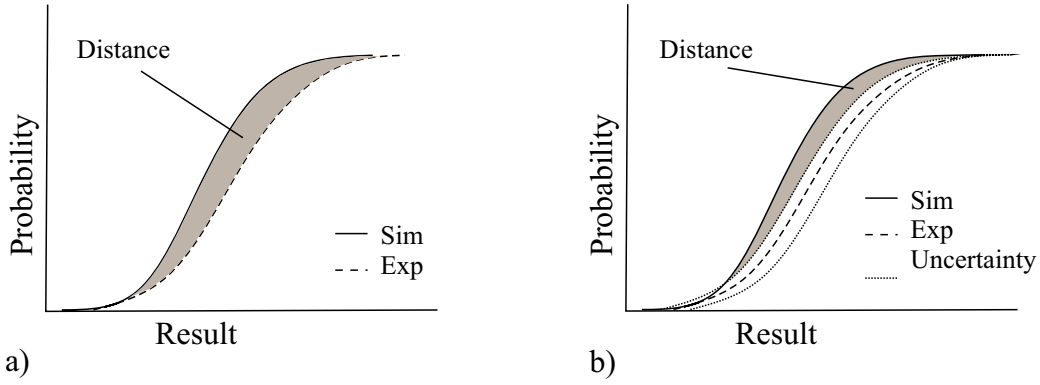


Figure 3.7: Distance between two cumulative density functions: a) Integrated difference, b) Difference according to the P-box method.

Difference between arbitrary multivariate distributions

The difference between two arbitrary distributions can be measured based on the probability density functions of the multivariate distributions. Likewise the univariate probability density function, a discretisation of the multivariate probability density functions is essential to determine the probability that a data point falls within an interval. Such a discrete probability density function is an array of probabilities whose structure corresponds to the structure of a multidimensional image where the pixel intensity is equal to the probability that a data point falls in this interval (pixel). To measure the difference between two images classical graphical techniques can be used. Biomedical researchers use these methods for example for colocalisation analysis between images [53, 54, 55, 56]. Image based difference measurement methods are typically based on the correlation between two images. A widely used colocalisation measure is the liner correlation coefficient ρ_0 which is calculated using equation 3.37 [53, 55, 54, 56],

$$\rho_0 = \frac{\sum\limits_{i=1} (S_{1\mathrm{i}} - \overline{S}_1)(S_{2\mathrm{i}} - \overline{S}_2)}{\sqrt{\sum\limits_{i=1} (S_{1\mathrm{i}} - \overline{S}_1)^2 \sum\limits_{i=1} (S_{2\mathrm{i}} - \overline{S}_2)^2}} \tag{3.37}$$

where $S_{1\mathrm{i}}$ and $\overline{S}_1$ are respectively the intensity of pixel i of image one and the mean pixel intensity of image one. Consistent notations are used for image two. This distance measure is used to analyse if a linear correlation exists between the images [19]. The main disadvantage of this distance measure is its sensitivity for the mean pixel intensity, which is highly dependent on the discretisation domain. To tackle this issue, Manders et. al introduced a modified variant of the linear correlation method by removing the subtraction of the average value as presented in equation 3.38 [53, 55, 54, 56],

$$\rho_m = \frac{\sum\limits_{i=1} (S_{1\mathrm{i}})(S_{2\mathrm{i}})}{\sqrt{\sum\limits_{i=1} (S_{1\mathrm{i}})^2 \sum\limits_{i=1} (S_{2\mathrm{i}})^2}} \tag{3.38}$$

where $S_{1\mathrm{i}}$, $S_{2\mathrm{i}}$ are respectively the intensity of pixel i of image one and the intensity of pixel i of image two. This overlap coefficient is independent of the discretisation domain, but it has become sensitive to the absolute value of a discretisation step (pixel). For image analysis in general, this could be a problem. However, it is not a problem when comparing two discrete probability density functions, because the integral of a probability density function is equal to one.

3.4.3 Distribution of distance measures

To determine if a difference is stochastically significant, it is essential to perform hypothesis testing [19, 20, 40]. In order to conduct a hypothesis test to determine whether the samples are different, it is essential to know the distribution of the distance measure. In this section, the distributions of the distance measures are determined. These distributions are well documented, therefore the reader is referred to literature for additional information about the distributions e.g. [19, 20, 40].

The mean value distance between two univariate samples is normally distributed for large samples according to the central limit theory [20, 40, 57, 58]. However, if the sample contains only a few data points (less then 30 data points) the distribution of the mean value difference is not normally distributed, due to significant fluctuations of the variance from sample to sample. If the populations are normally distributed then the distribution of the mean value difference between two samples is Student-t distributed. Even if small samples are taken from a non-normal distribution with a bell shape, the difference between mean values of samples is still approximately

t-distributed. [20] Hence, the mean value difference calculated with equation 3.29 is Student-t distributed with n_1+n_2-2 degrees of freedom if the mean value difference is normalised using the sample size as presented in equation 3.39 [18, 19, 20, 40],

$$t = \frac{|\bar{x}_2 - \bar{x}_1|}{s_{\text{pl}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \tag{3.39}$$

where s_{pl}, n_1, n_2, $\bar{x_1}$ and $\bar{x_2}$ are respectively the pooled variance of the sample, the size of sample one and two and the mean value of sample one and two. The pooled variance is calculated with equation 3.30, which is based on the assumption that the samples have an equal variance. To verify if this assumption is true, it is essential to determine whether a difference between the sample variances is present. This difference is calculated with equation 3.31 and it is F_{ν_1,ν_2} distributed [18, 20, 40]. The variables ν_1 and ν_2 are respectively the degrees of freedom of sample one and two.

For multivariate samples, the mean value difference between two samples is determined using equation 3.33, which varies due to the sample to sample variation. The Mahalanobis distance is normalised using the sample size as expressed in equation 3.40 [18],

$$T^2 = \frac{n_1 n_2}{n_1+n_2} {D_{\text{M}}}^2 \tag{3.40}$$

where n_1, n_2 and ${D_{\text{M}}}^2$ are respectively the sample size of sample one, the sample size of sample two and the Mahalanobis distance. The variable T^2 is Hotelling-T^2 distributed with dimensionality d and n_1+n_2-2 degrees of freedom if the samples are normally distributed [18]. Since the Hotelling-T^2 distribution is related to the F-distribution, it is common practice to express the distribution of the Mahalanobis distance using equation 3.41 [18],

$$\begin{gathered} F_{d,v-d+1} = \frac{v-d+1}{vd} T^2_{d,v} \\ v = n_1+n_2-2 \end{gathered} \tag{3.41}$$

where d and v are the dimensionality of the samples and the degrees of freedom. [18] To verify the assumption that the covariance matrices are equal, which is used during definition of the pooled covariance matrix, it is essential to measure the distance between the covariance matrices. Since every sample drawn from one population is slightly different, also the variances vary from sample to sample. The distribution of the difference covariance matrices is approximately Chi square distributed with $v = \frac{1}{2}d(d+1)$ degrees of freedom if the difference is calculated with equation 3.42 [18],

$$m = -2\left(1 - \left(\left(\frac{1}{n_1-1}+\frac{1}{n_2-1}-\frac{1}{n_1+n_2-2}\right)\left(\frac{2d^2+3d-1}{6(d+1)}\right)\right)\right)\ln(M) \tag{3.42}$$

where d, n_1 and n_2 are the dimensionality of the samples, the sample sizes of sample one and two respectively. The variable m is the Box-M statistic and the variable M is the distance between the covariance matrices, which is calculated with equation 3.35.

In practice, numerical and experimental results are often not normally distributed. If the samples are non-normally distributed, the distribution of the mean value and covariance matrix difference are unknown. In order to use the distance measures for normally distributed samples, it is required to transform the results to normally distributed results. Some of these samples are transformable

to a normal distribution because they are analytically related to the gaussian distribution (e.g. the lognormal distribution) while others might be transformable to an approximately normally distributed sample using a power transformation. An effective power transformation is the Yeo-Johnson transformation [59, 60] which is based on the Box-Cox transformation [61]. The Yeo-Johnson transformation is presented in equation 3.43 [59, 60],

$$W(x,\lambda) = \begin{cases} \frac{(x+1)^{\lambda}-1}{\lambda} & \text{if } x \geq 0, \lambda \neq 0 \\ \log(x+1) & \text{if } x \geq 0, \lambda = 0 \\ \frac{-((-x+1)^{2-\lambda}-1)}{2-\lambda} & \text{if } x < 0, \lambda \neq 2 \\ -\log(-x+1) & \text{if } x < 0, \lambda = 2 \end{cases} \tag{3.43}$$

where x and W are the non-normally distributed sample and the transformed sample. Furthermore, the variable λ is the coefficient of the power transformation. Its value is commonly between -3 and +3 [59, 60, 61]. To determine if the transformation is successful, the goodness-of-fit is determined between the transformed sample and the gaussian distribution. The transformation is considered unsuccessful if either a significant difference is measured or the value of λ is outside the range $[-3;3]$.
If the samples are arbitrarily distributed, it is not possible to determine the distribution of the distance measures. Even the distribution of the K-S statistic and the A-D test statistic are unknown when the samples are univariate and arbitrarily distributed. Consequently, it is not possible to perform a parametric hypothesis test to determine if the distance is significant.

3.4.4 Tests of hypotheses

Statistical hypothesis test are used to make a decision in acceptance or rejection of some statement [20, 40]. The following definition is used for the term statistical hypothesis: A statistical hypothesis is an assertion or conjecture concerning one or more populations [20, 40]. The truth of a statistical hypothesis is never known with absolute certainty unless we examine the entire population, which is impractical in most situations.
Therefore, it is common practice to use a sample to decide whether the null hypothesis is likely to be true or false. Evidence from the sample which is inconsistent with the stated hypothesis leads to rejection of the hypothesis. Evidence supporting the hypothesis leads to acceptance of the hypothesis. An accepted hypothesis is a result of insufficient evidence to reject the hypothesis and does not necessarily imply it is true. The hypothesis we wish to test is called the null hypothesis, which is denoted by H_0. The rejection of H_0 leads to acceptance of an alternative hypothesis, which is denoted by H_1. [20, 40]
To decide whether the null hypothesis is accepted or rejected, the distance between the samples is compared to the critical distance (critical value of the test statistic), which is based on the significance level α of the hypothesis test. Distances larger than the critical distance will be in favour of H_1 while values less or equal to the critical value will be in favour of H_0 [19, 20, 40]. This procedure leads to two possible erroneous results, a type I error and a type II error. Rejection of the null hypothesis when it is true is called a type I error while acceptance of the null hypothesis when it is not true is called a type II error as presented in table 3.2 [20, 40]. The absolute value of the type I error is defined by the significance level of the hypothesis test and it is directly related to the critical value of the hypothesis test. The type II error on the other hand is never known unless we specify a specific alternative hypothesis and know the sample size. [20, 40]
To explain hypothesis tests in more detail, an example is presented below. In this example a mean

Table 3.2: Possible results of a statistical hypothesis test. [20].

	H_0 is true	H_0 is false
Accept H_0	Correct decision	Type 2 error
Reject H_0	Type 1 error	Correct decision

value hypothesis test is performed using univariate normally distributed samples. The two used populations are both normally distributed with the following parameters N(5.2,0.2). The difference between samples taken from these identical distributed populations is calculated with equation 3.39. This difference is Student-t distributed. To determine whether two random samples of 20 data points drawn from these populations are identical, a mean value hypothesis test is performed using the null hypothesis $\mu_1 = \mu_2$ and the alternative hypothesis $\mu_1 \neq \mu_2$. For this hypothesis test an type I error of 5% is assumed which results in the two critical values as shown in figure 3.8a. If the distance is larger than absolute critical value, the hypothesis will be falsely rejected (type I error).

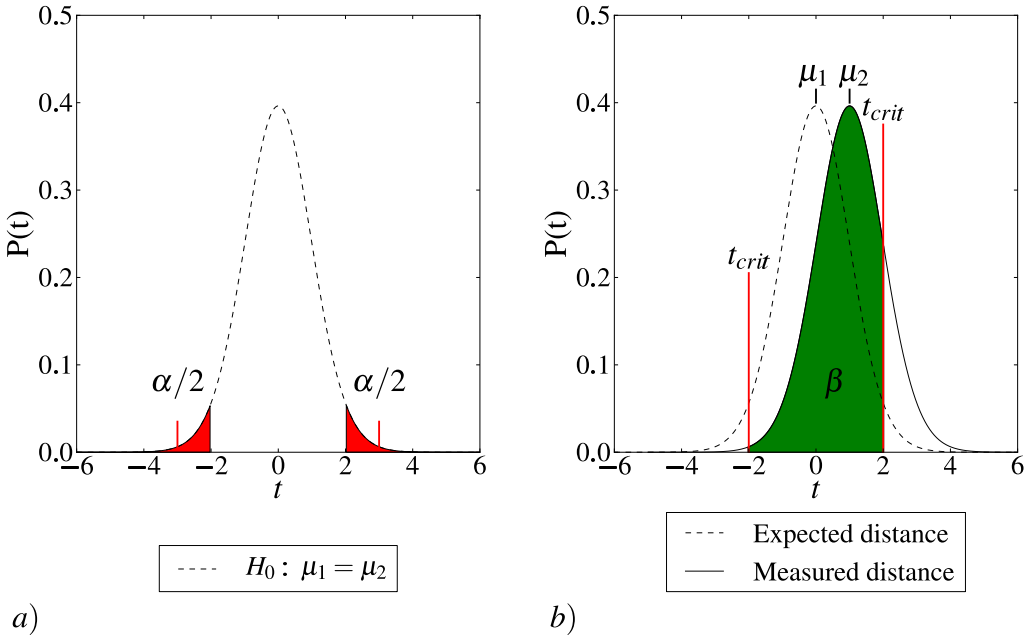


Figure 3.8: Error types of a hypothesis test based on a Student-t distributed distance using samples containing 20 values: a) Type I error, b) Type II error.

To calculate the type II error, it is necessary to select a specific alternative hypothesis. In this case, the alternative hypothesis $\mu_1 + 1.0 = \mu_2$ is selected where the constant is the distance calculated with equation 3.39. Based on this alternative hypothesis, the probability is calculated to obtain an absolute mean value difference less than the critical value if the populations were 1.0 separated using equation 3.44 [20],

$$\beta = \int_{t_{-\mathrm{crit}}}^{t_{+\mathrm{crit}}} P\left(t_{\mathrm{H_1}}(x)\right) \mathrm{d}t \tag{3.44}$$

where t_{crit} and $t_{\mathrm{H_1}}$ are respectively the critical value of the null hypothesis and the distribution of the distance if the populations were 1.0 separated. In this example, the distribution of distance

is non-central Student-t distributed because the populations are 1.0 separated. The variable β is called the type II error and is visually presented in figure 3.8b. To determine the probability to reject the null hypothesis, the discriminatory power is calculated, which is defined as $1-\beta$. Since the value of the type II error β is depending on the random distance between the samples, it is essential to determine the value of β for all possible differences, which can be plotted versus the sample mean value distance D_m. This diagram is known as the operation characteristic [62, 63] and is presented in figure 3.9.

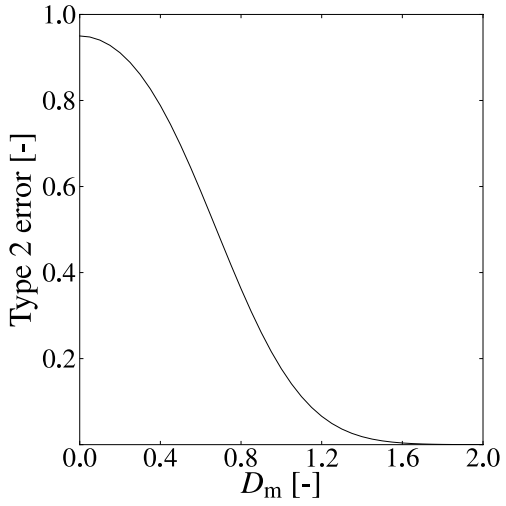


Figure 3.9: Operation characteristic of a mean value difference which is normalised using the standard deviation.

Similar to the hypothesis test presented in the example above, it is possible to create hypothesis tests for the other distance measures presented in subsection 3.4.2 if the distribution of the distance measure is known [18, 19, 20, 40, 64]. When the distribution of the distance measure is not known, numerically intensive hypothesis test methods can be used like bootstrapping [65] and randomisation based hypothesis tests [65]. For this work it is chosen to use the randomisation method which is explained in the next subsection.

Randomisation hypothesis test

Randomisation hypothesis tests are used to determine if a distance is significant when the distribution of the distance measure is unknown. To test whether a difference exists or not, so called pseudo samples are used for the randomisation hypothesis method to test the null hypothesis. These pseudo samples are generated by switching data points from one sample with data points from the other sample. To illustrate the principle of pseudo sample generation, all possible pseudo samples of the original samples A and B (see table 3.3) are presented in the second column of table 3.4 together with the original samples. The total number of possible pseudo samples PS is calculated with equation 3.45 [65],

$$PS = \frac{(n_1+n_2)!}{n_1!n_2!} \tag{3.45}$$

where n_1, n_2 are respectively the sample size of sample one and two. After generation of all possible pseudo samples, the difference between the pseudo samples is calculated. Subsequently, the number of pseudo samples is calculated for which the difference is greater than or equal to

the original difference. This number of pseudo samples is denoted as *nge* (Number Greater than or Equal to). The number of pseudo samples for which the difference is larger than or equal to the original difference is used to calculate the probability ϕ_0, which is the probability that the difference between the pseudo samples is lager than or equal to the difference between the original sample using equation 3.46 [65],

$$\phi_0 = \frac{nge+1}{PS+1} \tag{3.46}$$

where *nge* is the number of pseudo differences larger than or equal to the original difference and *PS* is the total number of possible pseudo samples. The null hypothesis is rejected if the p-value ϕ_0 is smaller than the selected significance level α of the hypothesis test [65].
To demonstrate the calculation of ϕ_0, an example is presented below for which two samples are taken from a normal distribution N(3.0,0.1). These samples are presented in table 3.3 and they are extraordinarily small to limit the amount of possible pseudo samples. The mean value difference

Table 3.3: Random samples from a normal distribution N(3,0.1).

Sample number	Value
A1	2.984
A2	2.780
A3	2.999
B1	3.053
B2	2.903

between the original samples and the pseudo samples are presented in table 3.4 together with the corresponding samples. In the first row of table 3.4, the original distance between the samples is presented. The probability for a distance at least as large as the original distance for this example is $7/10$, because 6 distances between pseudo samples are larger than the original distance as shown in table 3.4. Using a user defined significance level of $\alpha = 0.1$, it is concluded that the hypothesis is not rejected, because 0.7 is larger than the significance value. Since this sample has only a few variables and data points, all combinations can be calculated and used to determine the probability of the hypothesis test.

Table 3.4: Distance of all possible randomisations of the random samples presented in table 3.3.

Randomisation number	Data points nr. in sample A - B	Distance
1	A1,A2,A3 - B1,B2	0.057
2	A1,A2,B1 - A3,B2	0.012
3	A1,A2,B2 - A3,B1	0.137
4	A1,A3,B1 - A2,B2	0.170
5	A1,A3,B2 - A2,B1	0.045
6	A1,B1,B2 - A2,A3	0.090
7	A2,A3,B1 - A1,B2	0.001
8	A2,A3,B2 - A1,B1	0.124
9	A2,B1,B2 - A1,A3	0.079
10	A3,B1,B2 - A1,A2	0.103

For engineering problems which have more variables or data points, the number of possible combinations becomes too large to include all permutations. Instead of including all possible permutations, the Monte Carlo method can be used to generate a limited amount of pseudo samples NS. [65, 66, 24] The drawback is the possibility of an extra uncertainty (a false decision) in the probability to reject the hypothesis, which can be calculated using equation 3.47 [65],

$$\mathrm{P}\left(\frac{(nge+1)}{(NS+1)} \leq \alpha\right) \leq \alpha \tag{3.47}$$

where NS is the number of pseudo samples. If the data points in the original samples are mutually exclusive and $\alpha\,(NS+1)$ is an integer, it is shown with equation 3.47 that the Monte Carlo test is exactly valid. This implies that a Monte Carlo test is valid, because the probability to rejected the null hypothesis is never too large when the p-value ϕ is calculated with equation 3.48 [65].

$$\phi = \frac{nge+1}{NS+1} \tag{3.48}$$

Although the test is valid for every value of NS if $\alpha \cdot (NS+1)$ is an integer, the power of the test increases with increasing value of NS. This leads to the limit of the ratio, which is denoted as ϕ_l. The limit is calculated with equation 3.49 [65].

$$\phi_l = \lim_{NS\to\infty} \frac{nge+1}{NS+1} \tag{3.49}$$

The null hypothesis should be rejected if $\phi_l \leq \alpha$, but ϕ_l is unknown. It can only be estimated by using a finite number of pseudo samples by means of ϕ. Fortunately, the probability that $\phi \leq \alpha$ can be calculated for a given nge and NS using equation 3.50 [65],

$$\mathrm{P}(\phi \leq \alpha | nge, NS) = \frac{\sum_{\phi=0}^{\phi=\alpha} \mathrm{P}(nge|\phi, NS)\,\mathrm{P}(\phi)}{\sum_{\phi=0}^{\phi=1} \mathrm{P}(nge|\phi, NS)\,\mathrm{P}(\phi)} \tag{3.50}$$

When assuming ϕ is uniformly distributed, which is true if the null hypothesis is true, equation 3.50 simplifies to equation 3.51 [65],

$$\mathrm{P}(\phi \leq \alpha | nge, NS, \text{uniform prior}) = \frac{\int_{\phi=0}^{\phi=\alpha} \binom{NS}{nge} \phi^{nge} (1-\phi)^{NS-nge}\, d\phi}{\int_{\phi=0}^{\phi=1} \binom{NS}{nge} \phi^{nge} (1-\phi)^{NS-nge}\, d\phi} \tag{3.51}$$

where the expression on he right-hand side is the cumulative incomplete beta function with the parameters nge and $NS - nge + 1$ [65].
To determine which hypothesis test is more powerful, it is again required to determine the probability to reject the hypothesis. This probability is calculated with equation 3.52 [65].

$$\mathrm{P}(\text{reject}|\alpha, NS, \phi) = \sum_{nge=0}^{\alpha(NS+1)-1} \binom{NS}{nge} \phi^{nge} (1-\phi)^{NS-nge} \tag{3.52}$$

Since the probability is depending on ϕ, it is only valid for the used samples and pseudo samples. This is not what we would like to know when comparing hypothesis tests. We would like to know the probability to reject the hypothesis among all samples that could be drawn from the populations. To determine the probability to reject the null hypothesis based on all possible samples of the distributions, the conditional ϕ must be eliminated by including the probability density function of ϕ $f(\phi)$, which results in equation 3.53 [65].

$$\mathrm{P}(\text{reject}|\alpha,NS) = \sum_{nge=0}^{\alpha(NS+1)-1} \left(\begin{array}{c} NS \\ nge \end{array} \right) \int_0^1 \phi^{nge} (1-\phi)^{NS-nge} f(\phi)\, \mathrm{d}\phi \tag{3.53}$$

3.4.5 Optimisation

As explained in chapter 2, during validation it is determined whether a numerical model is valid. However, an engineer or scientist is interested in a valid model after execution of the validation process. Consequently, if the model is declared "invalid", it is essential to decrease the distance between the experimental results and the numerical results to obtain a valid numerical model by changing the unverified parameters (assumptions). To find the settings of the unverified parameters in an automatic manner, an optimisation method is used to minimise the distance. Since the function of the distance between the numerical and experimental results is not differentiable with respect to the unverified parameters, only zero order methods and response surface methods can be used. Where the zero order methods work directly with the distances, the response surface method first approximates the function of the distance around the original distance. The main advantage of the response surface method is that a gradient based optimiser can be used at the response surface making it more effective than a standard zero order method. For this work the adaptive response surface methodology of the software OptiSLang 3.2.3 is used to minimise the difference between the numerical and experimental results [67].
The first step of the adaptive response surface method is defining a subregion around the original distance within the limits of the parameters of the numerical model. This sub-domain is probed with Design of Experiments to generate support points. Based on these support points, a linear or quadratic approximation of the response function is created which can be varied locally with the moving least square routine. The adaptive response surface method starts with a relative large sub-domain of the bounded parameter space that continuously decreases till the solution has converged to an optimum. The starting sub-domain is calculated with equation 3.54, [67]

$$\mathbf{x}_0^l = \mathbf{x}_0 - \gamma_{\text{start}}\mathbf{r}_0, \qquad \mathbf{x}_0^u = \mathbf{x_0} + \gamma_{\text{start}}\mathbf{r}_0 \tag{3.54}$$

where $\mathbf{x}_0$, $\mathbf{r}_0$and γ_{start} are the start point of the response surface, the size of the global parameter domain and the ratio of the sub-domain based to the global parameter domain respectively. The default response surface domain is equal to 50% of the global parameter domain if the response surface remains withing the global parameter boundaries else the parameter boundary is used as response surface boundary. Subsequently, the support points of the response surface are defined using the Design of Experiment based on the D-optimal method. At every support point an generic optimisation is performed to obtain best possible approximation function for the response surface. Subsequently, the minimum of the approximation function is determined. Next, a new sub-domain is generated based on the range of the previous sub-domain with its centre at the optimum of the previous iteration. New sub-domains are generated till one of the termination criteria is fulfilled. To prevent oscillation between two sub-domains, a oscillation reduction factor is introduced which slows down the zooming of the sub-domain. [67]

4 Development of a validation method for arbitrary multivariate problems

The currently available validation methods should not be used to analyse non-normally distributed multivariate samples, because they assume that the samples are normally distributed which could lead to incorrectly validated models. To validate numerical models with non-normally distributed multivariate results, a new validation method has to be developed. In this chapter the development of a distribution independent validation method is presented. This method is suitable to validate n-dimensional numerical models with arbitrarily distributed results using experimental results. To successfully validate a set of arbitrary results, the validation method has to meet the following requirements:

1. Differences should be measured using a generally applicable distance measure.
2. Decisions should be based on a generally applicable decision criterion.
3. Calculation procedures should be numerically efficient.

To develop a method to validate a numerical model with arbitrarily distributed results, it is essential to follow the steps presented in figure 4.1. Before determining the distance between multivariate numerical and experimental results, it is required to know, estimate or assume the underlying distributions. For the currently used methods, a multivariate normal distribution is assumed. However, the underlying distributions of the numerical and experimental results are seldom multivariate normal and they are in general not known. Consequently, it is required to estimate the underlying

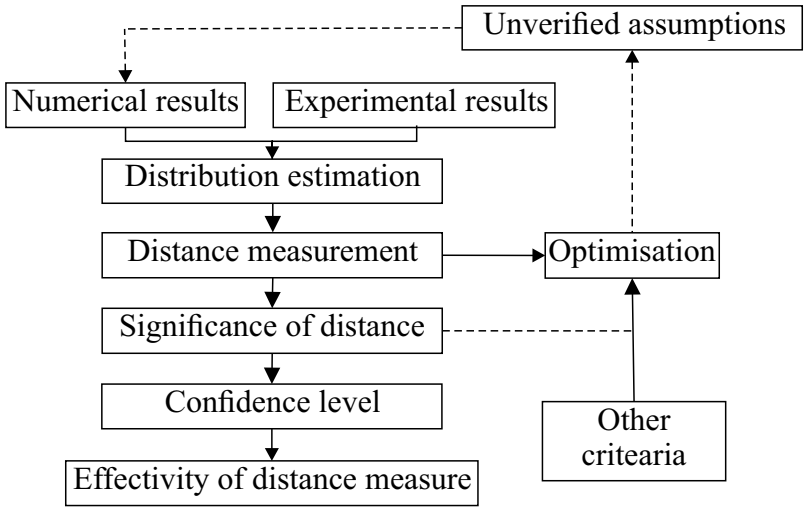


Figure 4.1: Flowdiagram of validation for arbitrary multivariate problems.

distribution of the numerical and experimental results using a distribution estimation method. The distribution estimation methods relevant for this work have been presented in section 3.4.1. Further details concerning the settings of the used methods to estimate the underlying distributions are presented in section 4.1.
Next, the distance between the estimated distribution functions has to be determined. To determine the distance between two distributions, both the cumulative distribution function and the probability density function can be used. Depending on the information to emphasise, the one or the other is in favour. Properties such as heaviness of the tails and modes are easier recognised in a probability density function, because they are directly represented by the densities. Since these properties are of interest for validation, a validation analysis based on the probability density function is used. The difference between two estimated probability density functions is calculated by solving equation 2.3 numerically, because a closed form solution does not exist for this equation if the samples are multivariate and arbitrarily distributed. To solve this integral numerically, the integration domain must be limited and the estimated probability density function must be discretised. Such a multivariate discrete probability density function is mathematically identical to a N-dimensional pixel based gray-scale image. Therefore, it is also possible to express the difference between two estimated probability density functions using image correlation. Such an image correlation method is used to measure the distance for the developed validation method, which leads to the name Image Based Validation Method (IBVM). Further details about the used distance measure as well as the domain and the resolution of the images are presented in section 4.2.
Subsequently, it is determined whether the measured distance between simulation and experiment is stochastically significant by means of a randomisation hypothesis test. Since the sample might be an extraordinary one, it is insufficient to decide if a difference is present based on only the significance level of the hypothesis test. It is required to determine the confidence level of a decision based on the hypothesis test. The details of the hypothesis test and its confidence level are presented in section 4.3. In addition, it is determined whether IBVM is more effective than the currently used methods by comparing the power of the methods with another. Details concerning the effectivity of a validation method are presented in section 4.4.
When one decides on basis of the validation criteria that the measured difference is too large and there are one or more unverified model assumptions, it might be possible to minimise the difference between simulation and experiment. Minimising the difference between simulation and experiment improves the validity of the model provided that the experimental results are not biased and the optimisation parameters remain in their physical restricted domain. However, care must be taken with optimisation, because neither the experiments nor the simulations are exact. One can only gain confidence in the experimental results by performing measurement uncertainty analysis, which are presented in section 3.1. If these uncertainty analyses show that the experimental results are measured sufficiently accurate, minimisation by changing the unverified parameters within a meaningful range could lead to a more accurate model. In this work, minimisation of the distance measured between simulation and experiment is performed with the Adaptive Response Surface method of the optimisation software OptiSLang, which is described in subsection 3.4.5. The problem specific settings of the optimisation routine are presented in the specific example, see chapter 6.

4.1 Distribution estimation of the numerical and experimental results

The underlying probability density function of a sample can be estimated with different methods as presented in section 3.4.1. For validation it is required that the estimated probability density function is independent of the chosen origin of the diagram used to present the estimated probability density function. This is only true for the kernel based estimation methods and the balloon estimator, which is a special case of the kernel estimator. For this work, the gaussian kernel estimator is used, because it is an effective kernel estimator [15]. The width of the Gaussian kernel can be defined using two different approaches. The first approach defines the kernel width based on a theoretically derived optimal width while the second method defines the kernel width based on the measurement uncertainties.
Using the first approach, the kernel width is chosen autonomously. The optimal width of the Gaussian kernel density estimator is calculated using equation 3.28, which is substituted in equation 3.26 to calculate the estimated probability density function. The second method defines the kernel based on the measurement uncertainty. Since measurement uncertainties are normally distributed according to the central limit theory if many sources contribute to the measurement uncertainty [20, 40, 57, 58], the kernel width is defined using the confidence intervals of the measurement uncertainty. Furthermore, it is meaningful to define a minimum number of data points per sample when using the measurement uncertainty to define the Gaussian kernel estimator to obtain an accurate estimated probability density function. The minimum number of data points required to estimate the probability density function is calculated by solving equation 3.28 for n.
To illustrate the influence of the kernel width on the estimated probability density function, the probability density function of a bimodal distribution is estimated using the gaussian kernel estimator for different sample sizes. The distribution parameters of the bimodal distributions are presented in table 4.1. From this distribution a sample is drawn, which contains 7, 20 or 1000 data points. These data points are listed in table 4.2 for the sample containing 7 and 20 data points. To estimate the underlying distribution function, a Gaussian kernel is positioned at every data point as shown in figure 4.2. As presented in figure 4.2a, the kernel N(0,0.8) results in a relatively wide estimated probability density function with little detail. Consequently, the estimated probability

Table 4.1: Parameters of the bimodal distributions which are used to demonstrate the influence of number of data points per sample and the influence of the kernel width on the estimated probability density function.

Distribution type	Mean value	standard deviation	weight [%]
Normal	3.65	0.55	0.75
Normal	2.45	0.45	0.25

Table 4.2: Data points of the samples drawn from the distribution specified in table 4.1 which contain 7 or 20 data points.

	i	x_{i0}	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	x_{i7}	x_{i8}	x_{i9}
Sample 1	0	2.4	2.5	3.4	3.5	3.52	3.8	4.1			
Sample 2	0	2.4	2.5	3.4	3.5	3.52	3.8	4.1	3.73	1.81	4.61
	1	4.02	3.76	3.53	4.24	2.26	3.17	3.25	3.06	2.81	2.7

density function is an inaccurate representation of the underlying bimodal distribution. In contrast to the wide kernel, the small kernel N(0,0.1) does not necessarily create a more accurate estimated probability density function of the underlying distribution as shown in figure 4.2b. The lack of sufficient measurements results in a spiked probability density estimate. To improve the estimated density function, more data points have to be measured as presented in figure 4.3a. Fortunately, one does not have to iterate till the estimated probability density function looks good. Solving equation 3.28 for n results in the minimum amount of data required to estimate the probability

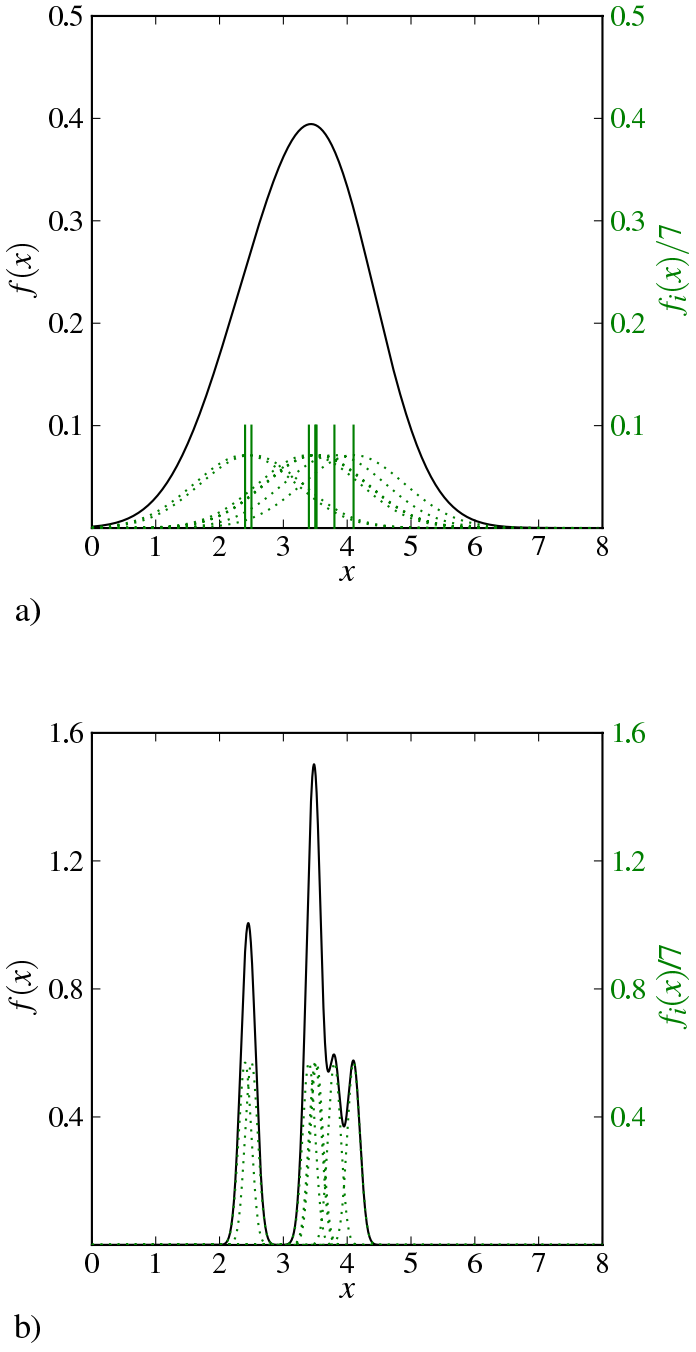


Figure 4.2: Influence of the width of the Gaussian kernel on the estimated probability density function: a) Distribution density estimate based on 7 data points using the kernel N(0,0.8), b) Distribution density estimate based on 7 data points using the kernel N(0,0.1)

density function accurately based on the mean value measurement uncertainty. This equation is generally applicable if the data are not extremely skewed or multimodal as discussed in section 3.4.1. According to equation 3.28, approximately 1000 data points are required to estimate the probability density function using the kernel N(0,0.1). The estimated probability density function based on 1000 data points is presented in figure 4.3b.

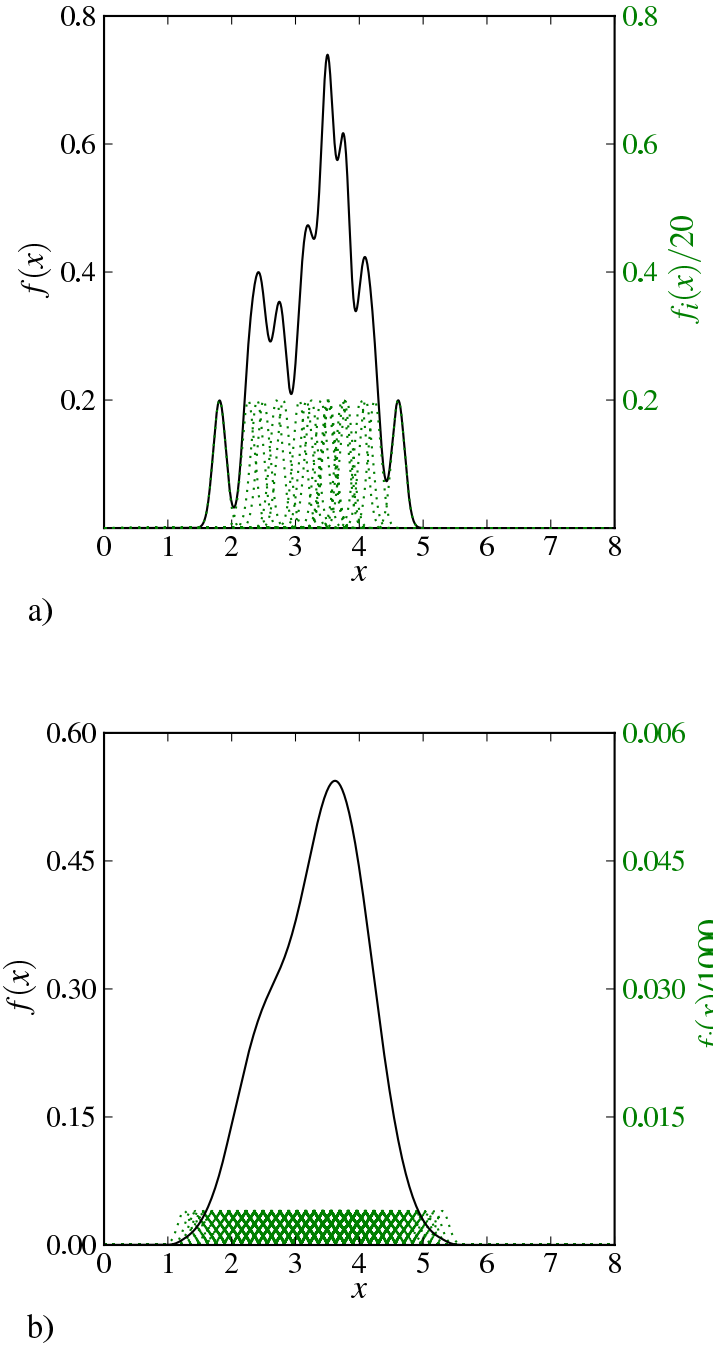


Figure 4.3: Influence of the number of data points on the estimated probability density function: a) Distribution density estimate based on 20 data points using the kernel N(0,0.1), b) Optimal distribution density estimate based on 1000 data points using the kernel N(0,0.1).

In conclusion, validation becomes valuable if the experiments and simulations are performed accurately and repeated many times, because only then the distributions are estimated accurately. If no additional measurements can be performed to avoid spiked estimated probability density functions, it is preferred to estimate the distribution using the autonomously defined gaussian kernel width presented at the beginning of this section.

4.2 Distance measurement

The distance between the numerical and experimental results is measured using an image based distance measurement technique, which lead to the name Image Based Distance Measure (IBDM). This distance measure uses the overlap coefficient according to Manders ρ_m (see equation 3.38) to determine the distance between two discrete estimated probability functions, because it is insensitive to the size of the domain of discretisation as shown in appendix A. Although image correlation coefficients can be used to determine the difference between multidimensional discrete probability distributions, the computational recourses required to process the probability distributions increase rapidly with the dimensionality. Therefore, it is preferred to use 2-dimensional projections of n-dimensional estimated probability density function to limit the required computational resources. This requires $d\,(d-1)/2$ projections on a 2-dimensional space to fully represent the original n-dimensional distribution, where d is the dimensionality of the original distribution. Each projection results in one distance measure, which can be multiplied with each other, summed or averaged to obtain the difference between the n-dimensional estimated probability density functions. In this work, the averaging method is used.
To generate an image of an estimated probability density function, it is required to define the domain of interest of the probability density function and the number of discretisation steps. This domain must be defined such that the integral of the probability density function is approximately one. Since the gaussian kernel has a low density at the tails of the distribution, the domain is defined as $[\mathbf{x}_{\text{min}} - 3\sigma_{\text{kernel}},\, \mathbf{x}_{\text{max}} + 3\sigma_{\text{kernel}}]$. Other domain definitions that result in a larger domain are permitted.
Concerning the number of pixels required to ensure that the discretised estimated probability density function is an accurate representation of the estimated probability density function, it can be said the more the better. Increasing the amount of pixels will increase the resemblance between the discretised probability function and the continuous estimated probability density function, but it will increase the demand on the computational resources, because more discrete probability values have to be processed during the difference calculation. To reduce the discretisation error to an insignificant influence without consuming too much computational resources, it is mandatory to determine the number of pixels required to measure the difference between two estimated probability density functions. The necessary amount of pixels depends on two factors: the maximum slope of the estimated probability density function (continuous function) and the euclidean distance between the two samples. The larger the slope of the estimated probability density function, the more pixels are required to represent the continuous probability density function accurately as illustrated in figure 4.4a and figure 4.4b. To obtain the same discretisation error in both figures, it is necessary to increase the amount of intervals in figure 4.4b by a factor two. Moreover, the further the samples are separated, the larger the domain of the discrete probability function is. To represent a large domain with the same accuracy as a small domain, it is necessary to use more pixels for the large domain. The ratio between the two domains gives the required increase of pixels to maintain equal accuracy. The domain of the estimated probability density functions is

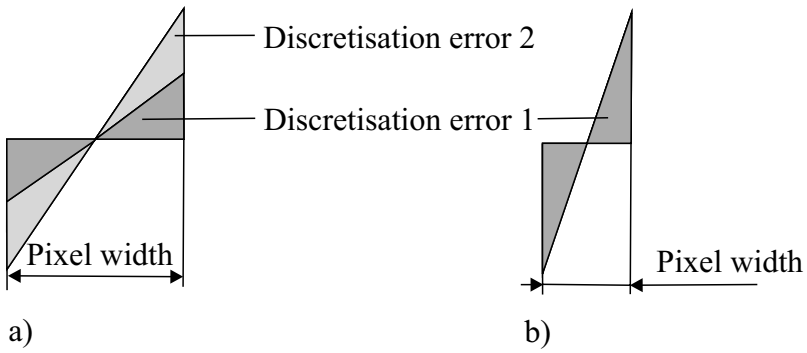


Figure 4.4: Influence of the pixel width on difference between the discrete probability density function and the estimated probability density function: a) Discretisation of two slopes using equal pixel width, b) Discretisation of the steeper slope of sub-image a) using more pixels.

defined using the kernel width. Consequently, the required amount of pixels also depends on the kernel density estimator.
To determine the required amount of pixels to create an accurate discretisation of the estimated probability density function, a convergence study is performed using an example. For this example the difference between two estimated probability densities is measured using the overlap coefficient according to Manders (see equation 3.38). Both samples contain 45 data points and are drawn form a bivariate normal population. The parameters of the distribution are presented in table 4.3 and the samples drawn from this population are presented in appendix C. To estimate the probability density function, the width of the Gaussian kernel estimator is varied from 0.1 to 1.0 times the standard deviation of the population. This kernel range covers all optimal gaussian kernels of practical importance ($1 < N < 1000000$ according to equation 3.28). Since the samples only contain 45 data points, a spiked estimated probability density function is created when a small kernel is used. This spiked probability density function results in a measured distance which is too large. Although a spiked estimated probability density function is undesired for validation, it is perfect to investigate the convergence of the distance measure for different number of pixels due to the steep slope of the spikes.
For this example the domain of the samples is defined as $[\mathbf{x}_{\min} - 5\sigma_{\text{kernel}}, \mathbf{x}_{\max} + 5\sigma_{\text{kernel}}]$, where the maximum and minimum values of the data points are obtained from appendix C. This results in the following domain of the descretised estimated probability density function: ([3.25, 6.91], [6.11, 9.81]). This domain is filled with pixels, where the pixel intensity is calculated by integration of the estimated probability density function over the domain of the pixel. To determine when the discretisation error is insignificant for this example, the overlap coefficient according to Manders ρ_m is plotted versus kernel width and number of pixels as shown in figure 4.5a.

Table 4.3: Parameters of the bivariate normal distribution used to demonstrate the influence of discretisation on distance measurements.

Variable nr.	Mean value	Standard deviation
1	5.0	0.25
2	8.0	0.25

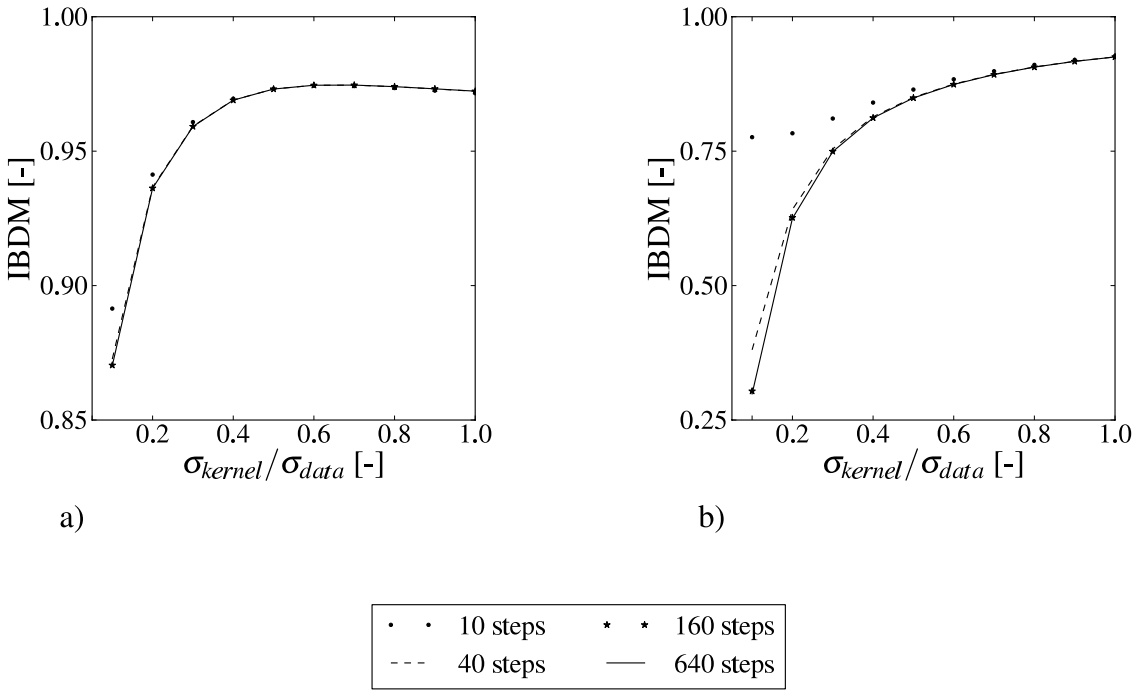


Figure 4.5: Measured distance between two samples with the Image based Distance Measure as function of the kernel width and the number of discretisation steps: a) Bivariate normally distributed samples, b) Bivariate lognormally distributed samples.

It can be seen in figure 4.5a that the correlation coefficient converges with increasing kernel width. More precisely, the distance between the samples approaches its minimum value when the kernel width is half of the standard deviation of the data. This value is equal to the optimal kernel width, which is calculated using equation 3.28. Furthermore, figure 4.5a shows that as few as 40 discretisation steps (pixels) per direction could be just enough to represent the estimated normal distribution sufficiently accurate. Certainly 160 discretisation steps are sufficient to accurately represent this estimated probability density function, because there is no significant difference between the curve based on 160 discretisation steps and the one based on 640 discretisation steps.
To investigate the influence of a skewed probability density function on number of discretisation steps required to accurately represent the estimated probability density function, the convergence study is repeated using a lognormal distribution. For this example the samples are drawn form a bivariate lognormally distributed population which is defined using the parameters shown in table 4.4. The domain of the discretised estimated probability density function is defined as $[\mathbf{x}_{\text{min}} - 5\sigma_{\text{kernel}},\ \mathbf{x}_{\text{max}} + 5\sigma_{\text{kernel}}]$, where the maximum and minimum values of the data points are obtained from appendix C. This results in the following domain of the discretised estimated

Table 4.4: Parameters of the bivariate lognormal distribution used to demonstrate the influence of discretisation on distance measurements.

Variable nr.	Mean value	Standard deviation
1	5.0	1.25
2	8.0	2.0

probability density function: ([1.72, 10.59], [2.91, 14.84]). Based on figure 4.5b, it is concluded that 160 discretisation steps in each direction are sufficient to represent the estimated probability density function accurately, because all curves generated using more discretisation steps are identical to the curve based on 160 discretisation steps. Furthermore, it can be seen that the distance between the two estimated probability density functions does not converge to a constant value, but approaches a linearly increasing function. This effect is caused by the shape difference between the lognormal distribution and the symmetrical kernel. Although the correlation coefficient does not converge to a constant value, the optimal value according to equation 3.28 is still a good compromise between maintaining the asymmetrical character of the population in the estimated probability density function and the spikiness of the estimated probability density function.
Based on the results of both examples, it is concluded that 160 discretisation steps in each direction are sufficient to discretise the estimated probability density function. However, this is only true if the discretisation domain is not significantly larger than the domain of each sample (simulation and experiment) on its own. If a mean value difference is present between the experimental and numerical results, the number of discretisation steps (pixels) required to accurately represent the estimated probability density functions will increase. Since the maximum distance is measured between two estimated probability density functions if they are barely overlapping, the required amount of pixels is increased with a factor 2 which results in 320 discretisation steps in each direction. Moreover, data points could be present further in the tails than the data points in these two examples. Therefore, it is chosen to enlarge the domain by defining the maximum and minimum values of the data points at six standard deviations from the mean value of the distribution. To maintain the resolution identical to the resolution used in the two examples, it is necessary to use 500 discretisation steps in each direction. It is assumed that 500 discretisation steps per direction are sufficient to accurately represent any estimated probability density function.

4.3 Hypothesis testing

To determine whether the distance measured with the Image Based Distance Measure is stochastically significant, it is necessary to perform a hypothesis test, see subsection 3.4.4 for further general information. Since neither the sample distribution nor the distance distribution of a multivariate arbitrarily distributed problem are known, a distribution independent hypothesis test is required. Such a test method is the approximate randomisation method, which is presented in section 3.4.4. As explained in section 3.4.4, a null hypothesis and an alternative hypothesis have to be defined to test whether the null hypothesis is rejected or not. The null hypothesis is rejected if the distance between simulation and experimental results is larger than the critical distance $D_{crit}(\alpha, N_{\text{exp}}, N_{\text{sim}})$, which depends on the sample sizes $N_{\text{exp}}, N_{\text{sim}}$ and the significance level α. Unfortunately, the critical distance is unknown, because the distribution of IBDM is unknown. Fortunately, it is also possible to decide if a distance is significant on basis of the p-values [65]. Consequently, the null hypothesis is rejected if the p-value is equal or smaller than the value of α. When the p-value is larger than α, it merely means that there is not enough evidence to reject the null hypothesis. [65] These p-values are calculated using equation 3.48, where the value of nge is calculated using NS pseudo samples that are generated by switching data points of the original sample as presented in figure 4.6a. The algorithm used to calculate the value of nge is presented in figure 4.6b. If the distance between the pseudo samples is larger than the distance between the original samples, the value of nge is increased by one. This process is repeated for every pair of the NS pseudo samples. Which data points are switched and how many are switched at once is not relevant for the method

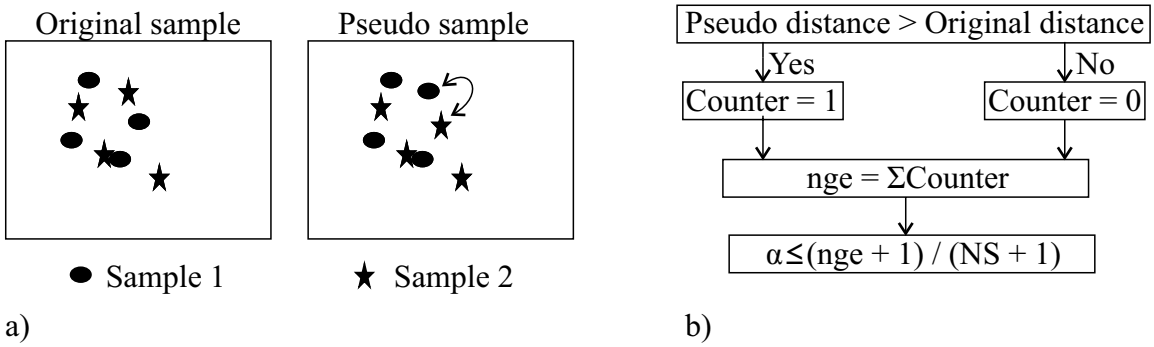


Figure 4.6: a) Principle of the randomisation to create pseudo samples, b) Randomisation hypothesis testing algorithm using NS pseudo samples.

as long as no clone of an already used sample pair is generated. For further details, the reader is referred to section 3.4.4 on page 30.
Since the hypothesis test is performed using one sample pair, it is of interest to know how confident one can be that the correct decision is made. To calculate the confidence level that the null hypothesis is rejected correctly, equation 3.51 is used. When using this equation to calculate the confidence level of a p-value, one assumes that the p-values are uniformly distributed. This assumption holds only if the null hypothesis is true.

4.4 Effectivity of a validation method

To determine whether the new validation method is more effective than the currently used validation methods, it is required to compare the power of the methods with each other. The power of a validation method is best presented using the operation characteristic of the distance measure, which presents the probability to accept the null hypothesis despite an existing difference between simulation and experiment. The validation method having the smallest probability to accept the null hypothesis when a difference is present is the most powerful method among the methods being compared [24]. To generate the operation characteristic for the Image Based Validation Method [62, 63], the probability to accept the null hypothesis is calculated for several support points as illustrated in figure 4.7. At every support point, the probability to accept the null hypothesis

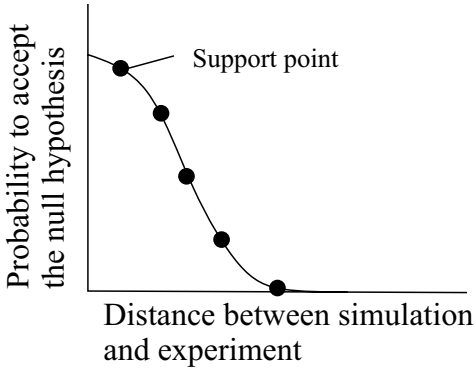


Figure 4.7: Sketch of an operation characteristic and the support points which are calculated using equation 3.53.

is calculated by means of equation 3.53, which is solved numerically using many sample pairs. These sample pairs are used to estimate the distribution of ϕ, which is required to solve equation 3.53. Consequently, the accurateness of the estimated distribution of ϕ depends on the amount of samples used. Further details concerning the comparison of the developed Image Based Validation Method with the currently used validation methods are presented in chapter 5.

5 Effectiveness of the image based validation method

As explained in chapter 4, the developed Image Based Validation Method (IBVM) should be able to detect differences between arbitrarily distributed data. To determine under which circumstances its Image Based Distance Measure (IBDM) is more effective than the currently used distance measures for validation, the operation characteristic of IBDM is compared to the operation characteristic of the currently used distance measures. An operation characteristic represents the probability to accept the null hypothesis although it is false, i.e. the probability to accept the null hypothesis versus the difference between the samples (simulation and experiment). Thus, the lower the probability to accept the null hypothesis when a difference is present, the more effective the validation method is to detect the difference. Since IBDM is developed for arbitrarily distributed results while the currently used methods are limited to normally distributed results, it is necessary to determine the effectiveness of IBDM to detect differences between normally distributed data and non-normally distributed data. Determining the effectiveness of IBDM to detect differences between normally distributed data is important to find out whether the new method performs equally well as the currently used methods in this case. This analysis of the normally distributed data provides the only available reference case. To determine the effectiveness of IBDM for non-normally distributed data, a test case is selected with lognormally distributed data. The lognormal distribution is chosen, because of its practical relevance for engineers [68, 38, 69]. The difference between either two normally distributed samples or two lognormally distributed samples can be expressed as a mean value difference and a covariance matrix difference. Consequently, this chapter is mainly focused on these reference cases, i.e. pure mean value differences and pure covariance matrix differences respectively. For the sake of completeness, also the case of a combined difference between two samples will be briefly discussed.

5.1 Problem description

In case of multivariate data sets, a mean value difference is commonly measured using the Mahalanobis distance while a covariance matrix difference is measured with the Box-M statistic [18]. The Mahalanobis distance is a mean value distance, which is normalised using the covariance matrix. Furthermore, the Mahalanobis distance is only insensitive for a difference between the covariance matrices of the samples when the amount of data points in both samples are equal and large [18]. Therefore, it has to be assumed that the covariance matrices of the underlying populations are equal and that the samples are drawn from Gaussian distributions when calculating the probability to accept the null hypothesis $\boldsymbol{\mu}_{\text{sim}} = \boldsymbol{\mu}_{\text{exp}}$ based on the Mahalanobis distance. To calculate the probability to accept the null hypothesis $\boldsymbol{\sigma}_{\text{sim}} = \boldsymbol{\sigma}_{\text{exp}}$ using the Box-M statistic, it must be assumed that the samples are normally distributed [18].

Since the hypothesis tests based on the Mahalanobis distance and the Box-M statistic have been developed specifically for multivariate normally distributed data, it is expected that these two distance measures are the most effective methods to determine such differences if no assumptions of the hypothesis tests are violated. For cases where the assumptions are violated, measuring the distance using IBDM could be more effective. To investigate whether this is the case, the operation characteristic is calculated for the following cases:

1. Two normally distributed data sets that differ in mean value, but have an identical covariance matrix.
2. Two normally distributed data sets that differ in covariance matrix, but have an identical mean value.
3. Two normally distributed data sets that differ in mean value and covariance matrix.
4. Two lognormally distributed data sets that differ in mean value, but have an identical covariance matrix.
5. Two lognormally distributed data sets that differ in covariance matrix, but have an identical mean value.
6. Two lognormally distributed data sets that differ in mean value and covariance matrix.

To create the operation characteristic for the used methods (Mahalanobis distance, Box-M statistic and IBDM), the probability to accept the null hypothesis has to be calculated for each of these methods using specific null hypotheses for each distance measure. The null hypothesis for the Mahalanobis distance is defined as $\boldsymbol{\mu}_{\text{sim}} = \boldsymbol{\mu}_{\text{exp}}$ and the null hypothesis for the Box-M statistic is defined as $\boldsymbol{\sigma}_{\text{sim}} = \boldsymbol{\sigma}_{\text{exp}}$ while the null hypothesis for IBDM is defined as $f_{\text{exp}}(\boldsymbol{x}) = f_{\text{sim}}(\boldsymbol{x})$. Since the distance measurement methods are developed to detect a specific difference, it is necessary that the difference between the samples is defined such that one can compare the methods with each other (the six cases defined above). These differences are specified for each case separately in sections 5.4 and 5.5. Using the defined difference between simulation and experiment, the probability to accept the null hypothesis is calculated for each case by means of hypothesis tests. For the classical hypothesis tests of the currently used distance measures, it is assumed that the samples are normally distributed, which is incorrect when the samples are drawn from a bivariate lognormal distribution. To avoid violation of the assumptions of the hypothesis tests for non-normally distributed data, randomisation hypothesis tests are used in this work. Another option would be transforming the lognormal distribution to the normal distribution. However, this transformation has some side effects. For example, a mean value test performed on the transformed normal distribution is actually a median value test in the original space [70, 71, 72]. Therefore, only untransformed data sets are analysed in this work. In the next section, the numerical model is presented which is used to calculate the probability to accept the null hypothesis.

5.2 Numerical model

As presented in table 3.2 on page 29, the probability to accept the null hypothesis is dividable into two cases: a case where the null hypothesis is true and a case where the null hypothesis is false. When the null hypothesis is true, the probability to accept the null hypothesis is directly defined by the chosen significance level α. Therefore, it is of interest to determine the probability to accept the null hypothesis when a difference between the populations is present. To determine this probability, the steps presented in figure 5.1 are required.

First, it is required to define a null hypothesis, an alternative hypothesis and the distance measurement method. Subsequently, the numerical samples and the experimental samples are drawn from their respective populations, which are either both bivariate normally distributed or bivariate lognormally distributed in this case. To simplify the procedure to calculate the operation characteristic, both samples are generated using random numbers. Here, the Monte Carlo Method (explained in subsection 3.3.4) has been proven to be a viable tool to generate random samples. To increase the efficiency of the numerical process, it is chosen to use the latin hyper cube method instead of crude Monte Carlo to generate the uniformly distributed random numbers. These uniformly distributed random numbers are transformed to either normally distributed samples or lognormally distributed samples using the procedure presented in subsection 3.3.4. Furthermore, a seed value is used for the random number generator to maintain the predefined nominal difference between the support points of the operation characteristic as well as for reproducibility of the results, because a seed value forces the random number generator to generate the same set of random numbers on every run. In this work the seed value 123457 is used.
Next, the distance between the numerical sample and the experimental sample is measured and a hypothesis test is performed to calculate the p-value corresponding to the measured distance. This hypothesis test can only be performed as a classical hypothesis test when the distribution of the distance measure is known. Otherwise, the hypothesis test must be performed using the

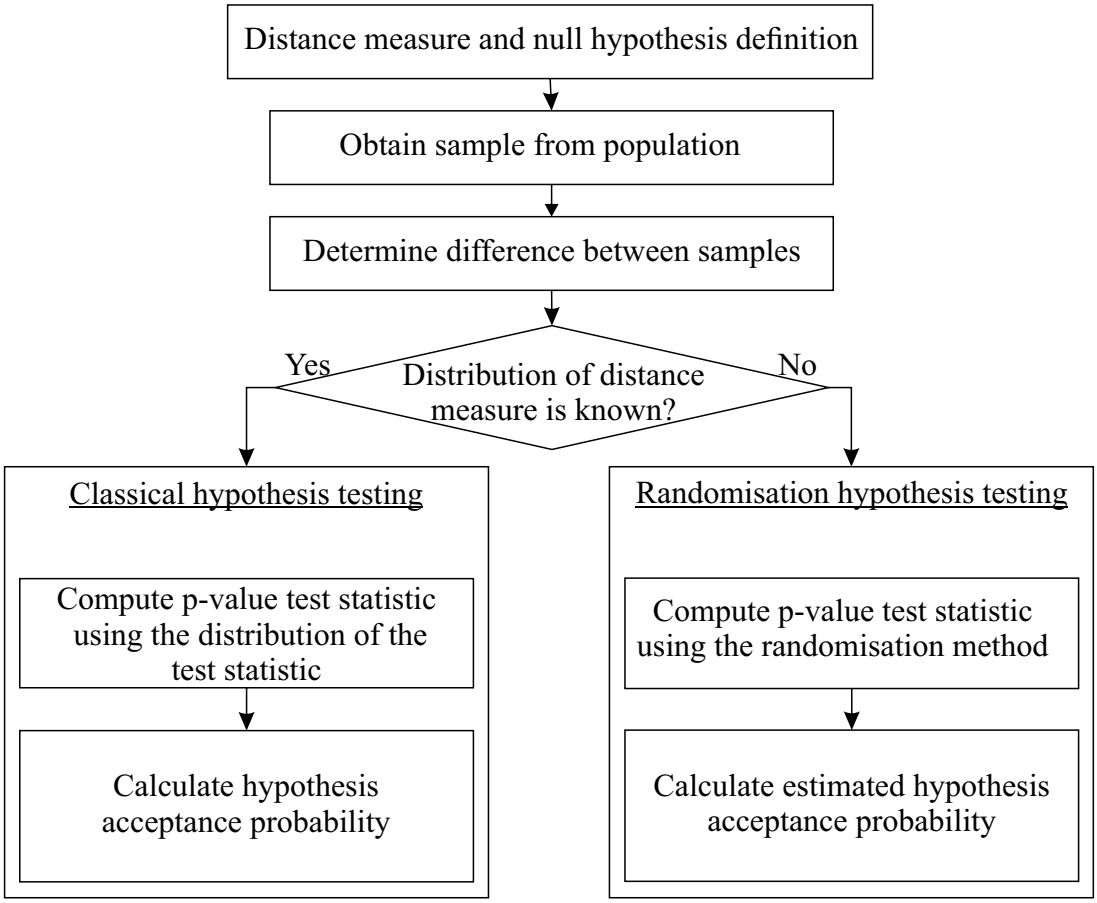


Figure 5.1: Flow diagram of the procedure to calculate the probability to accept the null hypothesis for support points of the operation characteristic.

randomisation hypothesis testing method presented in section 4.3. At this point, a note about the randomisation hypothesis testing method has to be made. Of course, this randomisation hypothesis testing method can also be used when the distribution of the distance measure is known. Each hypothesis test results in a p-value which is compared to the significance level α. For this work, a significance level of 5% is chosen, which is a common significance level for engineering problems [19, 20, 40]. When the hypothesis test results in a p-value smaller than the chosen significance level, the null hypothesis is rejected [19, 20, 40].
Subsequently, the probability to accept the null hypothesis is calculated, which depends on the distance $\boldsymbol{c}$ between the populations and the number of data points in the samples. Since the probability to accept the null hypothesis depends on the distance between the populations, it is commonly presented as operation characteristic [63, 62]. Here, $\boldsymbol{c}$ could be varied from zero to infinity. Nevertheless, it is not necessary to cover the entire range of $\boldsymbol{c}$, because the operation characteristic approaches zero quickly with increasing population separation. To further reduce the numerical calculation time, the probability to accept the null hypothesis is only calculated at several equidistant support points. These support points define the alternative hypotheses, which are used to calculate the corresponding values of β. Subsequently, the operation characteristic is created by linear interpolation of β between the support points. To illustrate the influence of the size of the samples on the probability to accept the null hypothesis, the operation characteristic is calculated for samples containing 5, 10 and 100 data points.
The procedure to calculate the probability to accept the null hypothesis depends on the distance measure and the distribution of the populations. In case the distribution of the populations is normally distributed, it is possible to determine the probability to accept the null hypothesis analytically for the Mahalanobis distance and the Box-M statistic. When the populations are not identical, the probability to accept the null hypothesis is equal to the type II error β. To calculate the value of β, equation 3.44 has to be solved for a specific number of data points per sample while assuming the alternative hypothesis is a specific alternative hypothesis. The type II error of the Mahalanobis distance is calculated with equation 5.1 [73],

$$\beta_{{D_M}^2} = \int_0^{x_{\text{crit}}} f_{d,v-d+1,\lambda}(x)\mathrm{d}x, \qquad x = \frac{v-d+1(n_1 n_2)}{vd(n_1+n_2)}{D_M}^2 \qquad v = n_1+n_2-2 \tag{5.1}$$

where $f_{d,v-d+1,\lambda}$ is the probability density function of the non-central F-distribution, ${D_M}^2$ is the Mahalanobis distance, d is the dimensionality of the problem, and λ is the non-centrality parameter of the non-central F-distribution. Furthermore, the Mahalanobis distance is calculated with equation 3.33 and the value of λ depends on the specific alternative hypothesis. For a mean value difference, the specfic alternative hypothesis is of the form $\boldsymbol{\mu}_{\text{sim}} = \boldsymbol{\mu}_{\text{exp}} + \boldsymbol{c}$, where $\boldsymbol{c}$ is the mean value difference between the two populations. The type II error for the Box-M statistic is calculated with equation 5.2,

$$\beta_m = \int_0^{m_{\text{crit}}} f_{\frac{1}{2}d(d+1),\lambda}(m)\mathrm{d}m \tag{5.2}$$

where $f_{\frac{1}{2}d(d+1),\lambda}$ and λ are the probability density function of the non-central Chi square distribution and the non-centrality parameter, respectively. The variable m must be non-central Chi square distributed when the covariances matrices are unequal, because m is Chi square distributed when the covariance matrices of simulation and experiment are identical [18]. The value of λ is equal to the difference between the populations of the alternative hypothesis. The Box-M statistic m is calculated with equation 3.42. For a covariance matrix difference, the specific alternative hypothesis is of the form $\boldsymbol{\sigma}_{\text{sim}} = \boldsymbol{\sigma}_{\text{exp}} + \boldsymbol{c}$, where $\boldsymbol{c}$ is the covariance matrix difference between the two

populations.
In case the randomisation hypothesis test is used to calculate the probability to accept the null hypothesis, equation 3.53 is used where the probability to accept the null hypothesis is equal to one minus the probability to reject the null hypothesis. To solve this equation, it is required to estimate the distribution of the p-values $f(\phi)$. This distribution is estimated by repeating the randomisation hypothesis test many times with new samples. Thus the result of equation 3.53 is approximated by using equation 5.3, where R is the amount of repeated hypothesis tests to estimate the distribution of ϕ.

$$\mathrm{P}\left(\text{reject}|\alpha,NS,R\right) = \sum_{nge=0}^{\alpha(NS+1)-1} \left(\begin{array}{c} NS \\ nge \end{array} \right) \frac{1}{R} \sum_{j=0}^{R} {\phi_j}^{nge} \left(1-\phi_j\right)^{NS-nge} \tag{5.3}$$

The flow diagram of the numerical model used to solve equation 5.3 is presented in figure 5.2. First one experimental sample and one numerical sample are drawn from their respective populations. Next, the distance is determined between these samples and NS pseudo samples are generated from

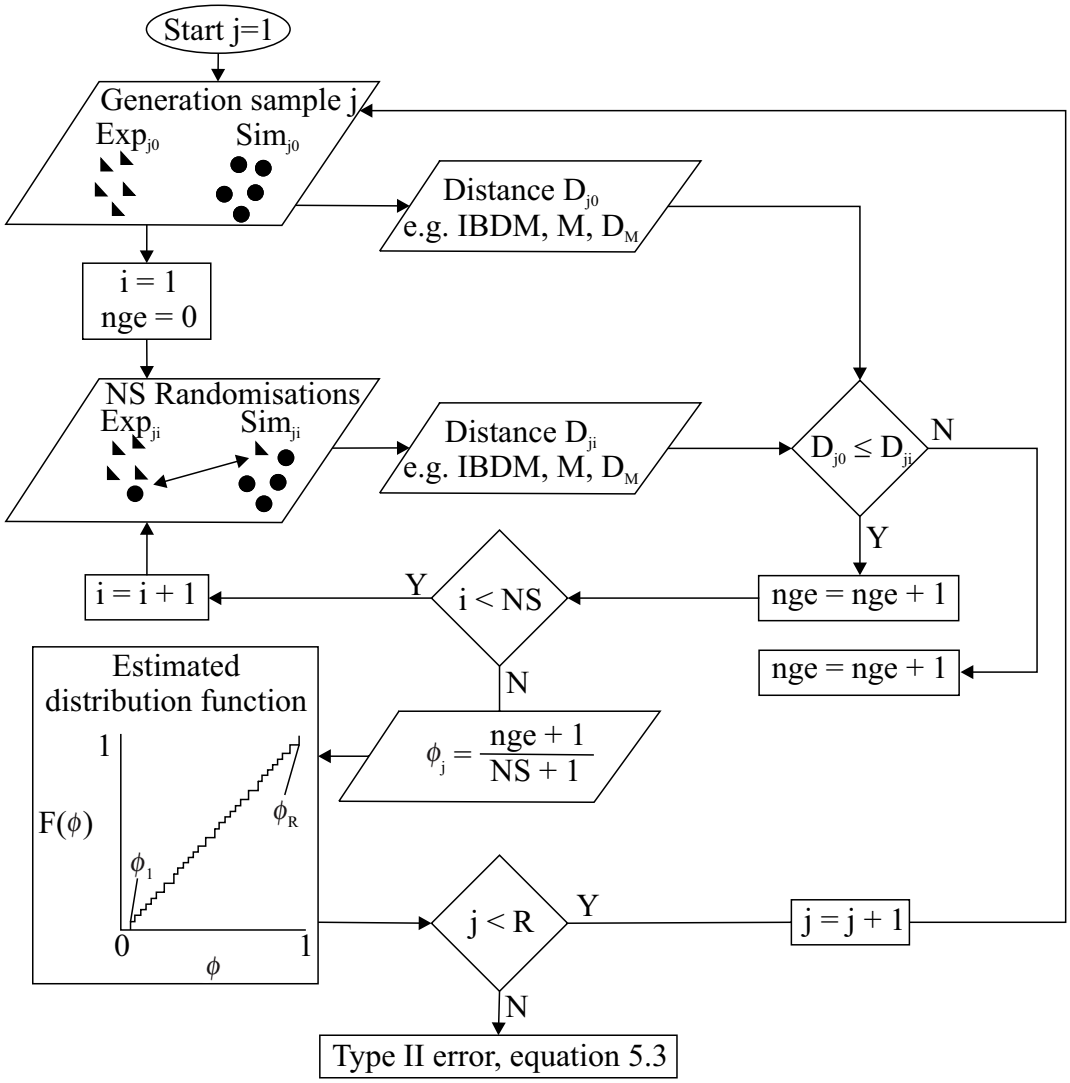


Figure 5.2: Flow diagram of the numerical model to calculate the probability to accept the null hypothesis using the randomisation hypothesis testing method.

these samples using randomisation. Subsequently, the distances between the NS pseudo samples are calculated. Next, it is calculated how many of pseudo samples yield in a distance larger than the distance between the original samples. Based on NS and the number of distances greater or equal than the original distance nge, the p-value ϕ_j is calculated. This process is repeated R times which results in the approximated distribution of ϕ, which is used to solve equation 5.3.
When the distribution of ϕ is estimated based on R values of ϕ, the distribution of ϕ is subjected to two errors. Namely, a discretisation error in the probability of ϕ which is related to R and an error in the values of ϕ due to inaccurate hypothesis tests. This second error can be minimised by increasing the number of randomisations used for the randomisation hypothesis tests. The other error can be minimised by increasing R. According to [65], 999 randomisations should result in accurate p-values ϕ. Consequently, the accuracy of the estimated distribution function of ϕ depends on R. To estimate the required amount of repeated hypothesis tests R to estimate the distribution of ϕ sufficiently accurate, the distribution of ϕ is estimated using different amounts of hypothesis tests using samples drawn from identically distributed populations. Since the populations are identically distributed, the distribution of ϕ is uniform distributed regardless of the used distance measure and the type of hypothesis test [18, 65]. To ensure that the classical hypothesis test and the randomisation hypothesis test are implemented adequately, the distribution of ϕ is investigated for both types of hypothesis test. This verification is presented in the next section.

5.3 Verification of the numerical model to calculate the probability to accept the null hypothesis

To investigate the required number of repeated hypothesis tests to calculate the type II error accurately, equation 5.3 is solved using 500 and 2000 repeated hypothesis tests. These hypothesis tests are based on the Mahalanobis distance between two bivariate normally distributed samples drawn from two identical populations. This reflects case 1 listed in section 5.1. Each sample contains 10 data points to obtain stable numerical results for the hypothesis tests [18]. Since the samples are drawn from identical populations, a uniform probability density function for ϕ is expected [18, 65]. Thus, 5% of the R ϕ values should be smaller than the significance level of 0.05. Consequently, an increased number of ϕ values smaller than 0.05 indicates a numerical error in this algorithm.
First the algorithm of the classical hypothesis test is verified. The estimated p-value distribution based on 500 repeated classical Mahalanobis hypothesis tests is presented in figure 5.3a whereas the p-value distribution based on 2000 repeated tests is presented in figure 5.3b. It is observed from figure 5.3a and figure 5.3b that the distribution of the p-values approaches the uniform distribution when the number of repeats is increased, as expected. Furthermore, it is shown that the numerical algorithm produces p-values smaller than 0.05 more frequently than the expected 5% of the R p-values. This error is caused by the Monte Carlo simulation and decreases with the amount of samples being generated, as presented in figure 5.3a and figure 5.3b. Since more p-values are smaller than 0.05 as expected, the probability to reject the null hypothesis is slightly larger than expected. Nevertheless, the amount of p-values smaller than 0.05 is only 2 percent points larger than expected for 500 repeated hypothesis tests. Therefore, it is assumed that the algorithm of the classical hypothesis test used in this work is sufficiently accurate to determine the probability to accept the null hypothesis.
Next, the p-value distribution of the Mahalanobis distance is calculated using 500 repeated randomisation hypothesis tests to verify the algorithm of the randomisation hypothesis test. As shown in figure 5.4b, the distribution of the p-values is approximately uniform distributed. Furthermore, the amount of p-values smaller than 0.05 is equal to 6.4% of the 500 p-values, which is slightly

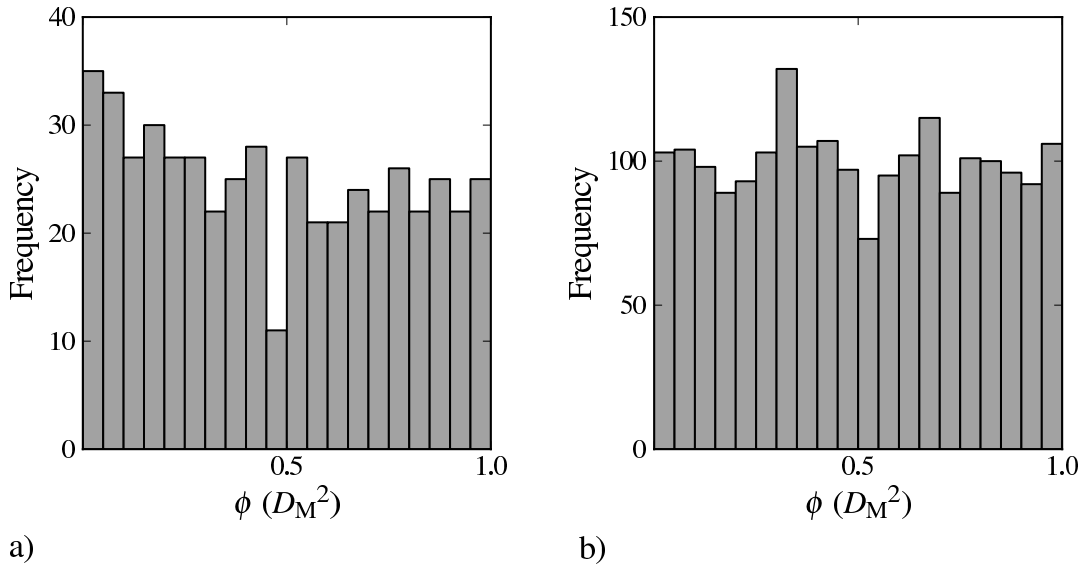


Figure 5.3: The estimated p-value distribution of: a) the Mahalanobis distance using the classical hypothesis test based on 500 samples, b) the Mahalanobis distance using the classical hypothesis test based on 2000 samples.

less compared to the 7% of the classical hypothesis test (see figure 5.4a). Since both methods only differ by 0.6 percent points, it is assumed that the randomisation hypothesis test is performed sufficiently accurate to calculate the probability to accept the null hypothesis.

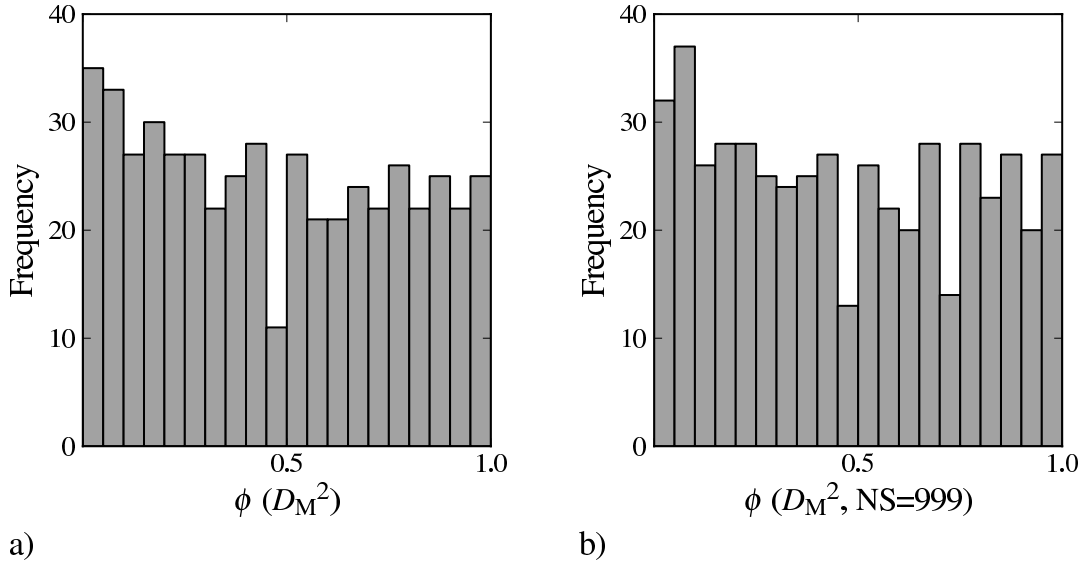


Figure 5.4: The estimated p-value distribution of: a) The Mahalanobis distance using the classical hypothesis test based on 500 samples, b) The Mahalanobis distance using the randomisation hypothesis test with 999 randomisations based on 500 samples.

Subsequently, it is verified that the result of the randomisation hypothesis test is independent of the distance measure, which should be the case for samples drawn from identically distributed populations. To test this case, the p-value distribution is determined using randomisation hypothesis test based on IBDM. The resulting p-value distribution is presented in figure 5.5b. It is observed from figure 5.5b that 7.4% of the 500 p-values are smaller than 0.05. Since the number of p-values smaller than 0.05 is only 1 percent point larger in figure 5.5b compared to figure 5.5a, it is assumed that the algorithm of the randomisation hypothesis test and IBDM are sufficiently accurate.
Since the algorithms based on 500 repeated hypothesis tests contains an error which could be up to approximately 2 percent points, the difference between two operation characteristics must be significantly greater than 2 percent points to call one distance measure more effective than another distance measure, which is expected for the used distance measures. Therefore, it is concluded that the numerical algorithms using 500 repeated hypothesis test are sufficiently accurate to compare the operation characteristics of two distance measures to another.
Nevertheless, due to computational limitations, it is required to limit the settings of the algorithms to 99 randomisations and 100 repeats to generate operation characteristics based on 20 support points for samples containing 100 data points. To avoid accuracy differences between the operation characteristics for different sample sizes, these less accurate settings are used in the next sections to illustrate the effect of the sample size on the operation characteristic using samples containing 5, 10 and 100 data points. When the difference between the operation characteristics is in the order of the numerical error of the algorithms, the calculation will be repeated using 500 repeated hypothesis tests and 999 randomisations.

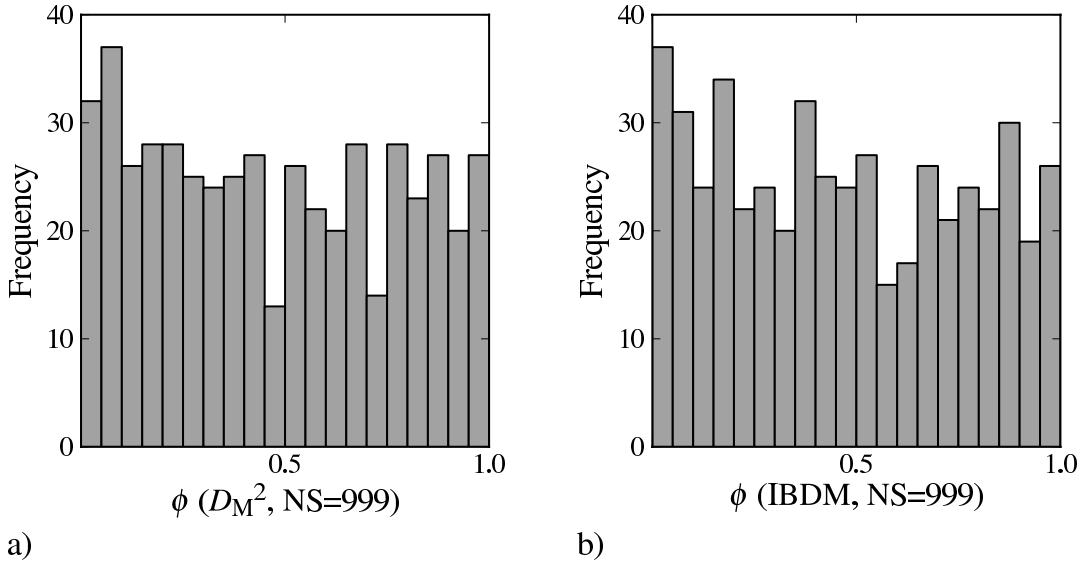


Figure 5.5: The estimated p-value distribution of: a) The Mahalanobis distance using the randomisation hypothesis test with 999 randomisations based on 500 samples, b) IBDM using the randomisation hypothesis test with 999 randomisations based on 500 samples.

5.4 Operation characteristics based on bivariate normally distributed samples

In this section, the operation characteristics are presented for the Mahalanobis distance, the Box-M statistic and the Image Based Distance Measure to determine whether IBDM is as effective as the two currently used distance measures to detect a difference between two bivariate normally distributed samples. To investigate whether this is the case, the first three cases defined in section 5.1 are investigated. More precisely, the effectiveness of the distance measures to detect the following three differences is determined: a mean value difference, a rotation and a combination of both differences.

5.4.1 Operation characteristics concerning a mean value difference

To determine if one method is more effective than another to detect a mean value difference (case 1 listed in section 5.1), an actual mean value difference must be present between the populations from which the samples are taken. Since every bivariate normal distribution can be normalised to a normal distribution with unit variance while the probability to accept the null hypothesis is identical for all mean value difference vectors of equal length between such samples, it is sufficient to use the normalised distributions to generate the numerical and experimental samples that are used to test the null hypothesis $\boldsymbol{\mu}_{\text{sim}} = \boldsymbol{\mu}_{\text{exp}}$. The parameters of these normalised distributions are presented in table 5.1. The null hypothesis concerning equal sample mean values is false, when the predefined difference between the populations is unequal to zero. As explained in section 5.2, the probability to accept the null hypothesis is equal to the type II error β when the populations are not identical. To calculate the type II error of an hypothesis test based on samples drawn from normally distributed populations of which the parameters are listed in table 5.1, a specific alternative hypothesis must be defined, which is presented in equation 5.4,

$$\boldsymbol{\mu}_{\text{sim}} = \boldsymbol{\mu}_{\text{exp}} + \begin{pmatrix} 2c\sigma_{1,\text{exp}} \\ 0 \end{pmatrix} \tag{5.4}$$

where c is the normalised mean value distance between the populations. The constant c is varied in the interval [0.05, 1.0] in steps of 0.05 to create a diagram of β versus c for a specific sample size, which is called an operation characteristic. Using the numerical model presented in section 5.2, the probability to accept the null hypothesis $\boldsymbol{\mu}_{\text{sim}} = \boldsymbol{\mu}_{\text{exp}}$ is calculated for every combination of the above specified sample sizes and values for c. More precisely, the probability to accept the null hypothesis is calculated using either equation 5.1 or equation 5.3 for the Mahalanobis distance and equation 5.3 for the Image Based Distance Measure. Connecting the support points with lines results in a linear approximation of the operation characteristics as presented in figures 5.6 and 5.7 for samples containing respectively 5, 10 data points and 100 data points. Comparing the operation characteristic of two methods to another reveals which method is more effective than the

Table 5.1: Parameters of the bivariate normal distributions.

	μ_1	μ_2	σ_1	σ_2
Simulation	2c	0.0	1.0	1.0
Experiment	0.0	0.0	1.0	1.0

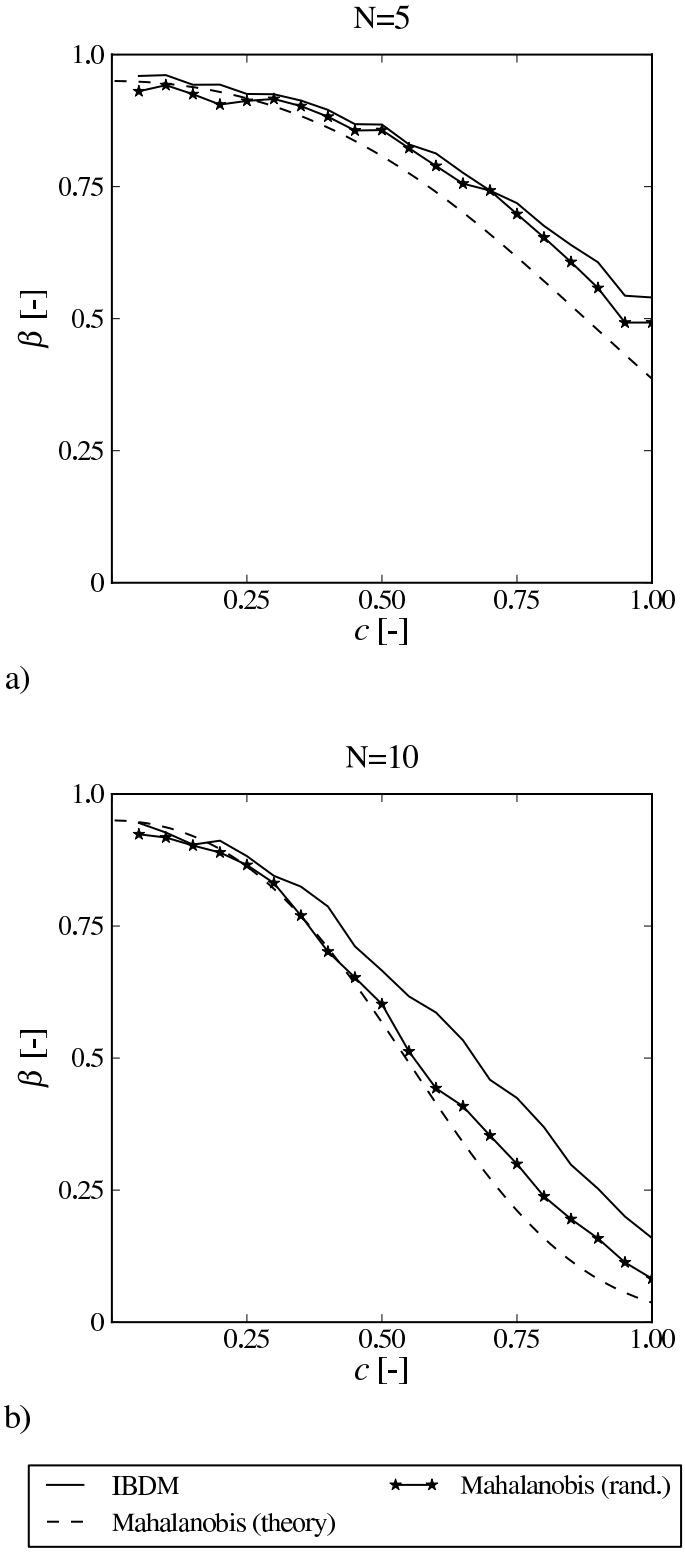


Figure 5.6: Operation characteristics for bivariate normally distributed samples which differ in mean value. Diagrams generated using the following settings for the numerical algorithm: a) N = 5, NS = 99, R = 100, b) N = 10, NS = 99, R = 100.

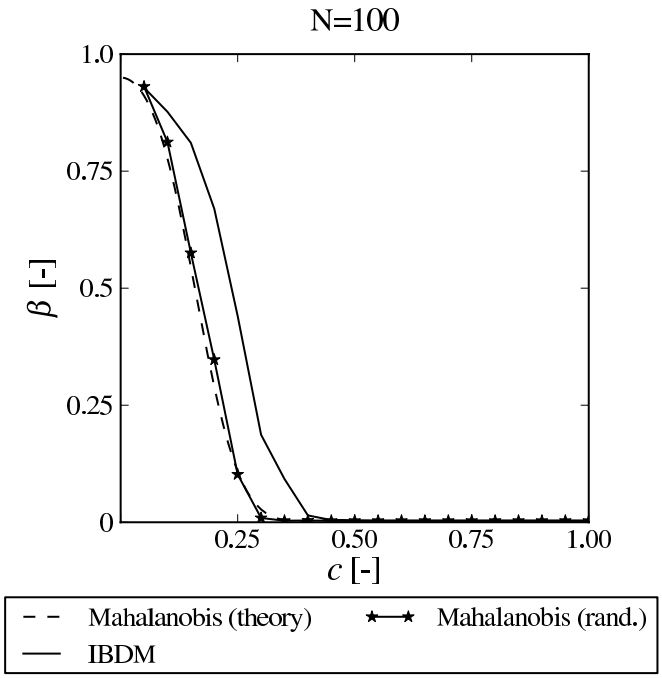


Figure 5.7: Operation characteristics for bivariate normally distributed samples which differ in mean value. Diagram generated using the following settings for the numerical algorithm: N = 100, NS = 99, R = 100.

other. The method belonging to the operation characteristic that approaches $\beta = 0$ the fastest is the most powerful method among the methods being compared to detect a mean difference between multivariate normally distributed samples of the specified size.
It is shown in figure 5.6 and 5.7 that the type II error decreases significantly with increasing number of data points per sample. Furthermore, it is shown that the type II error decreases with increasing predefined mean value difference between the samples. Especially in figure 5.7 it is shown that the type II error decreases rapidly with increasing predefined distance between the mean values of the samples, as expected [63]. Consequently, it is only possible to detect small mean value distances with confidence if the samples contain a large number of data points. This result was expected, because the type II error decreases with increasing number of data points [20, 63]. Moreover, it can be seen in figures 5.6 and 5.7 that the operation characteristic of the Mahalanobis distance based on the randomisation hypothesis test deviates from the analytical solution. Especially in figure 5.6a and 5.6b, a significant difference is present for $c > 0.4$. This difference must be caused by a numerical error, due to the limited amount of repeats (100 repeats) and the limited amount of randomisations (NS = 99) performed to obtain the p-value distributions of the distance measures. Since the analytical solution is correct, the numerical error can be quantified by determining the difference between the analytical operation characteristic and the operation characteristic based on the randomisation method of the Mahalanobis distance. Although a significant error is present in the absolute value of β when operation characteristics are based on the randomisation method, it is still possible to conclude which method is more effective as long as the error is smaller than the difference between the operation characteristics being compared. Furthermore, it is shown in section 5.3 that the difference between two probabilities to accept the null hypothesis based on the randomisation method is affected less by the numerical error than the absolute values of the probability to accept the null hypothesis.

For small distances, i.e. $c < 0.4$, the largest difference between the effectivity of IBDM and the Mahalanobis distance is observed for samples containing 100 data points. More precisely, IBDM is up to 40 percent points less effective than the Mahalanobis distance, see figure 5.7. Reducing the number of data points per sample to 10 data points decreases the difference in effectiveness between IBDM and the Mahalanobis distance as presented in figure 5.6b. Although a significant numerical error is present in the randomisation based solutions, the difference between IBDM and the Mahalanobis distance is larger than the difference between the two operation characteristics of the Mahalanobis distance. Consequently, IBDM is up to 8 percent points less effective than the Mahalanobis distance when c is larger than 0.3. Concerning samples containing 5 data points, it is observed that there is no significant difference between the operation characteristic of IBDM and the operation characteristic of the Mahalanobis distance based on the randomisation method as illustrated in 5.6a. Consequently, it is concluded that IBDM is approximately equally effective to detect a mean value difference between normally distributed samples containing 5 data points.
Since the results based on 5 and 10 data points contain a significant numerical error, it will be demonstrated that the difference between the operation characteristics based on the randomisation method is not affected significantly by the numerical error to ensure that the conclusions based on figure 5.6 are correct. For this purpose, the operation characteristics based on samples containing 10 data points are recalculated more accurately. To calculate the operation characteristics more accurately, equation 5.3 is solved with 999 randomisations and 500 repeated hypothesis tests. Furthermore, to illustrate the influence of both parameters on the operation characteristic, different combinations of the number of repeated hypothesis tests R and the amount of randomisations NS are used.
To determine the influence of the parameter NS on the operation characteristic, the amount of randomisations has been increased to 999 without changing the value of R, which results in the operation characteristics presented in figure 5.8b. Comparing figure 5.8a to figure 5.8b reveals that the numerical error has been decreased significantly for $c > 0.65$. Furthermore, it is also observed that the numerical error at very small differences has been increased. Using 500 repeated hypothesis tests instead of 100 reduces the numerical error related to the estimated distribution of ϕ without increasing the accuracy of the randomisation hypothesis test. This also increases the accuracy of the operation characteristic as shown in figure 5.9a. More precisely, increasing the number of repeats increases the smoothness of the operation characteristics as can be observed by comparing figure 5.8a to figure 5.9a. To reduce the numerical error to an insignificant value, both improvements have to be applied simultaneously, which results in the operation characteristics shown in figure 5.9b.
It is shown in figure 5.9b that the operation characteristic based on the Mahalanobis distance created with the randomisation hypothesis test converges to the theoretical operation characteristic, which confirms that the operation characteristic can be calculated accurately using the randomisation method at the cost of computational effort. Furthermore, it is observed that the difference between the operation characteristic of IBDM and the operation characteristic of the Mahalanobis distance has not changed significantly in figure 5.9a compared to figure 5.8a. Therefore, it is concluded that the numerical error has no significant influence on the difference between the operation characteristics based on the randomisation method, as expected. Thus, it is most likely sufficient to use 99 randomisations and 100 repeated hypothesis tests to determine whether IBDM is more effective than currently used distance measure for case 2 and 3 specified in section 5.1. Consequently, the operation characteristics based on randomisation hypothesis tests presented in the following two subsections are created using 99 randomisations and 100 repeated hypothesis tests.

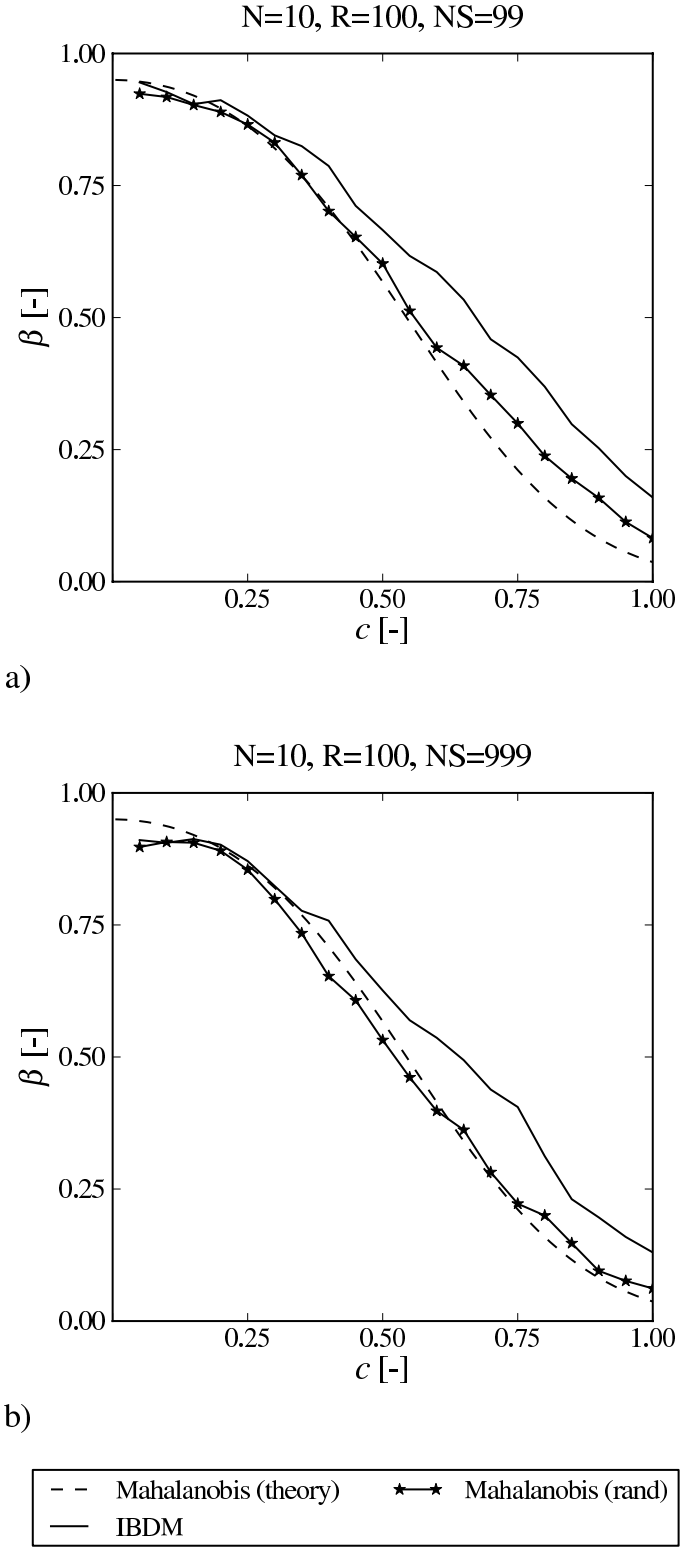


Figure 5.8: Operation characteristic for bivariate normally distributed samples containing 10 points which differ in mean value. Diagrams generated using the following settings for the numerical algorithm: a) N = 10, NS = 99, R = 100, b) N = 10, NS = 999, R = 100.

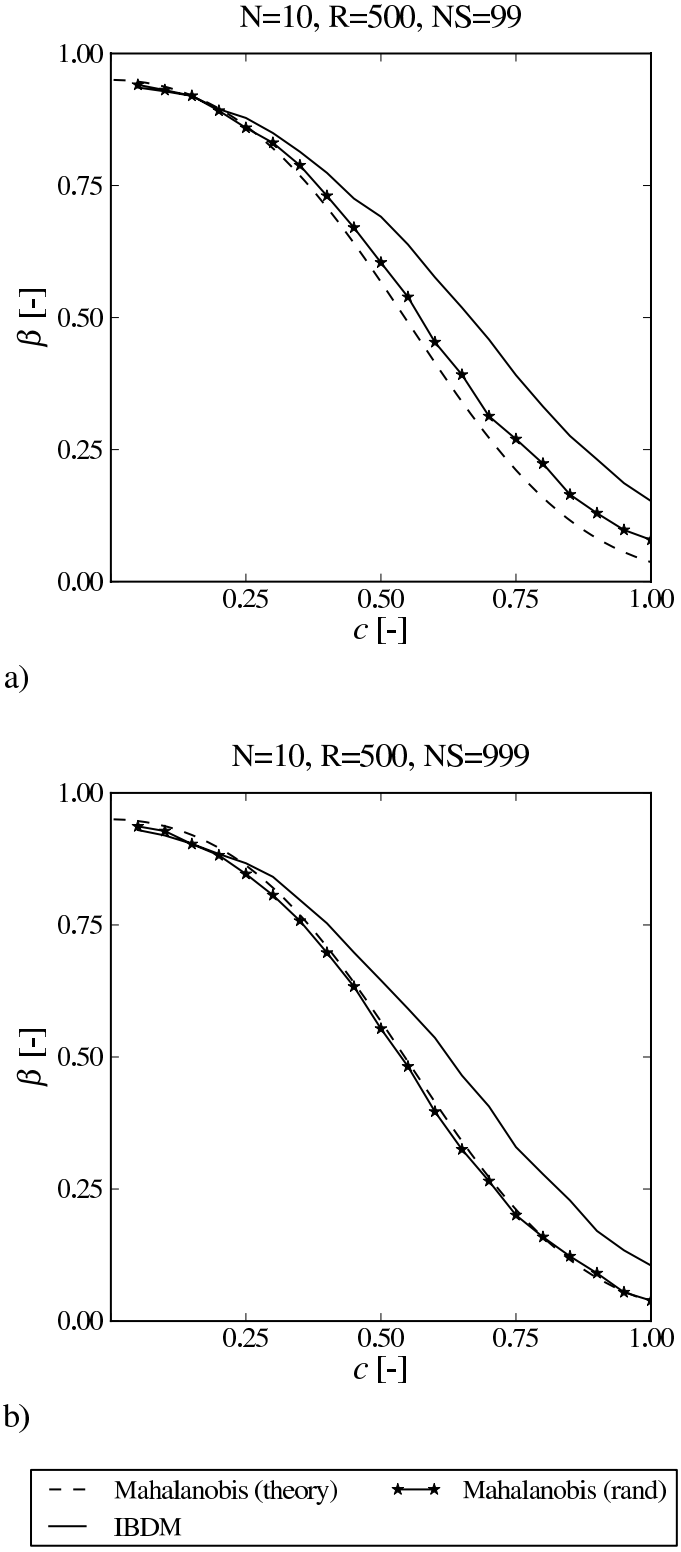


Figure 5.9: Operation characteristic for bivariate normally distributed samples containing 10 points which differ in mean value. Diagrams generated using the following settings for the numerical algorithm: a) N = 10, NS = 99, R = 500, b) N = 10, NS = 999, R = 500.

5.4.2 Operation characteristics concerning a covariance matrix difference

To examine whether the Image Based Validation Method is more or less effective to detect a covariance matrix difference than the currently used validation method based on the Box-M statistic (case 2 listed in section 5.1), such a difference must be present between the populations from which the samples are taken. For this case the difference is defined as rotational difference, which simulates a correlation matrix difference between the results of the experiments and the numerical model. To test the null hypothesis $\boldsymbol{\sigma}_{\text{sim}} = \boldsymbol{\sigma}_{\text{exp}}$, samples are drawn from the populations of which the parameters are presented in table 5.2. When the predefined rotational difference θ is unequal to zero, the null hypothesis is false. Consequently, the probability to accept the null hypothesis once more represents the type II error β, which again depends on the sample size and on the specific alternative hypothesis. This alternative hypothesis is defined using equation 5.5,

$$\boldsymbol{\sigma}_{\text{sim}} = \begin{pmatrix} cos^2(\theta) & -sin^2(\theta) \\ sin^2(\theta) & cos^2(\theta) \end{pmatrix} \boldsymbol{\sigma}_{\text{exp}}, \qquad \theta = \frac{c}{2}\pi \tag{5.5}$$

where $\boldsymbol{\sigma}_{\text{exp}}$ is the covariance matrix of the experimental population, θ is the angle of rotation (see figure 2.4) and c is the normalised angle of rotation. Varying c from 0.05 to 1.0 in steps of 0.05 gives the operation characteristic for one sample size. Once again, to include the influence of the number of data points per sample, the operation characteristic are calculated for samples containing 5, 10 and 100 data points as defined in section 5.2.

To create the operation characteristic for the Box-M statistic and IBDM, the probability to accept the null hypothesis is calculated for both methods at every support point. Currently the covariance matrix difference is measured using the distance measure M^2, which is calculated using equation 3.35. This distance measure can be transformed to the Box-M statistic (see equation 3.42) of which the probability to accept the null hypothesis can be calculated using equation 5.2. Alternatively, the probability to accept the null hypothesis based on the currently used distance measure can be calculated with equation 5.3. For the Image Based Distance Measure, the probability to accept the null hypothesis has to be calculated using equation 5.3, because the distribution of the distance measure is unknown. The resulting operation characteristics are presented in figures 5.10 and 5.11. It can be observed from figures 5.10 and 5.11 that the type II error decreases with increasing number of data points per sample as well as with increasing distance between the samples. Furthermore, it can be concluded from figures 5.10a and 5.10b that it is difficult to detect the rotational difference when the samples contain few data points, because β decreases slowly with increasing rotational difference. Even in case of Box-M (theory), which is the most effective method to detect a covariance matrix difference, β is still approximately equal to 50% for $c = 1.0$, i.e. a rotational difference of 90 degrees. The problem of being unable to detect a rotational difference between small samples arises from the significant sample to sample covariance matrix variation, which leads to a poor performance of all distance measures. Since it is recommended in literature

Table 5.2: Parameters of the bivariate normal distributions.

	Mean value 1	Mean value 2	covariance matrix
Simulation	0.0	0.0	$\begin{pmatrix} \cos^2(\theta) & -9.0\sin^2(\theta) \\ \sin^2(\theta) & 9.0\cos^2(\theta) \end{pmatrix}$
Experiment	0.0	0.0	$\begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 9.0 \end{pmatrix}$

to use samples that contain at least 10 data points to detect a covariance matrix difference between bivariate normally distributed samples [18], one could already have been expecting in advance that the distance measures would be ineffective for samples containing 5 data points.

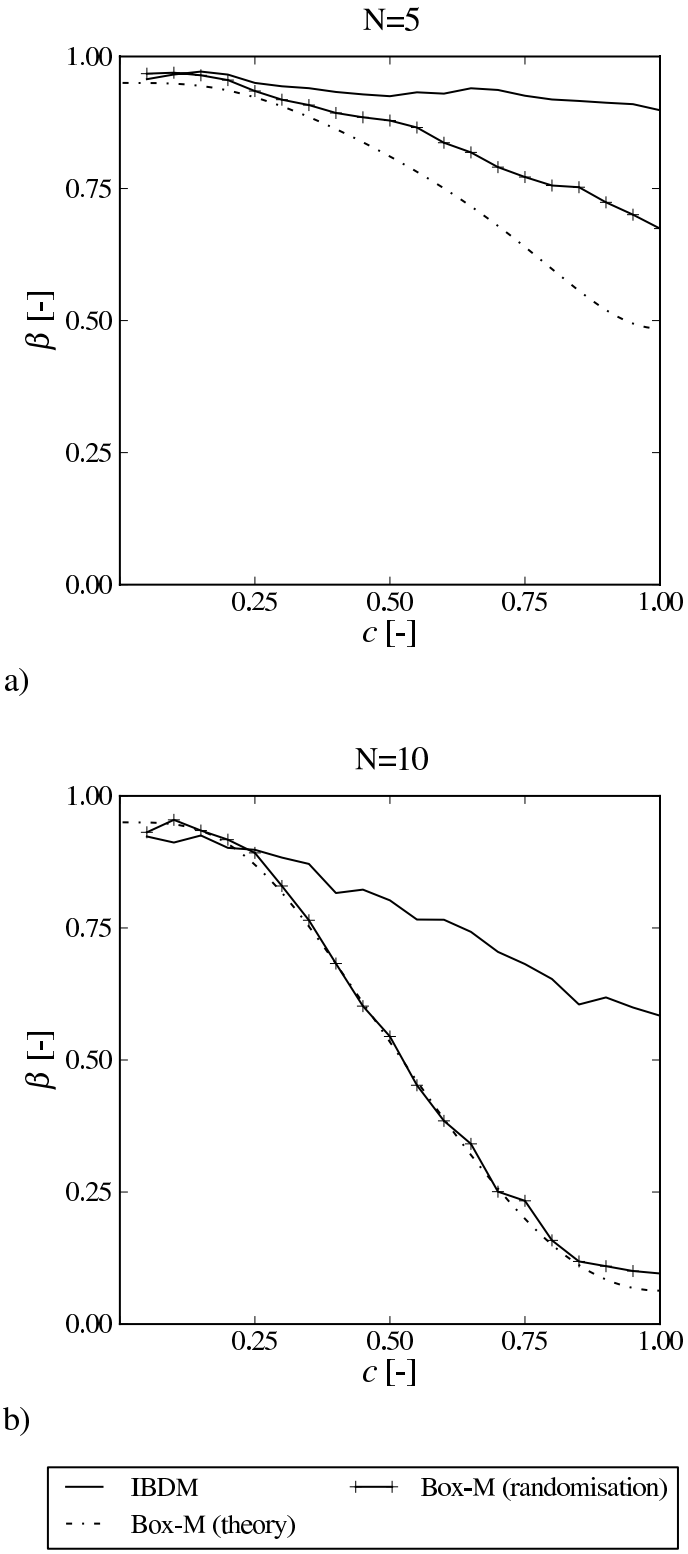


Figure 5.10: Operation characteristics for bivariate normally distributed samples which are rotated relatively to each other. Diagrams generated using the following settings for the numerical algorithm: a) N = 5, NS = 99, R = 100, b) N = 10, NS = 99, R = 100.

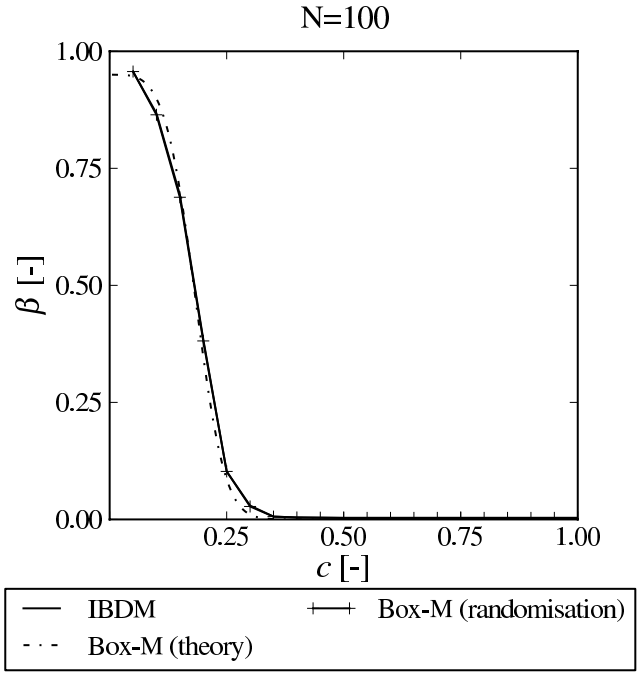


Figure 5.11: Operation characteristics for bivariate normally distributed samples which are rotated relatively to each other. Diagram generated using the following settings for the numerical algorithm: N = 100, NS = 99, R = 100.

Increasing the amount of data points per sample from 5 to 10 increases the probability to detect a rotational difference as presented in figure 5.10b. Furthermore, the operation characteristic of the Box-M statistic based on the randomisation hypothesis test converges to the analytical solution, which indicates that the numerical algorithm is sufficiently accurate if the samples contain 10 or more data points. Moreover, it is observed from figures 5.10a and 5.10b that IBDM is equally effective compared to the Box-M statistic to detect a rotational difference up to $c = 0.25$, which corresponds to a rotational difference of 22.5 degrees. For small samples and large rotational differences IBDM is less effective than the Box-M statistic to detect large rotational differences between normal distributions for small samples.
It was expected that IBDM is less effective than the Box-M statistic to detect a rotational difference between small samples, because the underlying distribution of the samples cannot be estimated accurately using only a few data points. When the underlying distribution function is estimated inaccurately, it is logical that a distance measure based on these estimated distributions is ineffective. Likewise, it is expected that IBDM is as effective as the Box-M statistic to recognise a rotational difference between two samples taken from a bivariate normal distribution when the samples contain many data points. As shown in figure 5.11, this expectation is confirmed.

5.4.3 Operation characteristics concerning a combined difference

To investigate whether an established distance measure or IBDM is more effective to detect a combined mean value difference with a covariance matrix difference (case 3 listed in section 5.1), a test case is used. For this test case, the mean value difference and the rotational difference are changed simultaneously to violate one assumption of the classical hypothesis test of the Mahalanobis distance while the operation characteristic of the Box-M statistic is unaffected by the

additional mean value difference. Consequently, the experimental population is specified using the parameters listed in table 5.2, while the population of the simulation is specified using equation 5.4 and equation 5.5, where c is the difference between the populations. As in the prior subsections, the constant c is varied in the interval [0.05, 1.0] in steps of 0.05 to define support points for the operation characteristics. To calculate the probability to accept the null hypothesis $f_{\text{exp}}(\boldsymbol{x}) = f_{\text{sim}}(\boldsymbol{x})$ for the Mahalanobis distance, equation 5.3 has to be solved because the distribution of the Mahalanobis distance is unknown when the variances of populations are unequal. To calculate the type II error of the hypothesis test based on the Box-M statistic either equation 5.2 or equation 5.3 can be used, because the Box-M statistic is independent of the mean value of the samples. For the image based validation method, it is required to calculate the probability to accept the null hypothesis using equation 5.3. The resulting operation characteristics are presented in figures 5.12 and 5.13. From figures 5.12 and 5.13 it can be observed that the type II error β decreases with increasing distance between the samples and with increasing number of data points per sample. Furthermore, it is shown in figure 5.12a that a combined difference between two normally distributed samples containing only 5 data points is hardly detectable for any distance measure. Therefore, it is advisable to use samples containing sufficient data points, which can be defined as at least five times the problem dimensionality [18]. Consequently, a sample containing 5 data points is not meaningful to detect a difference between bivariate distributed samples. When the samples contain 10 data points, it is observed from figure 5.12b that the Image Based Distance Measure is up to 34 percent points less effective than the Box-M statistic and up to 21 percent points more effective than the Mahalanobis distance to detect the combined difference. Furthermore, IBDM is slightly more powerful than the Box-M statistic to detect this combined difference when c is smaller than 0.35. When the samples contain many data points, it is found that IBDM is more effective compared to the Mahalanobis distance and the Box-M statistic to detect this combined difference, especially for c smaller than 0.35 as shown in figure 5.13. Based on the previous observations, it is concluded that IBDM is more effective to detect moderate combined differences between normally distributed samples than the currently used distance measures.

Besides the increased effectivity to detect such a combined difference, there is another advantage of using IBDM to measure the distance: In contrast to the currently used distance measures, IBDM is sensitive to rotational differences and mean value differences. More precisely, the Box-M statistic is independent of the mean value difference, which leads to identical operation characteristics for the Box-M statistic in figures 5.12 and 5.10 and figures 5.13 and 5.11, respectively. The Mahalanobis distance is a mean value difference measure, which is normalised using the pooled covariance matrix. Thereby, it measures also the covariance matrix difference as a side effect. Comparing figures 5.12 and 5.13 to figures 5.6 and 5.7 reveals that the Mahalanobis distance is less powerful to detect the combined difference than a pure mean value difference. Although this loss of power decreases with increasing number of data points, it is advisable to reduce significant covariance matrix differences using an optimisation routine based on the Box-M statistic before a hypothesis test based on the Mahalanobis distance is performed. Consequently, it is always necessary to measure the distance between two samples using either the Box-M statistic and the Mahalanobis distance or IBDM, provided that the samples contain many data points, to determine whether a significant difference is present between two samples.

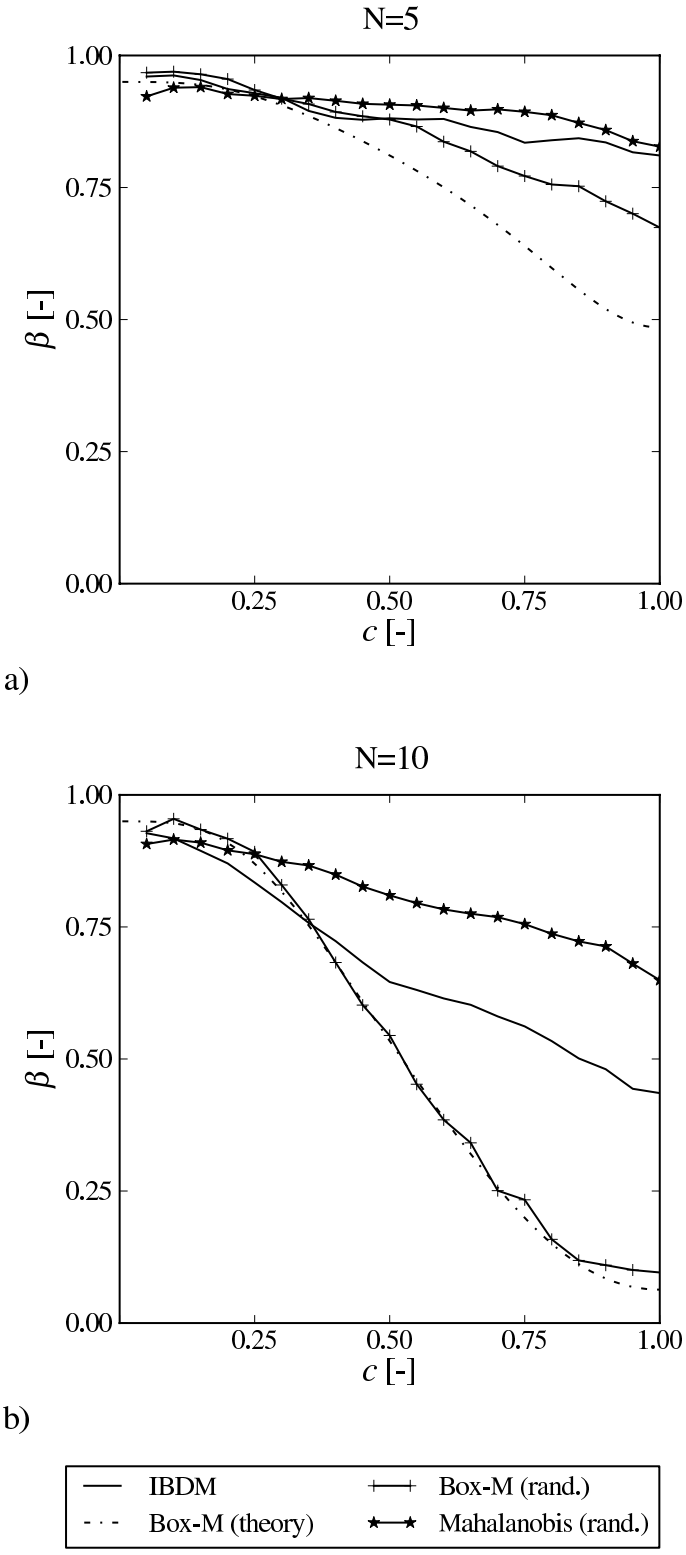


Figure 5.12: Operation characteristics for bivariate normally distributed samples which differ in mean value and are rotated relatively to each other. Diagrams generated using the following settings for the numerical algorithm: a) N = 5, NS = 99, R = 100, b) N = 10, NS = 99, R = 100.

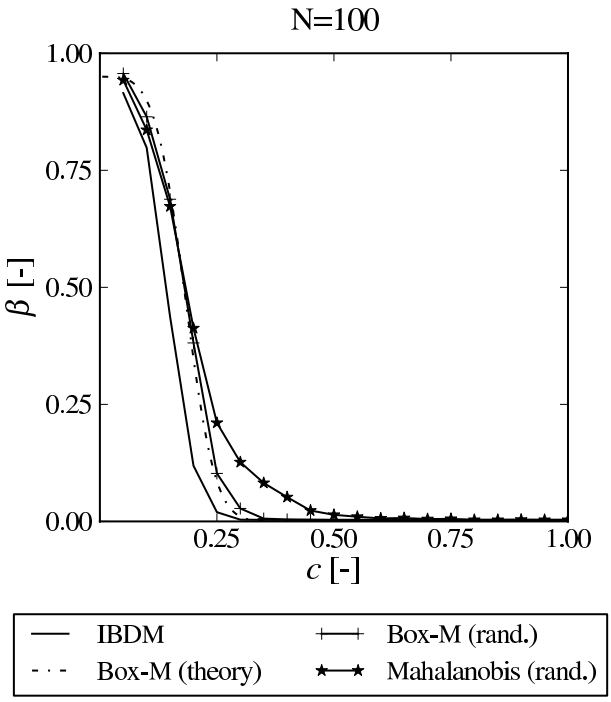


Figure 5.13: Operation characteristics for bivariate normally distributed samples which differ in mean value and are rotated relatively to each other. Diagram generated using the following settings for the numerical algorithm: N = 100, NS = 99, R = 10.

5.4.4 Discussion of the effectiveness of IBDM

In this chapter, the effectiveness of the IBDM has been compared to the effectiveness of the Box-M statistic and the Mahalanobis distance using the first three cases listed in section 5.1. In case 1, ideal conditions are created for the hypothesis tests based on the Mahalanobis distance to detect a mean value difference, whereas ideal conditions are created for the hypothesis tests based on the Box-M statistic in case 2. Consequently, the Image Based Distance Measure must be less effective than the Mahalanobis distance to detect a pure mean value difference between two normally distributed samples due to 100% fulfilled assumptions of the hypothesis test for this case (see figures 5.6 and 5.8). Similar, it was expected that IBDM is at most as effective as the Box-M statistic to detect a covariance matrix difference. As shown in figure 5.10, IBDM is less effective than the Box-M statistic for small samples and equally effective for large samples. Since the rotational difference between small samples is detected least ineffective using the Box-M statistic, it is recommended to detect a covariance matrix difference using the Box-M statistic if the samples contain few data points.

To investigate whether IBDM could be more effective than the Mahalanobis distance under less ideal conditions, case 3 has been defined such that one assumption of the hypothesis tests of the Mahalanobis distance is violated. Consequently, it is necessary to perform two hypothesis tests to determine if the difference is significant when using the currently used methods to detect the combined rotational and mean value difference. One to determine if the rotational difference is significant and another hypothesis test to determine if the mean values are identical. Both tests are required to determine if the samples are significantly different, because the Box-M statistic detects only covariance matrix differences whereas the Mahalanobis distance is insensitive to covariance matrix differences when the samples are equally large and contain many data points. In case the samples contain many data points, a combined difference is detected more effectively using IBDM

compared to the currently used distance measures as shown figure 5.12c. Even when the samples contain relatively few data points, it is useful to measure the distance using IBDM, because it is the only method that can detect a combined mean difference without violating the assumptions of the hypothesis test. To further investigate the effectiveness of the Mahalanobis distance, the Box-M statistic and IBDM, the probability to detect differences between bivariate lognormally distributed samples is calculated in the next section.

5.5 Operation characteristics based on bivariate lognormally distributed samples

As mentioned before, a mean value difference is typically measured using the Mahalanobis distance and a covariance matrix difference is commonly measured using the Box-M statistic. To determine whether IBDM is more effective to detect a difference between bivariate lognormally distributed samples than the currently used distance measures, the operation characteristic is calculated for cases 4 to 6 defined in section 5.1. These operation characteristics are presented in the following three subsections.

5.5.1 Operation characteristics concerning a mean value difference

To determine if one method is more effective than another method to detect a mean value difference between lognormally distributed samples, it is required that the populations from which the samples are drawn are unequal. Since every lognormal distribution can be transformed to a lognormal distribution with unit variance while the covariance matrix of both samples are equal for case 4 defined in section 5.1, it is sufficient to calculate the probability to accept the null hypothesis $\boldsymbol{\mu}_{\text{sim}} = \boldsymbol{\mu}_{\text{exp}}$ using samples drawn from lognormally distributed populations defined by the parameters presented in table 5.3.

Table 5.3: Parameters of the bivariate lognormal distributions.

	λ_1	λ_2	${\zeta_1}^2$	${\zeta_2}^2$
Simulation	$\ln(5.0)+2c-0.5{\zeta_1}^2$	$\ln(8.0)-0.5{\zeta_2}^2$	$\ln(2.0)$	$\ln(2.0)$
Experiment	$\ln(5.0)-0.5{\zeta_2}^2$	$\ln(8.0)-0.5{\zeta_2}^2$	$\ln(2.0)$	$\ln(2.0)$

As for the normally distributed case, the probability to accept the null hypothesis represents the type II error β, because the populations are not identical. The type II error again depends on the sample size and the specific alternative hypothesis, which is defined using equation 5.6,

$$\boldsymbol{\mu}_{\text{sim}} = \boldsymbol{\mu}_{\text{exp}} + \begin{pmatrix} \exp(\ln(5)+2\text{c})\text{ - }5) \\ 0 \end{pmatrix} \tag{5.6}$$

where c is the normalised mean value distance between the populations. As before, the constant c is variated in the interval [0.1, 1.0] in steps of 0.1 to create a diagram of β versus c for a specific sample size. This diagram is called an operation characteristic. To calculate the type II error for each support point of the operation characteristic, equation 5.3 is solved for each distance measure using samples containing 5, 10 and 100 data points. The resulting operation characteristics are presented in figures 5.14 and 5.15.

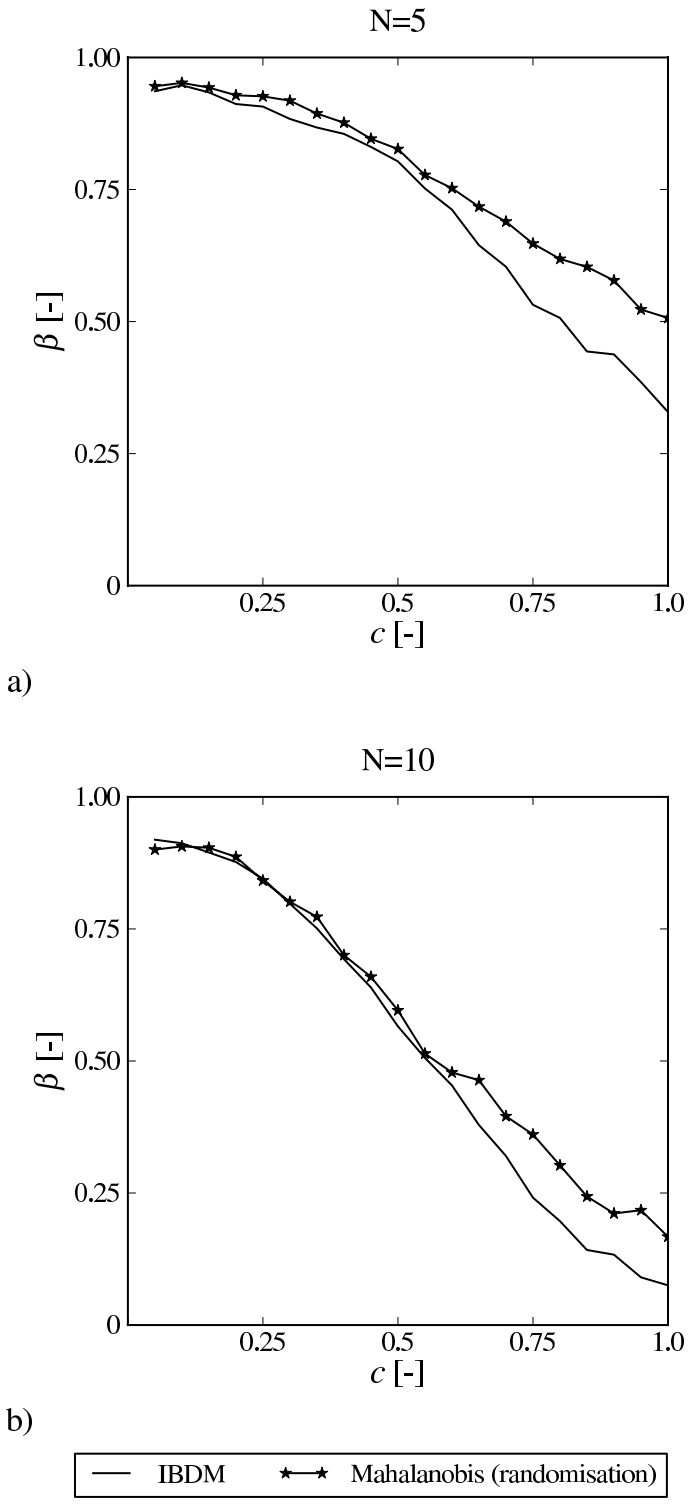


Figure 5.14: Operation characteristics for bivariate lognormally distributed samples which differ in mean value. Diagrams generated using the following settings for the numerical algorithm: a) N = 5, NS = 99, R = 100, b) N = 10, NS = 99, R = 100.

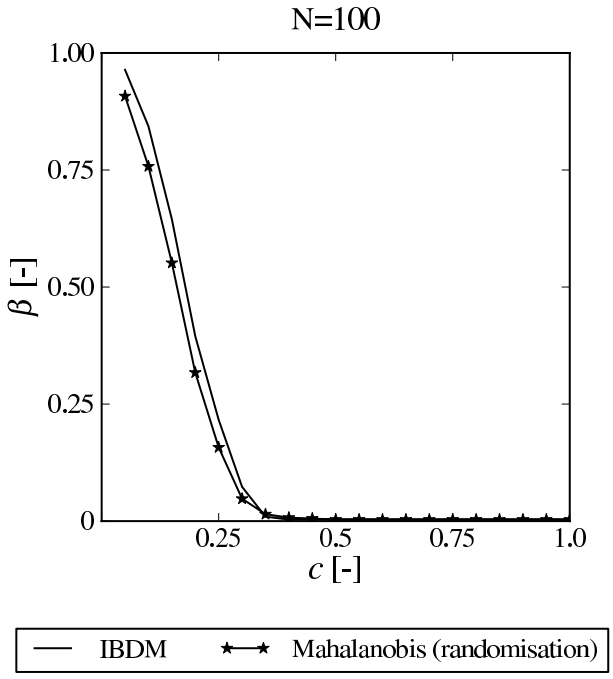


Figure 5.15: Operation characteristics for bivariate lognormally distributed samples which differ in mean value. Diagram generated using the following settings for the numerical algorithm: N = 100, NS = 99, R = 100.

It can be observed from figures 5.14 and 5.15 that the type II error β decreases with increasing mean value difference. Furthermore, the type II error decreases significantly when the number of data points per sample increases. Therefore, it is only possible to detect small differences with confidence using large samples, as already observed for normally distributed samples. When comparing the operation characteristics of IBDM to the operation characteristics of the Mahalanobis distance, it is observed from figure 5.14a and figure 5.14b that IBDM is more effective than the Mahalanobis distance to detect a mean value difference especially when c is larger than 0.6. Furthermore, for large samples, it is observed from figure 5.15 that the Mahalanobis distance is more effective than IBDM to detect a mean value difference. This result was expected, because a mean value difference is normally distributed regardless the distribution of the samples for large samples according to the central limit theorem [20, 40, 57, 58]. Since the Mahalanobis distance is the most effective mean value distance measure between normally distributed data (see section 5.4.1), a similar operation characteristic is expected for the Mahalanobis distance between lognormally distributed samples on basis of the central limit theorem. Although it is clear when IBDM is more effective than the Mahalanobis distance, the curves in figure 5.14 are unsmooth due to the less accurate settings used to calculate the operation characteristics.
To demonstrate that the numerical inaccuracy can be reduced to an insignificant value and to show that the numerical inaccuracy has no significant effect on the difference between the operation characteristics, the operation characteristics for samples containing 10 data points are recalculated using more accurate settings. Increasing the amount of randomisations NS performed during the randomisation hypothesis tests increases the accuracy of the p-values, which leads to slightly more powerful operation characteristics as can be seen by comparing figure 5.16a to 5.16b. Increasing the number of repetitions R of the randomisation hypothesis test increases the accuracy of the estimated distribution of the p-values. Thereby, it increases the accuracy of the operation characteristics, which results in smoother curves as shown in 5.17a. To reduce the numerical error to an

insignificant value for large values of c, e.g. $c = 1.0$, it is required to combine both improvements as presented in figure 5.17b. Since the tendencies in figure 5.16a and figure 5.17b are identical, it is concluded that the numerical error has no influence on the conclusions drawn based on figures 5.14 and 5.15.

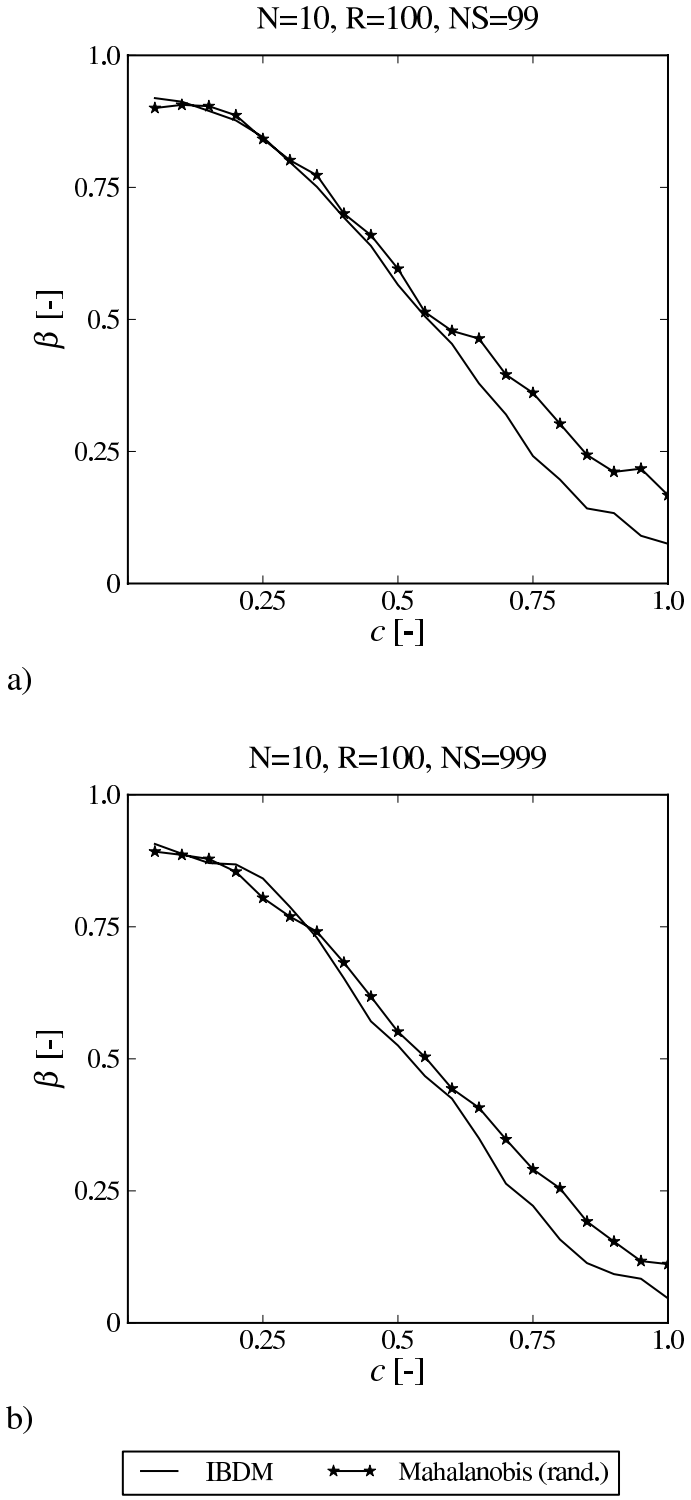


Figure 5.16: Operation characteristic for bivariate lognormally distributed samples containing 10 points which differ in mean value. Diagrams generated using different setting for the numerical alogrithm: a) N = 10, NS = 99, R = 100, b) N = 10, NS = 999, R = 100.

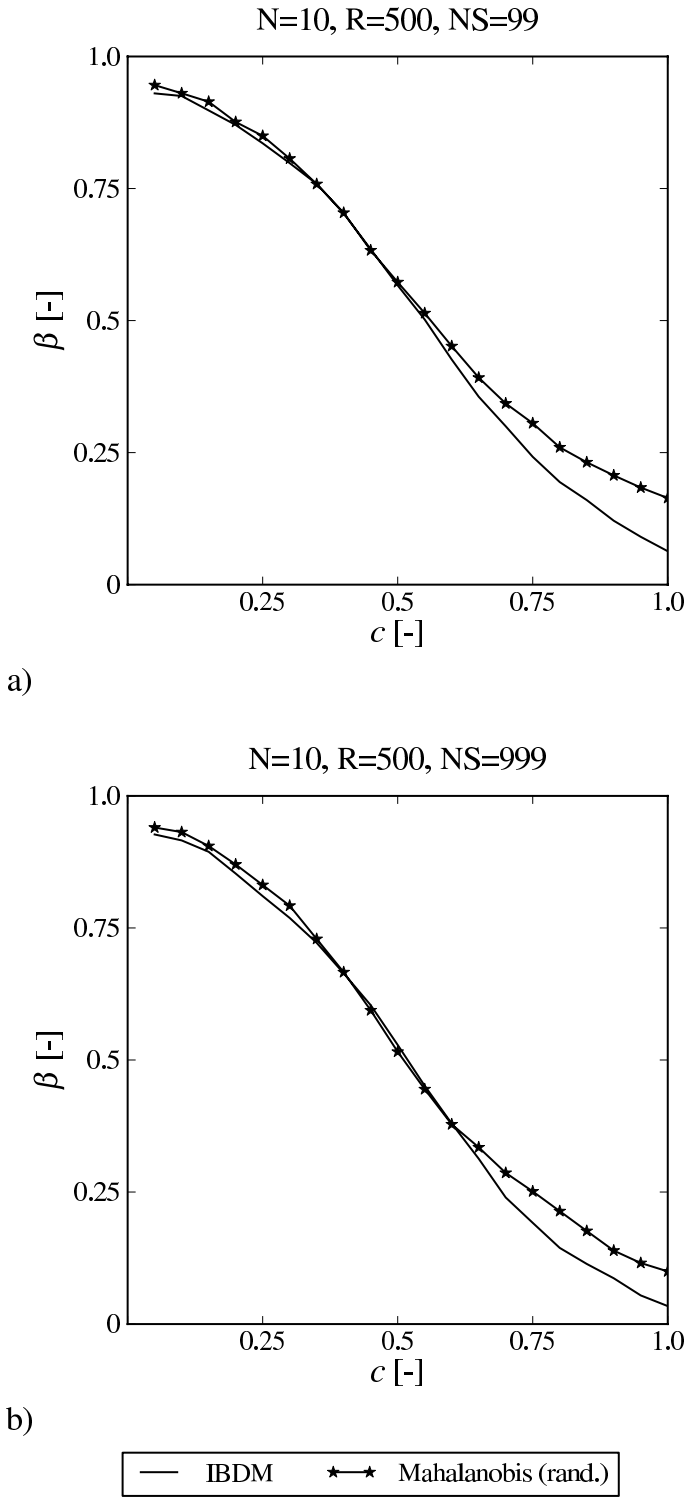


Figure 5.17: Operation characteristic for bivariate lognormally distributed samples containing 10 points which differ in mean value. Diagrams generated using different setting for the numerical alogrithm: a) N = 10, NS = 99, R = 500, b) N = 10, NS = 999, R = 500.

5.5.2 Operation characteristic of a covariance difference

To investigate whether IBDM is more powerful than the Box-M statistic to detect a covariance matrix difference between lognormally distributed samples (case 5 defined in section 5.1), the probability to accept the null hypothesis $\boldsymbol{\sigma}_{\text{exp}} = \boldsymbol{\sigma}_{\text{sim}}$ is determined. To calculate this probability, the covariance matrix difference is defined as a rotational difference between the bivariate lognormal distributions. The parameters of these distributions are presented in table 5.4.

Table 5.4: Parameters of the bivariate lognormal distribution.

	mean value 1	mean value 2	covariance matrix
Simulation	5.0	8.0	$\begin{pmatrix} \cos^2(\theta) & -9.0\sin^2(\theta) \\ \sin^2(\theta) & 9.0\cos^2(\theta) \end{pmatrix}$
Experiment	5.0	8.0	$\begin{pmatrix} 1.0 & 0.0 \\ 0.0 & 9.0 \end{pmatrix}$

Furthermore, the alternative hypothesis is specified by means of equation 5.5, where $\boldsymbol{\sigma}_{\text{exp}}$ is the covariance matrix of the experimental populations presented in table 5.4, θ is the angle of rotation and c is the normalised angle of rotation. Varying c from 0.05 to 1.0 in steps of 0.05 gives the operation characteristic for one sample size. To illustrate the influence of the number of data points per sample on the probability to accept the null hypothesis, the operation characteristics are calculated using samples that contain 5, 10 or 100 data points. Since the distribution of the populations is non-normal, it is necessary to calculate the probability to accept the null hypothesis using equation 5.3, which is solved by means of the numerical procedure presented in section 5.2. Calculating the probability to accept the null hypothesis for the defined values of c and N results in the operation characteristics, which are shown in figures 5.18 and 5.19.

It is observed from figures 5.18 and 5.19 that the type II error decreases with increasing rotational difference and increasing number of data points per sample. Furthermore, it is shown in figure 5.19 that IBDM is equally effective compared to the Box-M statistic for samples containing many data points. Moreover, the Box-M statistic is more effective than IBVM for samples containing 10 data points, because β is smaller for the Box-M statistic than for IBDM for $c > 0.15$ and the curve of IBDM for $c < 0.15$ is too low because it does not approach the value 0.95 when c goes to zero as shown in figure 5.18b. Since the operation characteristic does not approach the expected value of 0.95, a numerical error must be present. Nevertheless, it is not required to recalculate the operation characteristic using more accurate settings, because the slope of IBDM is significantly less steep than the one of the Box-M statistic. Further reduction of the number of data points will eventually lead to unreliable results of the hypothesis tests as shown in figure 5.18a. This figure again shows the importance of having enough data points to determine the distance. distance.

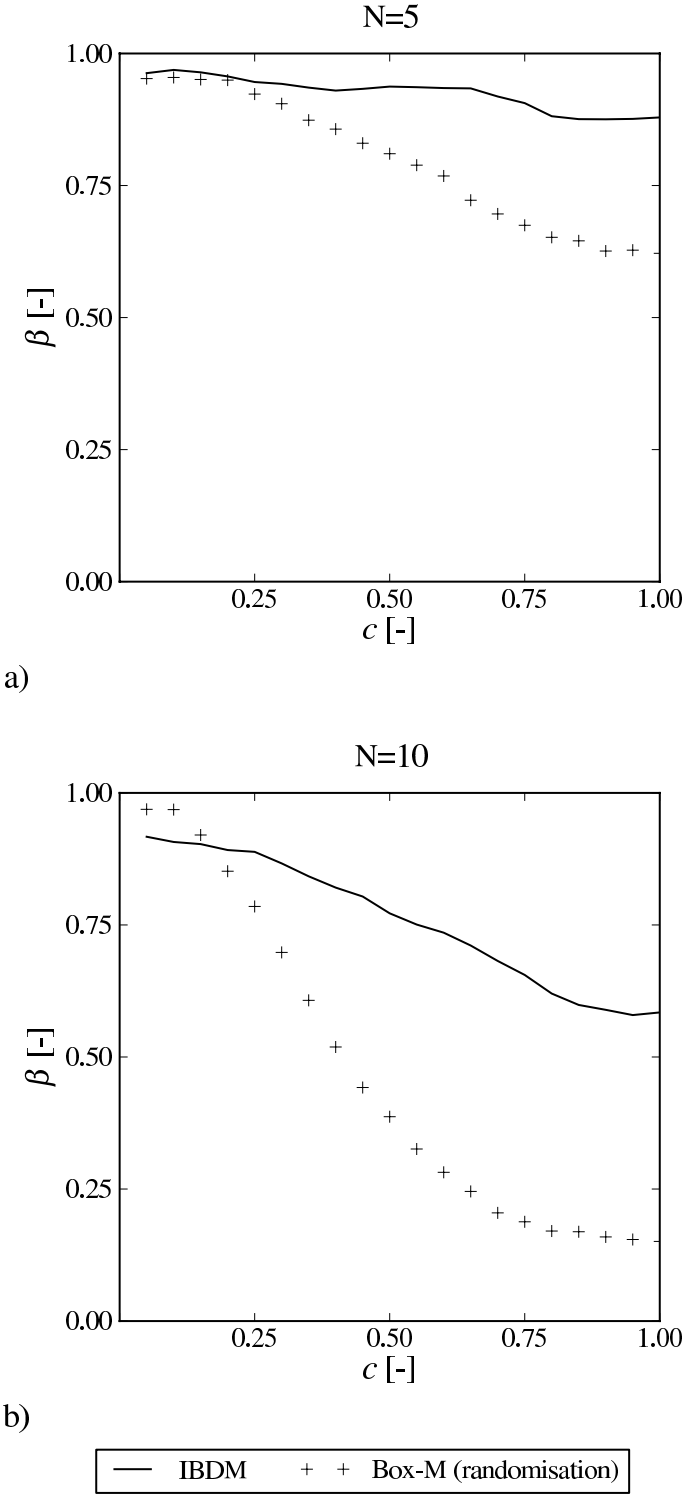


Figure 5.18: Operation characteristics for bivariate lognormally distributed samples which are rotated relatively to each other. Diagrams generated using the following settings for the numerical algorithm: a) N = 5, NS = 99, R = 100, b) N = 10, NS = 99, R = 100.

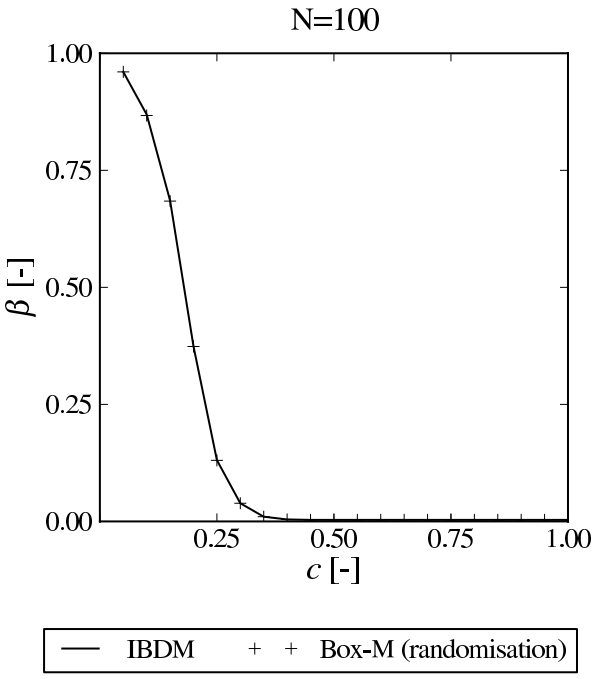


Figure 5.19: Operation characteristics for bivariate lognormally distributed samples which are rotated relatively to each other. Diagrams generated using the following settings for the numerical algorithm c) N = 100, NS = 99, R = 100.

5.5.3 Operation characteristic of a combined difference

To test whether a currently used distance measure is more effective than IBDM to detect a combined difference, the difference between the lognormally distributed populations for case 6 is defined similar to case 3 presented in section 5.4.3. Consequently, the difference is defined as a simultaneous translation and rotation of the simulation population. The experimental lognormally distributed population is defined using the parameters presented in table 5.4, while the population of the simulation is specified by means of equation 5.4 and equation 5.5. Using samples drawn from these distributions, the probability to accept the null hypothesis $f_{\mathrm{exp}}(\boldsymbol{x}) = f_{\mathrm{sim}}(\boldsymbol{x})$ is calculated, which is equal to the type II error when the samples are not identically distributed. To calculate the type II error of the Mahalanobis distance, the Box-M statistic and IBVM are calculated using equation 5.3, because the distributions of the distance measures are unknown. Again, this equation is solved for samples containing 5, 10 and 100 data points for values of c in the range [0.05, 1.0] in steps of 0.05 to obtain the operation characteristics. These operation characteristics are presented in figures 5.20 and 5.21.

It is shown in figures 5.20 and 5.21 that the type II error decreases with increasing distance between the samples (experiment and simulation). Furthermore, it decreases with increasing number of data points per sample. As shown in figures 5.20a and 5.20b, it is difficult to detect the difference using the Box-M statistic, because it is insensitive to a mean value difference. Furthermore, it is shown that the Mahalanobis distance and IBDM are approximately equal effective to detect this combined difference. When the samples contain 10 data points, it is shown that the difference can be measured most effective using the Mahalanobis distance or IBDM, which are equally effective. Further increasing the number of data points per sample to 100 increases the effectivity of all methods significantly, but IDBM and the Mahalanobis distance remain the most effective methods

to detect this difference. All in all, it is concluded from figures 5.20 and 5.21 that this combined difference can be detected equally effective with the Mahalanobis distance and IBDM.

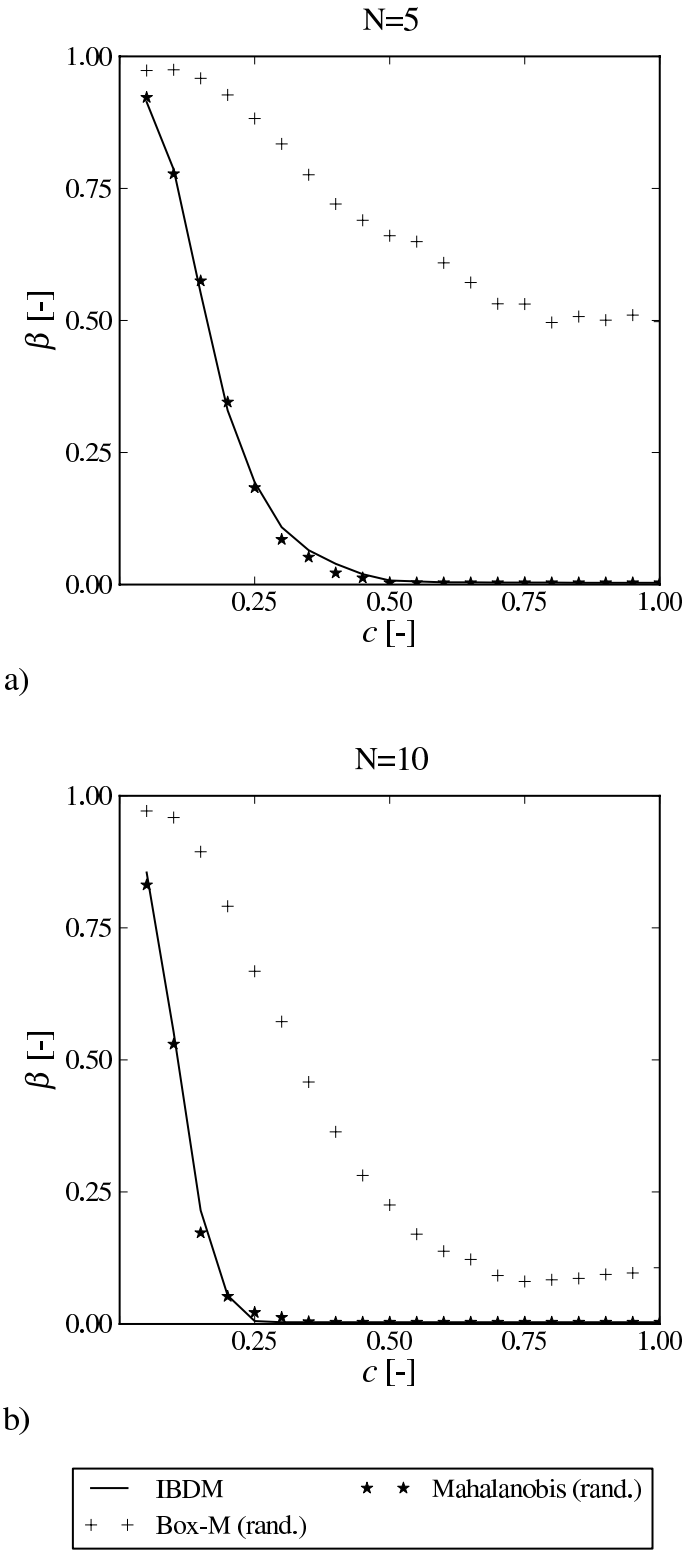


Figure 5.20: Operation characteristics for bivariate lognormally distributed samples which differ in mean value and are rotated relatively to each other. Diagrams generated using the following settings for the numerical algorithm: a) N = 5, NS = 99, R = 100, b) N = 10, NS = 99, R = 100

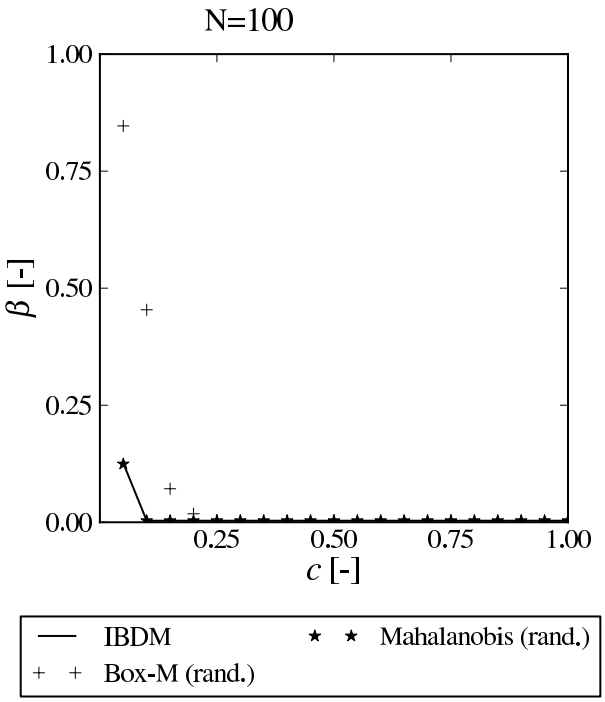


Figure 5.21: Operation characteristics for bivariate lognormally distributed samples which differ in mean value and are rotated relatively to each other. Diagram generated using the following settings for the numerical algorithm: N = 100, NS = 99, R = 100.

5.5.4 Discussion of the effectivity of IBDM

As expected, IBDM is more effective than the Mahalanobis distance to detect a pure mean value difference between lognormally distributed samples containing up to 10 data points as presented in figures 5.14 and 5.16. Furthermore, it is shown in figure 5.14c that IBDM is slightly less effective to detect a mean value difference compared to the Mahalanobis distance between large samples, as expected on the basis of the central limit theorem. Moreover, it is shown in figure 5.18c that IBDM can be an effective method to detect a rotational difference between lognormally distributed samples when the samples contain many data points. In addition, it is observed from figure 5.18 that the IBDM is less effective to detect a rotational difference than the Box-M statistic when the samples contain 10 data points or less.
More important is the effectiveness of IBDM to detect a mean value difference and covariance matrix difference simultaneously. It is shown in figure 5.20 that IBDM is equally effective to detect the combined difference for this test case, but this is only one of the many possible combinations of a mean value difference and a rotational difference. Furthermore, it has been shown in figure 5.14 that IBDM is more effective to detect a pure mean value difference for small samples while being equally effective for large samples to detect a pure mean value difference. Combining both results leads to the conclusion that IBDM is more effective to detect a difference with a dominant mean value distance for small samples. For large samples, it is concluded from figures 5.14 and 5.20 that IBDM is slightly less effective to detect a difference with a dominant mean value component. When the covariance matrix difference dominates the difference between two lognormally distributed samples, it is concluded from figure 5.20 and figure 5.18 that IBDM is less effective for small samples while being equally effective for large samples.

All in all, it can be concluded that IBDM is an effective method to detect a difference between lognormally distributed data when the mean value distance is dominant or the samples are sufficiently large. Furthermore, the currently used distance measures are only sensitive to either a covariance matrix difference or a mean value difference. Therefore, it is necessary to perform two hypothesis tests when the currently used distance measures are used. Since IBDM is at least as effective as the currently used methods for most differences while being sensitive to any difference between the samples, it is recommended to use IBDM to detect differences between non-normally distributed samples. In addition, it is recommended to use the Box-M statistic when a large covariance matrix difference is present between small samples, because IBDM is less effective than the Box-M statistic in this situation.

5.6 The effectivity of IBDM to detect a difference

To determine whether IBDM is more effective than the conventionally used Mahalanobis distance and Box-M statistic, the operation characteristics based on samples containing 10 and 100 data points of the normal distribution and the lognormal distribution are compared to each other. It is observed from figure 5.8 and figure 5.16 that IBDM is equally effective compared to the Mahalanobis distance to detect small mean value differences ($c < 0.25$) between samples containing 10 data points. Furthermore, it is shown in figure 5.16 that IBDM is at least as effective as the Mahalanobis distance to detect a mean value difference between lognormally distributed samples containing 10 data points. Moreover, it is shown in figures 5.6c and 5.14c that IBDM is less effective to determine a mean value difference compared to the Mahalanobis distance when the samples contain 100 data points. Since IBDM is only significantly less effective than the Mahalanobis distance when the samples are normally distributed or contain many data points, it can be concluded that IBDM is an effective method to detect a mean value difference between arbitrarily distributed results.
When a covariance matrix difference is present between the samples, it is concluded on basis of figure 5.10c and figure 5.18c that IBDM is equally effective to detect the difference compared to the Box-M statistic when sufficient data points are available to accurately estimate the distribution of the samples. However, IBDM is less effective than the Box-M statistic to detect rotational differences between small samples. Therefore, it is advisable to use the Box-M statistic to determine the covariance matrix difference between small samples, because this distance measure is least ineffective.
Concerning the effectivity of IBDM to detect a difference that consists of a mean value difference and a rotational difference, it is found that IBDM is at least as effective as the Mahalanobis distance when the rotational difference is smaller than approximately 30 degrees as presented in figures 5.12 and 5.20. However, when the covariance matrix difference dominates the distance and the samples contain less than 100 data points, it is more effective to use the Box-M statistic as shown in figures 5.10b, 5.18b and 5.12b. Therefore, it is advisable to use the Image Based Distance Measure to measure the difference between two arbitrarily distributed samples unless it is known that the samples mainly differ in covariance matrix.
In summary, the Image Based Distance Measure is an effective distance measure when the samples contain many data points to determine the distance between arbitrarily distributed results, but the advantages of using IBDM decrease when either the samples approach normality or the samples contain too few data points to detect shape differences between the estimated distribution function. Although IBDM can quantify any difference between normally distributed samples, the currently

used distance measures are in general more effective and require less computational effort compared to IBDM. Therefore, it should be investigated if the results are approximately normally distributed before using IBDM.

6 Validation of a sphere indentation model

As described in the introduction, the loads acting on a structural component, the manufactured geometry and the local material properties can all be subjected to scatter due to actual operating conditions, manufacturing conditions and tolerances. If these structural components are safety relevant, it could be necessary to further improve the state of the art life time prediction models by considering also local material variation.
In general, the variation of the material behaviour is caused by microstructural and morphological aspects in the lower millimetre range [38, 68, 74, 75]. For most structural components, these local properties cannot be determined using conventional uniaxial tests (e.g. tensile experiments), due to experimental limitations and difficulties. Instrumented sphere indentation is a method capable to determine local material properties in the lower millimetre range [35, 36, 37, 38]. To extract local material properties from indentation experiments, inverse parameter identification is performed using a numerical model of the experimental setup. Consequently, the numerical model must be an accurate representation of the actual indentation setup to reliably extract the relevant local material properties from an indentation test. Hence, the indentation model must be validated to ensure that the numerical model is an accurate representation of the experimental setup.
In this chapter, the validation process is presented that is used to determine whether the numerical model is an accurate representation of the experimental setup. In the first section, the flow chart of the validation process is explained. Next, the experiments are described and the experimental results are presented. Subsequently, the numerical model is outlined, which generates the numerical results that are compared to the experimental results. In the last section, the difference between numerical results and the experimental results is calculated and the steps taken to reduce this difference between simulation and experiment (i.e. model optimisation) are presented.

6.1 Validation process of the sphere indentation model

As stated in the introduction, the following three components are required to validate a numerical model: experimental results, numerical results and a validation method. Validation leads to the conclusion that the model is either valid or invalid. Since an engineer is not interested in an invalid model, it is very useful to extend the validation process by an optimisation loop to decrease the difference between the experimental results and the simulated results when the model is invalid. The validation process including an optimisation loop of the sphere indentation model is schematically presented in figure 6.1. The experimental results are measured during sphere indentation experiments, while the numerical results are calculated using a finite element based sphere indentation simulation. Validation in this work is always to be interpreted as validation in presence of scatter, e.g. the stress-strain behaviour of the material scatters in reality. Therefore, it is necessary to introduce random scatter in the input variables of the numerical model. To quantify and describe the variation of the stress-strain behaviour of the material, it is necessary to perform tensile exper-

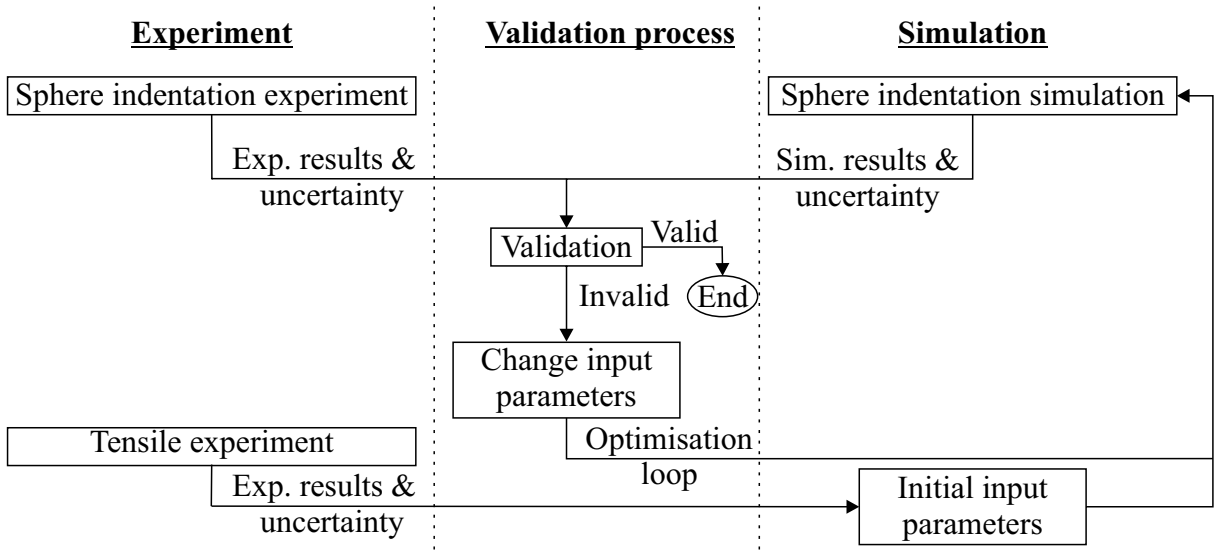


Figure 6.1: Flowchart of the validation process of a sphere indentation model.

iments as presented in figure 6.1. Their results provide the initial settings for the material model of the sphere indentation simulation.
During validation of the sphere indentation model, the simulated indentation curve (i.e. indentation depth versus load) is compared to the experimentally measured indentation curve to determine whether the numerical model is an accurate representation of the experiments. When the model is invalid, the unverified assumptions (see figure 2.1) are changed within their respective ranges using an optimisation process until a valid numerical model is obtained. This optimisation loop is performed using the adaptive response surface methodology of OptiSLang [67]. For further details concerning this optimisation algorithm the reader is referred to subsection 3.4.5. The box named validation in figure 6.1 is further broken down in figure 6.2.

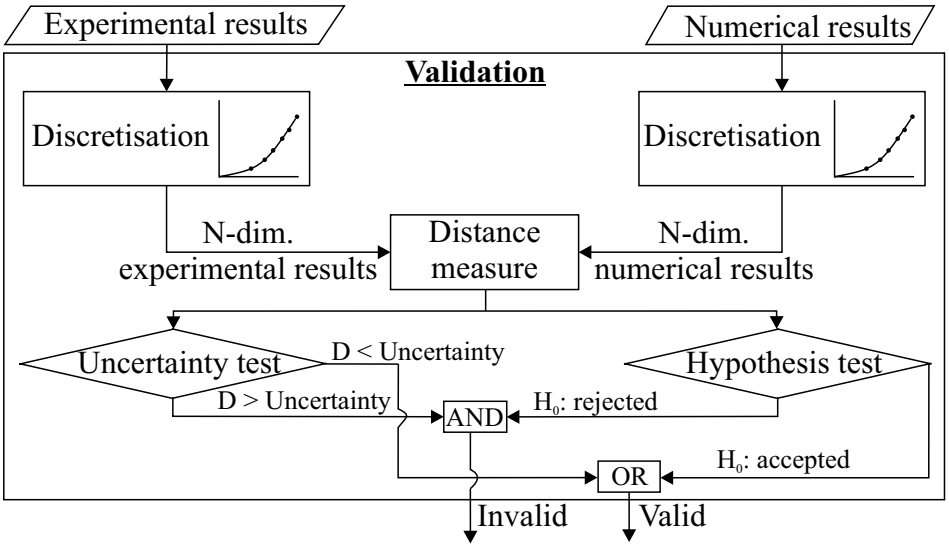


Figure 6.2: Flowchart of the steps to determine the validity of a sphere indentation model.

As shown in figure 6.2, the indentation curves are discretised at equidistant applied loads before measuring the distance between two indentation curves. Only the loading part of the indentation curve needs to be discretised to identify the local material hardening behaviour, because plastic deformation is a process taking place during loading and not during unloading. These discretised

loading curves yield the N-dimensional experimental and numerical results, i.e. the indentation depth at several loading levels. Using the N-dimensional results of the indentation curves, the distance between numerical results and experimental results is calculated. After computation of the distance between the numerical results and the experimental results, it is determined whether the distance is significant by means of a hypothesis test and a uncertainty based criterion. When the distance is statistically significant or larger than the uncertainty present in the numerical model and the measurements, an optimisation process is started to decrease the distance between simulation and experiment by changing the unverified assumptions of the numerical model (see figure 6.1).
The distance between numerical results and the experimental results is measured using the Image Based Distance Measure (IBDM) (see chapter 4) and the Mahalanobis distance, which is used as reference distance measure representing the state-of-the-art methods. The Image Based Distance Measure requires several steps to calculate the distance between the experimental and simulated results, as illustrated in figure 6.3.

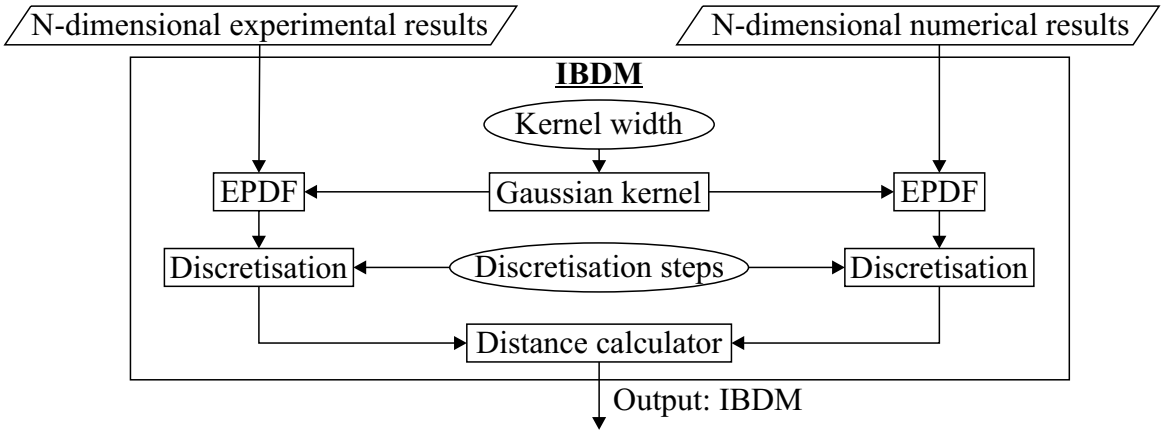


Figure 6.3: Flowchart of the steps to calculate the distance using IBDM.

During the first step, the probability density functions of the experimental and simulated results are estimated using the gaussian kernel estimator. The width of the kernel is based on either the measurement uncertainties or it is equal to the optimal Gaussian kernel width, which is calculated using equation 3.28 and the experimentally measured results. Subsequently, the Estimated Probability Density Functions (EPDFs) are discretised using 500 discretisation steps per direction as explained in section 4.2, which results in a probability distribution. The distance between two probability distributions is calculated using the overlap coefficient according to Manders, as explained in section 4.2. In case the dimensionality of the probability distribution is greater than two, all possible 2-dimensional projections of the probability distribution are used to measure the distance between the two multivariate probability distributions.
To calculate the Mahalanobis distance, the multivariate result spaces are used directly, as presented in figure 6.4. The mahalanobis distance is calculated using equation 3.33.

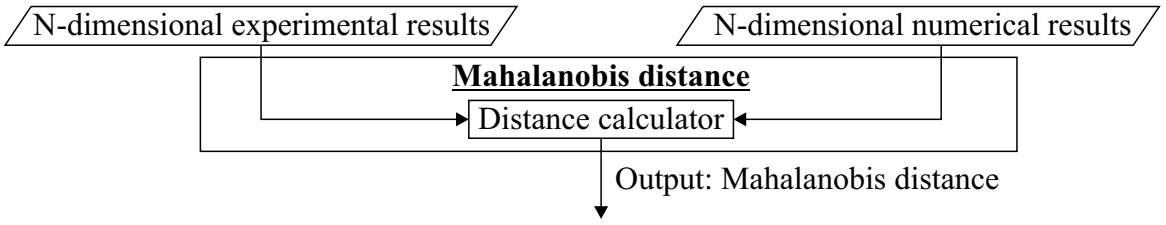


Figure 6.4: Flowchart of the steps to calculate the Mahalanobis distance.

As presented in figure 6.2, a hypothesis test is performed to determine whether the distance between simulation and experiment is significant. The steps required to perform the randomisation hypothesis test are schematically presented in figure 6.5. In the first step, randomisation is performed to obtain *NS* pseudo samples, as explained in section 4.3. Next, the distance between the pseudo samples are calculated and compared to the original distance. Based on the amount of pseudo distances larger than original distance, the p-value ϕ is calculated using equation 3.48. This p-value is compared to the user specified significance level α to determine whether the null hypothesis $f_{\text{exp}}(\boldsymbol{x}) = f_{\text{sim}}(\boldsymbol{x})$ has to be rejected. When the p-value ϕ is smaller than the significance level α, the null hypothesis being tested is rejected. Otherwise, the null hypothesis is accepted, which merely means that there is not enough evidence to reject the null hypothesis.

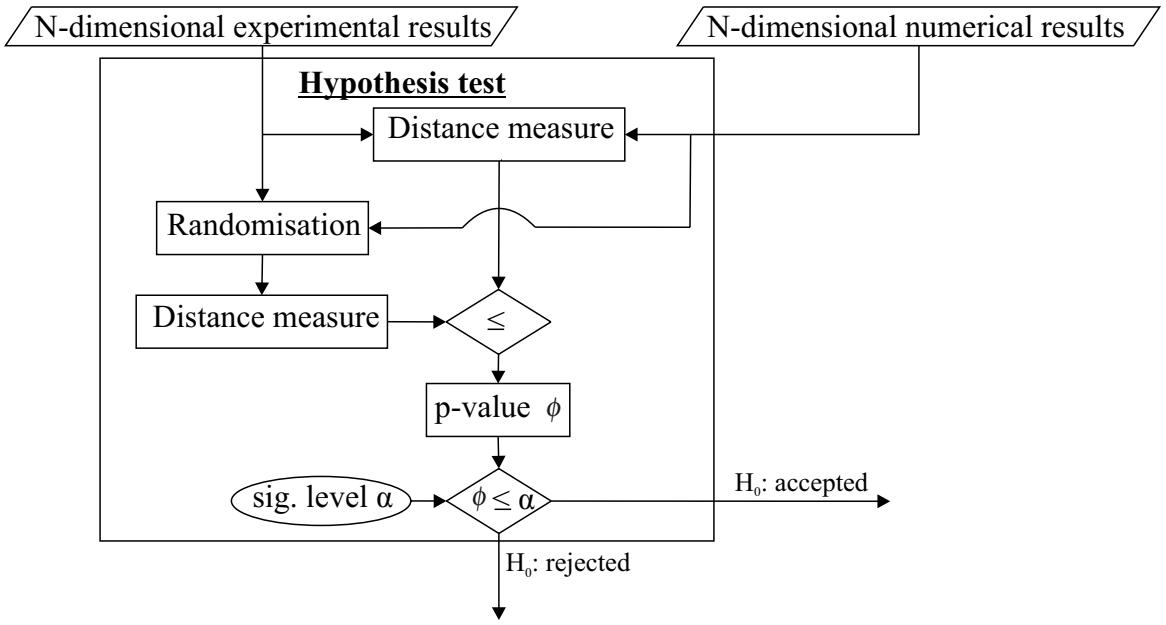


Figure 6.5: Flowchart of the randomisation hypothesis test, which is used to determine whether a difference is significant.

6.2 Experiments

To validate the indentation model, two kinds of experiments are performed as presented in figure 6.1. The sphere indentation experiments are used to validate the indentation model while the tensile experiments are performed to determine the variation of the stress-strain behaviour of the bulk material, which is an essential input for the sphere indentation simulations as presented in figure 6.1. To eliminate any batch-to-batch influence, the specimens are taken from the same plate. Furthermore, the specimens are extracted in a clustered pattern to reduce the systematic influence of the specimen location on the experimental results. The specimen extraction plan of the specimens tabulated in table 6.1 is illustrated in figure 6.6.
On the indentation specimen taken from the rolled plate, sphere indentation experiments are performed where the applied force versus indentation depth is measured. Using the tensile specimen, tensile experiments are performed where the specimen elongation versus applied force are measured. Furthermore, the measurement uncertainties are determined for both experiments to investigate their influence on the experimental results as shown in figure 6.1. It is necessary to determine

Table 6.1: Specimen overview containing specimen number, material name, type of specimen and orientation.

Specimen number	Material	Type of specimen	Orientation
1-20	Al2024-T351	Tensile specimen	perpendicular to rolling direction
21-40	Al2024-T351	Tensile specimen	parallel to rolling direction
41-60	Al2024-T351	Indentation specimen	-
61-80	Al2024-T351	Indentation specimen	-

the measurement uncertainties, because these will be used during validation to determine whether the distance between simulation and experiment is significant, see figure 6.2. Consequently, a detailed uncertainty analysis is performed. The experimental results are thus a combination of the experimental data and the measurement uncertainties as presented in figure 6.7.
In the following two subsections, the experimental setup, the experimental results and the measurement uncertainties of the sphere indentation experiments and the tensile experiments are presented.

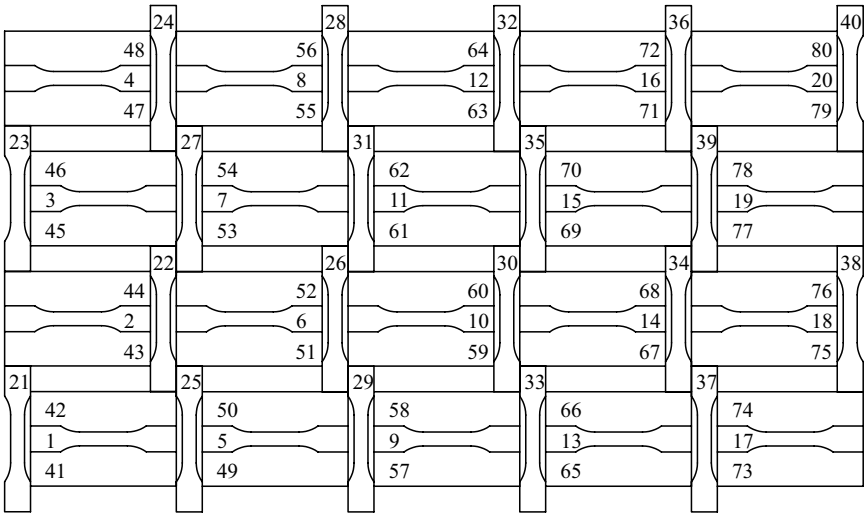


Figure 6.6: Sketch of the specimen extraction pattern.

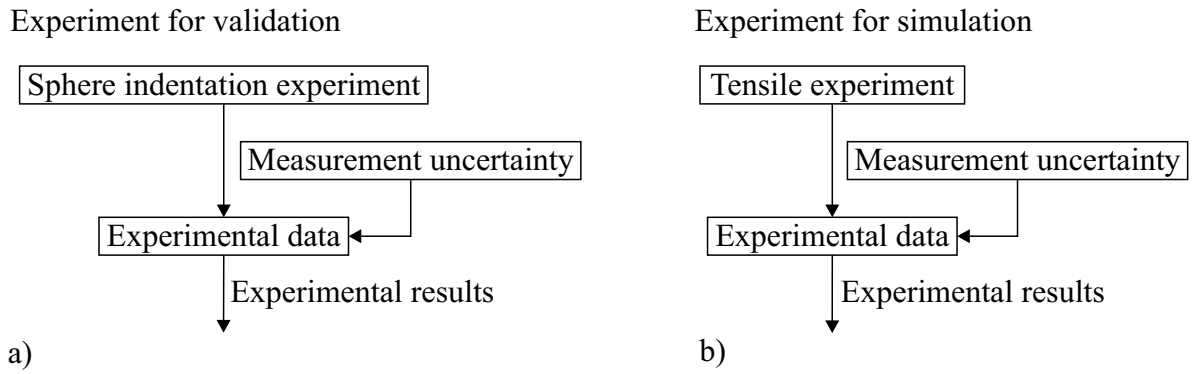


Figure 6.7: a) Flowchart of the sphere indentation experiments (experiment for validation), b) Flowchart of the tensile experiments (experiment for simulation).

6.2.1 Sphere indentation experiments

The experimental setup to measure the indentation depth during the indentation experiments is presented in figures 6.8a and 6.8b. It consists of a steel specimen support (I), a tungsten carbide indenter shaft with a diamond half sphere on the tip (II), a displacement transducer (III), and specimen clamps (IV). This experimental setup is installed in a universal testing rig (Instron 6665A) equipped with a 5 kN load cell. Load and indentation depth are measured simultaneously during the indentation experiments, which are performed using the load profile presented in figure 6.9. For this setup, a specimen with the nominal dimensions of 40 mm x 90 mm x 6 mm is used, see figure 6.8c. To minimise the measurement error, a high quality contact between specimen and anvil, and indenter and specimen are essential. This is obtained through an appropriate surface preparation process of the top and bottom side of the specimens after grinding both sides plain. The top surface (where the indents will be placed) is first finished using P2500 abrasive paper whereafter it is polished using a 3 μm diamond suspension and an oxide polishing suspension (0.05μm) while the bottom side is finished using P1200 abrasive paper. To ensure that the indents do not influence each other, a minimum distance of 10 mm (approximately 6x the indentation diameter) is maintained between the centres of the indentations as well as between the centres of the outer indentations and the edges of the specimen. This results in a total of 14 indents per indentation specimen, as shown in figure 6.8c.

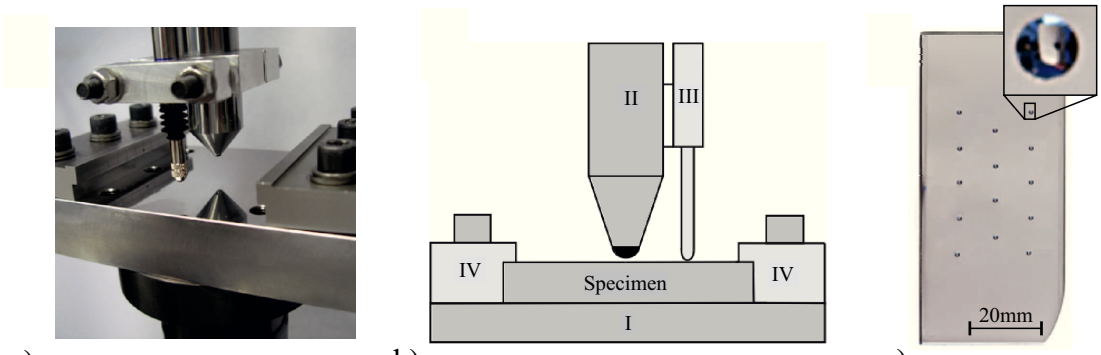


Figure 6.8: Experimental indentation setup: a) Photo of the experimental setup, b) Schematic overview of the components of the setup, c) Photo of a specimen after indentation experiments and a magnification of the indent.

To measure the indentation depth, an inductive displacement transducer is attached to the shaft of the indenter as shown in figure 6.8a. Consequently, the measured indentation depth is affected by

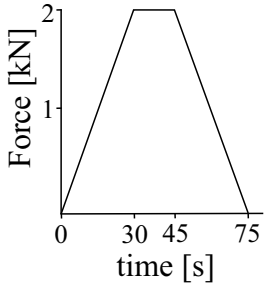


Figure 6.9: Loading curve of indentation experiments.

the elastic deformation of the indenter shaft. Therefore, it is necessary to correct the measured displacement for the deformation of the shaft to obtain the indentation depth accurately. This correction factor is determined from a finite element simulation of the indenter shaft. Using the displacement of the two points shown in figure 6.10a at several load levels, the compliance of the indenter shaft is calculated. Point P1 represents the point where the displacement transducer is attached to the shaft and point P2 represents the interface between the diamond tip and shaft. As indicated by the slope of the curve in figure 6.10b, the deformation behaviour of the shaft is linear, with a compliance equal to 0.65μm/kN.

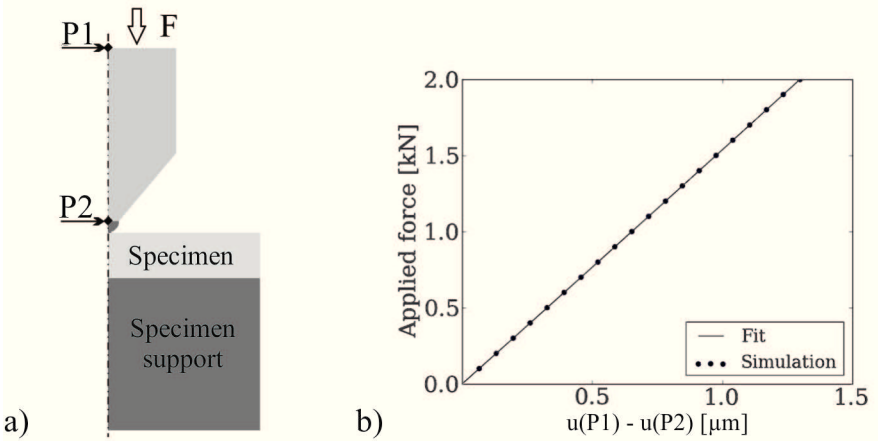


Figure 6.10: a) Shaft deformation model sketch for the section between diamond sphere P2 and displacement transducer attachment point P1, b) Deformation u of the shaft section.

To calculate material properties such as H_{IT} and E_{IT} from instrumented indentation experiments, it is necessary to know the relation between indentation depth and contact area. Usually the contact area is based on the indentation depth using the method of Oliver & Pharr [32]. For this it is required to know the actual indenter shape, which has been measured at Fraunhofer IWM Halle by means of white light interferometry [76, 77]. The result of this measurement is presented in figure 6.11a, where the colours represent the vertical height with respect to a reference plane. Based on figure 6.11a, it is concluded that the indenter is rotationally symmetrical, which is recognised from

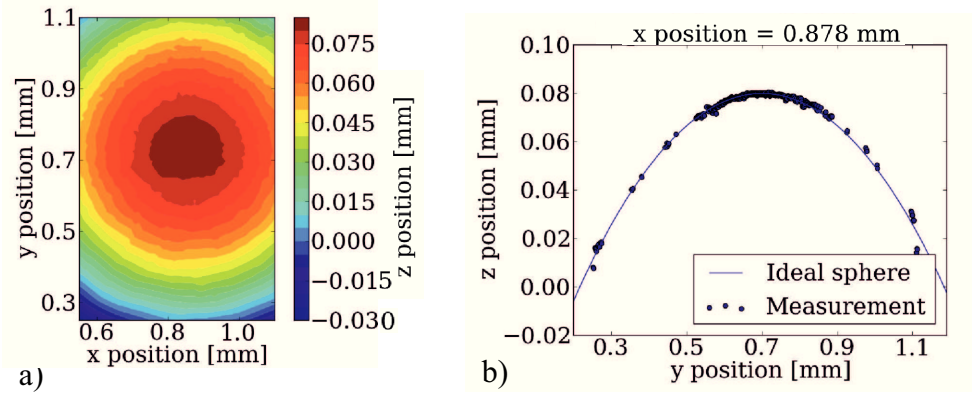


Figure 6.11: a) Measurement of the indenter topology, b) Typical cross-section of the indenter topology.

the circular pattern. To compare the topology with the ideal spherical shape, the data points on several cross-sections are compared to cross-sections of the corresponding ideal sphere. The data points on a typical cross-section and an ideal spherical shape are presented in figure 6.11b. Based on this typical cross-section, as well as other investigated cross-sections, it is concluded that there is no significant difference between an ideal sphere and the actual indenter shape. Therefore, it can be assumed that the indenter is an ideal sphere with a diameter of 3mm.
Using the instrumented indentation setup shown in figure 6.8, 560 indentation experiments are performed to obtain an extensive database to validate the indentation model. These indents are positioned on the 40 specimens according to the pattern shown in 6.8c. However, not every indentation experiment was successful, due to incorrect displacement measurements or simply human error during operation. Therefore, only 534 valid indentation curves are obtained which are presented in 6.12. From these indentation curves the indentation modulus is calculated using equation 3.9, with E_i=1050 [78], ν_i=0.2 [78] and ν=0.34 [79]. Furthermore, $A_c(h_c)$ is defined by the form of the indenter, which is an ideal sphere in this case. The mean value and standard deviation of the indentation moduli are shown in table 6.2. During sphere indentation experiments, the material is deformed in all three directions of the slightly anisotropic material. Hence, the resulting material response represents a kind of averaged response of the stiffness in transverse, longitudinal and thickness direction. Since the Youngs modulus in thickness is approximately equal to the Youngs modulus in transverse direction [79], it can be expected that the value of the indentation modulus falls between the values of the transverse and longitudinal macroscopic elastic modulus, which is confirmed by the results presented in table 6.2. Hence, it can be assumed that the indentation experiments are accurate enough to determine the local stress-strain behaviour from the indentation depth at several load levels during loading. These indentation depths are presented in table 6.3 in terms of the mean value and the standard deviation.

Table 6.2: Comparison of E_{IT} and the tension Young's moduli of aluminium 2024-T351.

E_{IT} [GPa]	Young's Modulus [GPa] transverse direction	Young's modulus [GPa] longitudinal direction
72.0 ± 1.5	70.3 ± 1.7	73.1 ± 1.3

Table 6.3: Mean value and standard deviation of the indentation depth of the indentation experiments performed on specimen of aluminium 2024-T351.

	Indentation depth	
Indentation load [N]	mean value [mm]	standard deviation [mm]
300	0.0310	0.0005
500	0.0472	0.0007
700	0.0624	0.0009
900	0.0771	0.0010
1100	0.0913	0.0011
1300	0.1052	0.0011
1500	0.1190	0.0012
1650	0.1294	0.0013
1850	0.1430	0.0013

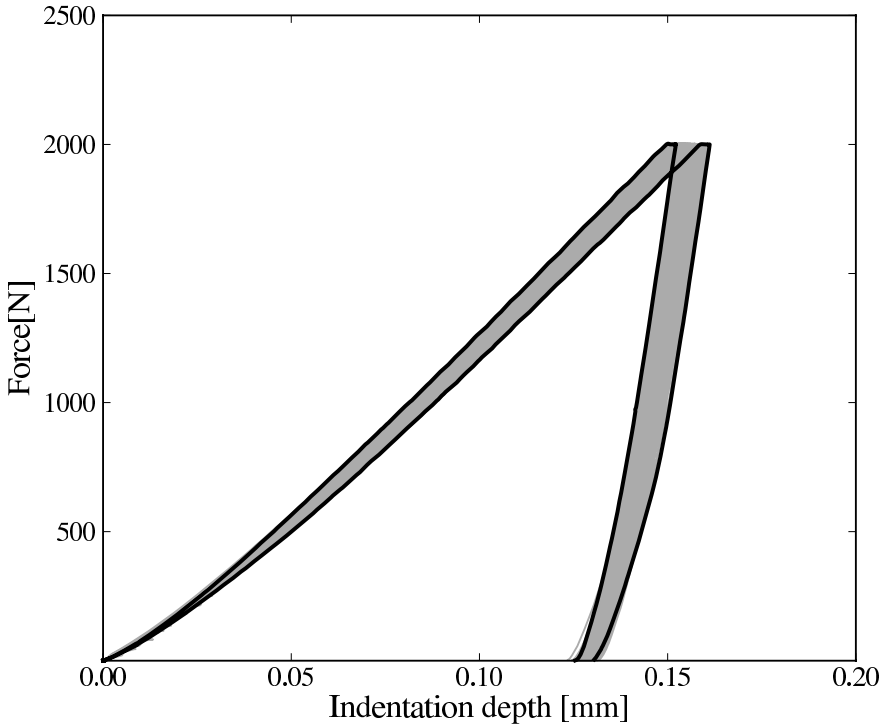


Figure 6.12: Illustration of the indentation curves of 534 indentation experiments in aluminium 2024-T351, the minimum indentation curve and the maximum indentation curve are highlighted in black.

Apart from the experimental results, it is important to determine the measurement uncertainties, because the measurement uncertainty of the indentation depth directly influences the decision whether a distance between simulation and experiment is significant. The measurement uncertainty of the load measurement system is smaller than 0.06% according to the calibration protocol of the device. Therefore, it can be concluded that the influence of the measurement uncertainty of the applied loads on the results is negligible. Consequently, it is assumed that the applied loads are measured exactly. The measurement uncertainty of the displacement measurement system is calculated using 6 calibration series. It is common practice to present measurement uncertainties in calibration protocols using a probability interval of 95% [26, 80]. The values of this probability interval are presented in table 6.4.

Table 6.4: Measurement uncertainty of the indentation measurements defined using the probability interval of 95%.

Displacement [mm]	Measurement uncertainty [%]
0.05	±2.42
0.10	±0.91
0.15	±1.23
0.20	±1.33
0.25	±1.22

6.2.2 Tensile experiments

The tensile experiments are performed using dog-bone shaped specimens (see figure 6.6) with a gauge section of 50 mm x 12.5 mm x 6 mm and a radius of 34 mm. These specimens are loaded in tension with a constant test velocity of 1 mm/min using an universal testing rig (Instron 4505). During the tensile experiments, the applied load is measured using a 100kN load cell and the specimen elongation is measured using a laser extensometer (Fiedler P-50) over a length of 46 mm centred at the gauge section of the specimen. Based on these two quantities, the engineering stress and strain are calculated using equations 3.3 and 3.4, see figure 6.13. For small strains,

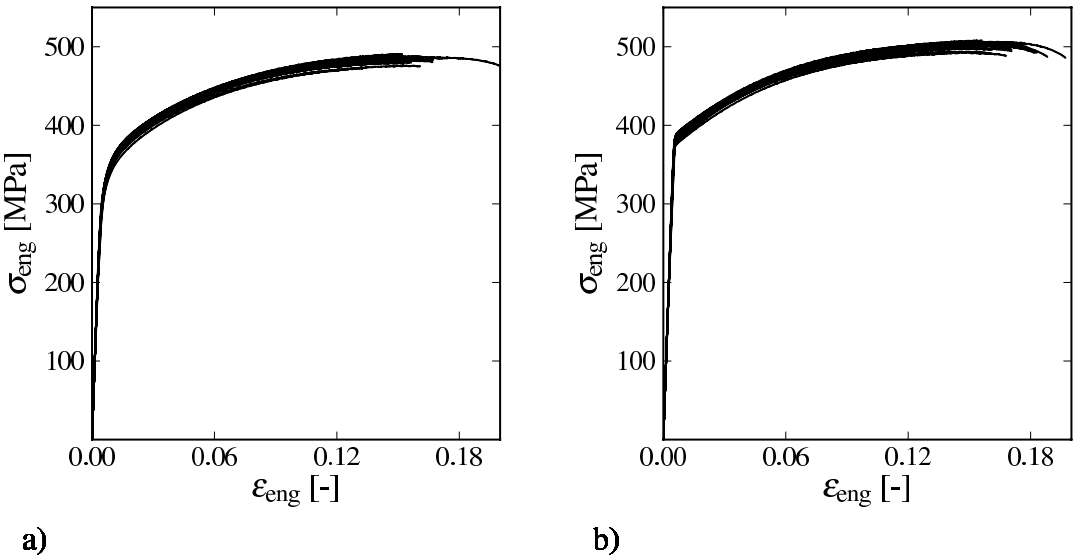


Figure 6.13: Tensile engineering stress - engineering strain behaviour of aluminium 2024-T351: a) Measured in transverse direction, b) Measured in longitudinal direction.

the stress-strain behaviour is linear. This linear elastic behaviour is expressed by means of the Youngs modulus, which is calculated using equation 3.5. The mean value and standard deviation of the Youngs modulus are presented in table 6.5 for the transverse and longitudinal direction of aluminium 2024-T351.

Table 6.5: Mean value and standard deviation the Youngs moduli of aluminium 2024-T351 in transverse and longitudinal direction.

Material parameter [Units]	2024 transverse	2024 longitudinal
E [GPa]	70.3 $\pm$1.7	73.1 $\pm$1.3

Although the tensile experiments are commonly evaluated using the engineering stress and strain, it is necessary here to calculate the true stress and strain behaviour, because the elastic-plastic material model of the sphere indentation model is defined using the true stress-strain behaviour. The true strain and the true stress value are calculated using equations 3.6 and 3.7. These quantities exhibit a monotonically increasing relation as shown in figure 6.14.
To calculate the uncertainty of engineering stress, engineering strain and Youngs modulus, the measurement uncertainties have to be propagated through equations 3.4, 3.3 and 3.5. Here, the

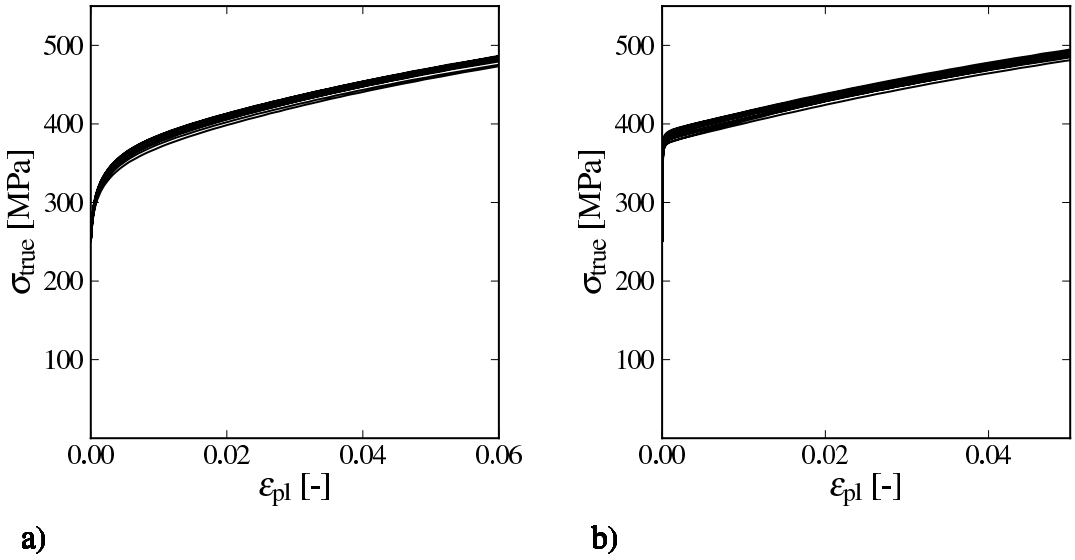


Figure 6.14: Tensile true stress - true plastic strain behaviour of aluminium 2024-T351: a) Determined in transverse direction, b) Determined in longitudinal direction.

uncertainties are calculated from the measurement uncertainties using the interval arithmetic according to equation 3.1. The measurement uncertainty of the applied load is a combination of the measurement uncertainty present in the measurement setup and the load cell, which are calibrated separately. Since the applied load varies during the experiments, it is necessary to determine the measurement uncertainty at several load levels. It is expected that the maximum relative uncertainty occurs at low loading. To calculate the measurement uncertainty of the applied loads, the measurement uncertainty of the load cell and the measurement setup have to be added together as presented in table 6.6. The values presented in table 6.6 are obtained from calibration protocols. It is beyond the scope of this work to explain the calibration procedures used to measure these measurement uncertainties. On basis of table 6.6, it is concluded that the maximum measurement uncertainty is present at 10% of the load cell capacity. At increased loading, the relative measurement uncertainty decreases as shown in table 6.6.
The elongation measurements are performed with a laser extensometer, whereof only the maximum measurement uncertainty is known. The maximum measurement uncertainty is a function of the measured displacements and it is specified by the device manufacturer in two classes. Namely, for elongations smaller than $300\,\mu$m the measurement uncertainty is equal to $0.6\,\mu$m while for elongations larger than $300\,\mu$m it is equal to 0.2% of the actual elongation.

Table 6.6: Measurement uncertainties of the tensile experiments.

Calibration load [kN]	Measurement setup [%]	Load cell [%]	Applied load [%]	Cross sect. area [%]	Eng. Stress [%]
10.0	-0.31 $\pm$ 0.65	$\pm$0.03	-0.31 $\pm$ 0.68	$\pm$ 0.2	-0.31 $\pm$ 0.88
20.0	-0.22 $\pm$ 0.21	$\pm$0.02	-0.22 $\pm$ 0.23	$\pm$ 0.2	-0.22 $\pm$ 0.43
30.0	-0.20 $\pm$ 0.16	$\pm$0.02	-0.20 $\pm$ 0.18	$\pm$ 0.2	-0.20 $\pm$ 0.38
40.0	-0.19 $\pm$ 0.19	$\pm$0.01	-0.19 $\pm$ 0.20	$\pm$ 0.2	-0.19 $\pm$ 0.40

To calculate the uncertainty of the engineering stress and Youngs modulus, it is necessary to determine the uncertainty of the cross sectional area. It is a combination of the geometrical variation due to manufacturing tolerances and the measurement uncertainty of the geometry measurements. The uncertainties of the geometry of the specimen in width and thickness direction are respectively $\pm 7\,\mu$m and $\pm 3\,\mu$m, which represent the 95% probability interval of scatter of the geometry of the specimen. This variation is measured using a device which has a maximum measurement uncertainty of $4\,\mu$m. Hence, the total uncertainty in width direction and thickness direction are respectively $\pm 11\,\mu$m and $\pm 7\,\mu$m. Using the nominal thickness of 6 mm and width of 12.5 mm, the uncertainty of the cross-sectional area is equal to $\pm 0.20\%$. Since the uncertainty in the cross sectional area is now known, the uncertainty of the engineering stress and Youngs modulus can be calculated using equation 3.1. Here, the measurement uncertainty in the Youngs modulus depends on the elongation of the specimen. Consequently, the uncertainty in the Youngs modulus is equal to $\pm 2.19\,\%$ for the used specimen gauge length, material and load range to calculate the Youngs modulus. The measurement uncertainties in the engineering stress are presented in the corresponding column of table 6.6.

6.3 Numerical model of the indentation tests

To simulate the indentation process under influence of uncertainties, a numerical model is used that consists of three parts as presented in figure 6.15. The deterministic model is used to simulate the indentation process as presented in figure 6.15a. This model requires, among other things, material parameters for the material model. To introduce scatter to the input parameters reflecting their scatter in reality, it is necessary to describe the scatter of the input parameters by means of a stochastic description as presented in figure 6.15b. Using the deterministic model and the stochastic description of the input parameters, a probabilistic model is created as illustrated in figure 6.15c. Basically, this stochastic model is a pre and post processing tool wrapped around the deterministic model to obtain scattering results by means of a Monte Carlo simulation. Using Monte Carlo simulation realisations, random values for the input parameters are generated according to their respective distribution function. For each deterministic simulation of the finite element model, the values of its input parameters are modified using random numbers which are based on the stochastic descriptions. This gives the stochastic results of the simulation as shown in figure 6.15c. In the following subsections the deterministic model, the stochastic description of the input parameters and the stochastic results of the model are presented.

6.3.1 Deterministic model

The deterministic model is a finite element model (FEM) created with Ansys14.0 [81], which simulates the indentation process in an elastic-plastic material. To create a deterministic model that fits its purpose, it is essential to identify the requirements put on the model by its intended use. In this case, the indentation model is created to determine the local hardening behaviour from indentation experiments via inverse parameter identification. Creating an indentation in the slightly anisotropic material (see figure 6.14) will result in one indention curve, but it is not possible to derive the anisotropic hardening behaviour in a unique manner from indents made with a sphere indenter as shown in figure 6.16a. Since the intended use of the model is identifying material parameters from sphere indents via inverse parameter identification, it is essential to simplify the material to an isotropic material to ensure a unique relation between the indentation depth and the flow stress as illustrated in figure 6.16b. Moreover, inverse parameter investigation is an optimisa-

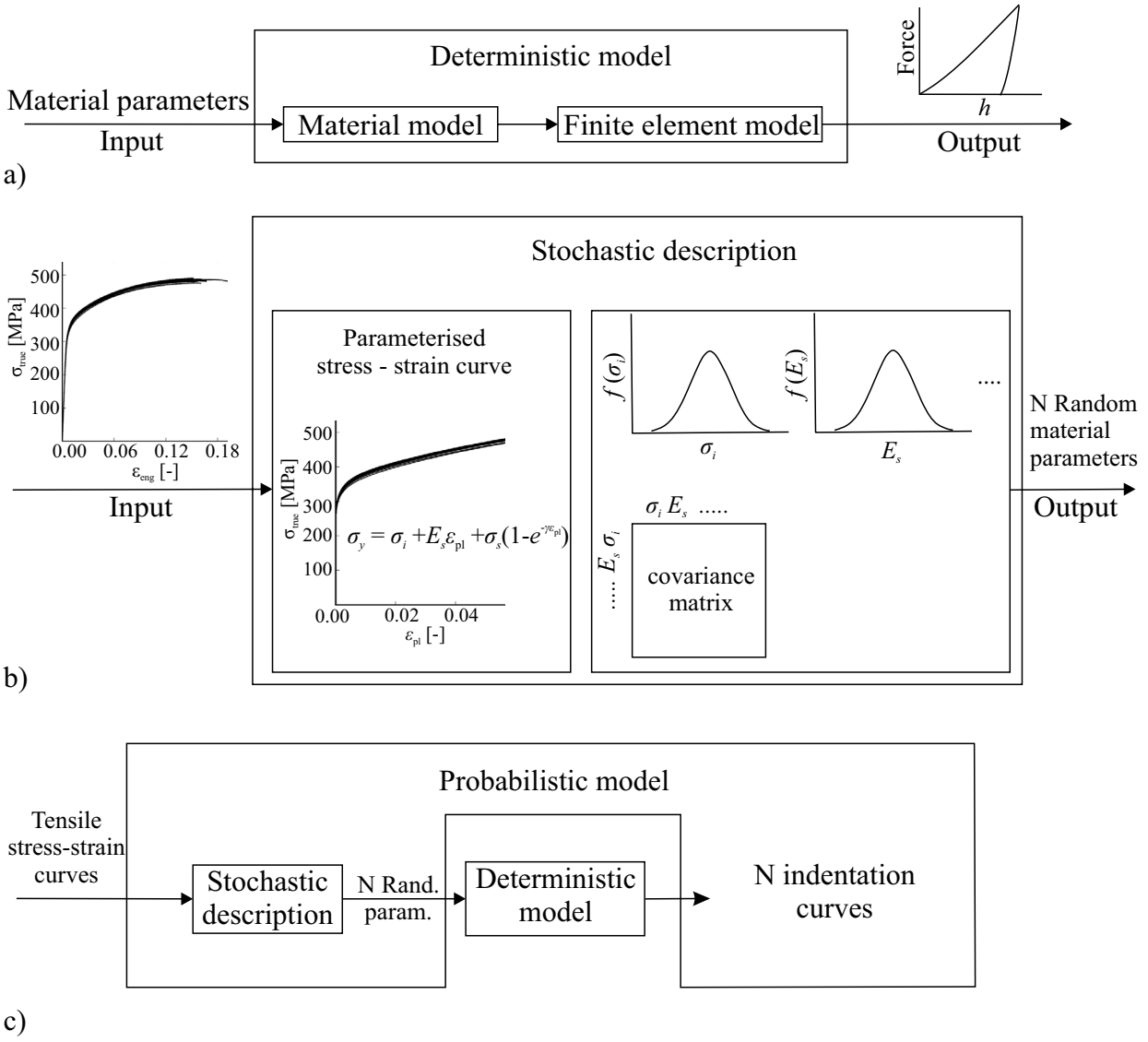


Figure 6.15: Flow chart of the stochastic sphere indentation model: a) The deterministic model, b) The stochastic description, c) The probabilistic model.

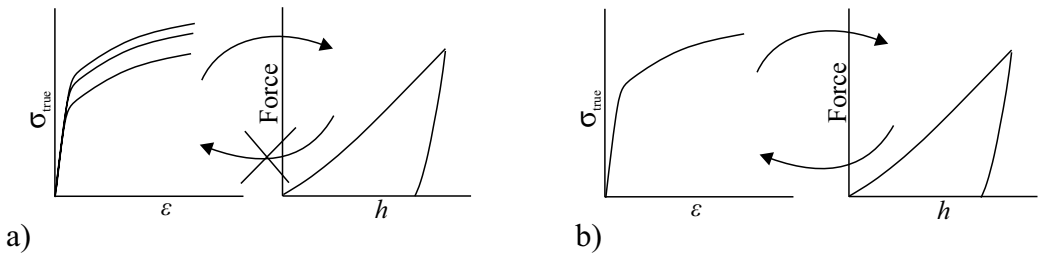


Figure 6.16: a) Illustration of the non-unique relation between the indentation depth and the anisotropic hardening behaviour of the material, b) Illustration of the unique relation between the indentation depth and the isotropic hardening behaviour of the material.

tion procedure, which requires a numerical efficient model to determine the local material properties quickly. Consequently, a rotational symmetric model is created, as presented in figure 6.17.
The material is modelled as isotropic elastic-plastic material with isotropic hardening. Since the material is modelled as isotropic material, a systematic error is made because the rolled plate of aluminium 2024-T351 is slightly anisotropic. Nevertheless, in context of sphere indentation simulation this error can be accepted, because the macroscopic anisotropy is relatively small (see table 6.5 and figures 6.13 and 6.14) and a unique relation must exist between the indentation depth and the flow stress for the intended use of the model, i.e. inverse parameter identification.
The finite element model used to simulate the indentations consists of three different parts made of different materials, namely a diamond indenter, a specimen and a steel specimen support as shown in figure 6.17. For this model quadratic elements (Ansys element name is 183 [82]) are used with an element size of 0.012 mm x 0.012 mm in the contact zone which is coarsened gradually outside the indentation zone. The load is applied to the topside of the diamond half sphere and the nodes on top of the diamond indenter are coupled to ensure a uniform displacement. The specimen and the specimen support are connected to each other using shared nodes while the specimen and the indenter are connected using contact elements with target and surface elements on both surfaces to ensure minimal penetration of the bodies. Moreover, the indenter is modelled as an ideal sphere, because no significant deviation has been found between the actual indenter topology and an ideal sphere as shown in figure 6.11. In addition, a frictionless contact is assumed, because the effect of friction is assumed to be small for such well polished surfaces. Furthermore, the coefficient of friction has a small influence on the indents for large indentation depths [34]. The behaviour of the contact elements is calculated using the augmented Lagrangian algorithm of Ansys14.0 with the default settings and initial gap closure [82]. Since this model is effectively a 2-dimensional model, it cannot represent the experimental boundaries of the specimen clamps. Therefore, the influence of this simplified symmetrical boundary condition is investigated during model verification.
Model verification has to be performed before validation, because it ensures that the model calculates the solution accurately. Verification of this model consist of a finite element size convergence study and an investigation of the boundary conditions. The finite element size convergence study is used to determine the optimal trade off between model accuracy and calculation efficiency. Based on table 6.7 and figure 6.18, it is concluded that the model with 3105 elements is the best

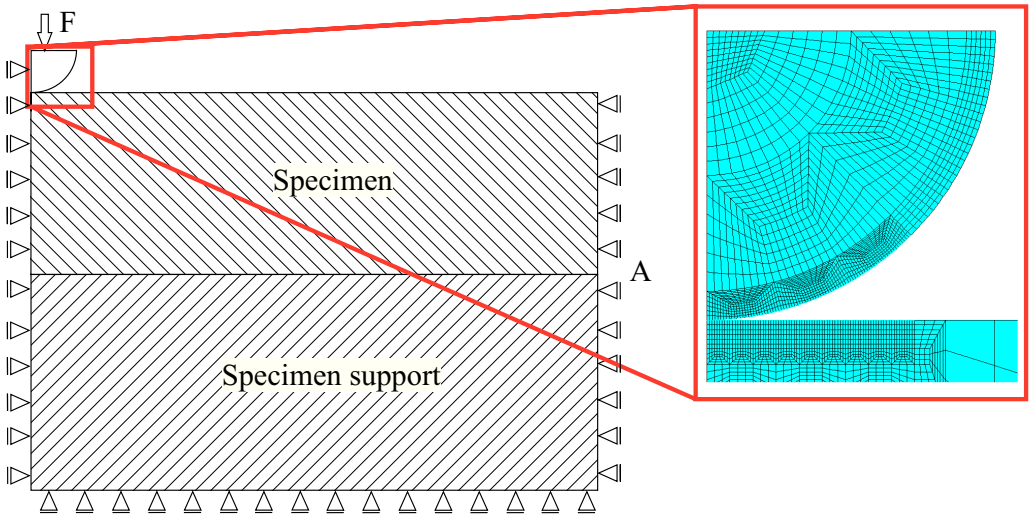


Figure 6.17: Schematic representation of the rotational symmetric finite element model.

compromise between computational efficiency and accuracy. It has a maximum model error of 0.24% compared to a model with 11124 elements (providing the reference solution), while being approximately 4.5 times faster.

Table 6.7: Indentation depth and error of the finite element model for different amounts of elements.

Elements	CPU time [s]	h_{max} at 600N [mm]	Error [%]	h_{max} at 1200N [mm]	Error [%]	h_{max} at 1800N [mm]	Error [%]
11124	1507	0.04440	-	0.07877	-	0.11099	-
8265	992	0.04432	0.18	0.07864	0.17	0.11097	0.02
5376	611	0.04426	0.32	0.07855	0.28	0.11089	0.09
3105	338	0.04434	0.14	0.07863	0.18	0.11072	0.24
1452	138	0.04426	0.32	0.07848	0.37	0.11047	0.47
417	44	0.04407	0.74	0.07790	1.11	0.10922	1.49

To estimate the influence of the simplified boundary condition on the indentation depth, both extreme boundary conditions (fixed in radial direction and free) are investigated. For aluminium 2024, the difference between both boundary conditions is 0.14% for the maximum indentation depth at 2000 N using the model with 11124 elements. Therefore, it is concluded that the boundary condition at edge A does not significantly influence the results. Since a fixed boundary condition in radial direction is closer to reality, this boundary condition is used for all following simulations. Consequently, all simulations related to validation of the indentation model are performed using the finite element model with 3105 elements and fixed boundary conditions at edge A.

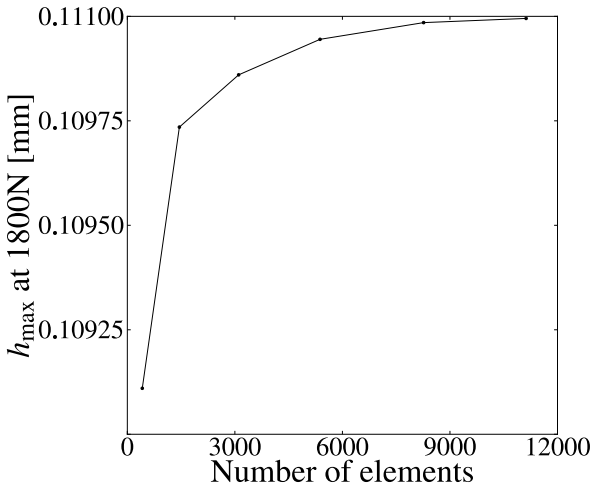


Figure 6.18: Simulated maximum indentation depth at 1800 N versus the number of finite elements.

6.3.2 Stochastic description of the random input parameters

The macroscopic hardening behaviour of aluminium 2024-T351 is slightly anisotropic as presented in figure 6.14. However for reasons mentioned in the previous section, it is necessary to simplify the material to an isotropic material. Although simplified, the material model should still represent the material accurately. Therefore, it is essential to estimate the hardening behaviour of the representative isotopic material. Since hardening is measured in the plastic zone, which can be up to 5.5x larger than the contact area [83], it is concluded from figure 6.19 that the plastic zone contains a sufficient number of grains. Since the plastic zone contains many grains, it is expected that the mean value of the local representative hardening behaviour is approximately equal to the mean value of the global representative hardening behaviour under equal loading conditions. During indentation, the material is loaded in all directions. Therefore, it is expected that the representative material hardening behaviour should fall between the individual hardening behaviour of the three material directions, namely transverse, longitudinal and thickness direction. Although the hardening behaviour in thickness direction is unknown, it is expected that the flow stress of the material in thickness direction is lower compared to flow stress in transverse direction on basis of the shape of the grains presented in figure 6.20a and figure 6.20b. Furthermore, it is known from the tensile experiments that the flow stress of the material is higher in longitudinal direction. Therefore, it is assumed that the hardening behaviour in transverse direction is an adequate rough estimate of the corresponding representative isotropic material.

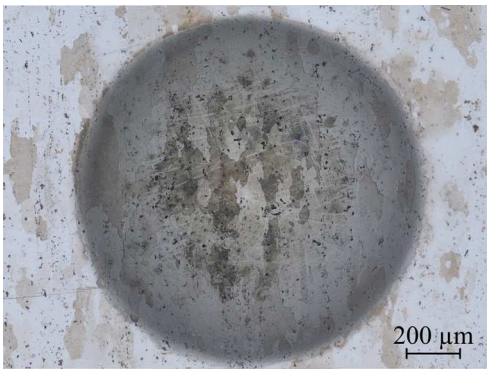
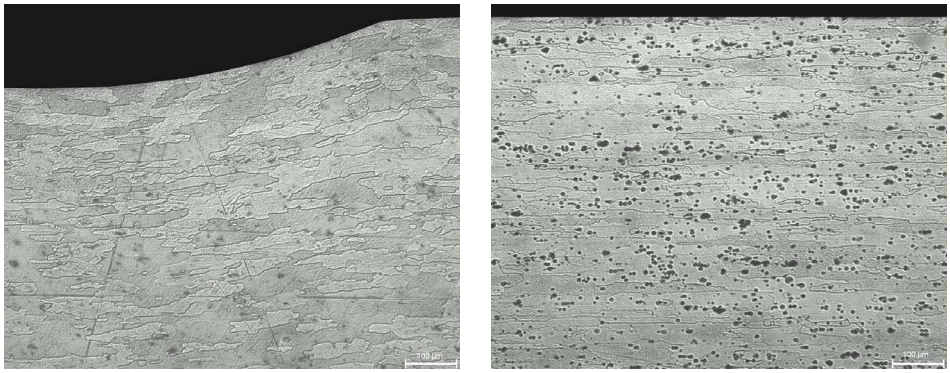


Figure 6.19: The size of a sphere indent relative to the size of the grains of aluminium 2024-T351.



a) b)

Figure 6.20: Shape and size of the grains of aluminium 2024-T351: a) Transverse direction, b) Longitudinal direction.

To parameterise the hardening behaviour of aluminium 2024-T351 the Voce hardening law [84] is used, which is an isotropic hardening model that fits the hardening behaviour of aluminium 2024-T351 (shown in figure 6.14) quite well. The Voce hardening law is presented in equation 6.1,

$$\sigma_y(\varepsilon_{pl}) = \sigma_i + E_s\,\varepsilon_{pl} + \sigma_s(1 - e^{-\gamma\varepsilon_{pl}}) \tag{6.1}$$

where ε_{pl} is the true plastic strain, σ_y is the flow stress of the material, σ_i is the initial yield stress, E_s is the slope of saturation, σ_s is the difference between initial yield stress and saturation stress and γ is the hardening saturation rate of the exponential term.
Using the optimisation routine "fmin" of the Python package SciPy, the Voce hardening model is fitted to the experimentally measured true stress plastic strain curves of aluminium 2024-T351 in transverse direction (shown in figure 6.14a). Minimisation of the difference between the experimentally measured flow stress and the flow stress based on the hardening law is performed using the following objective function presented in equation 6.2,

$$d_{\text{flow stress}} = \sum_{i=0}^{100} (\sigma_{exp,\varepsilon_{pl}} - \sigma_{y,\varepsilon_{pl}})^2, \qquad \varepsilon_{pl} = \frac{i}{10}\% \tag{6.2}$$

where $\sigma_{y,\varepsilon_{pl}}$ is the flow stress according to the hardening law at ε_{pl} plastic strain and $\sigma_{\text{exp},\varepsilon_{pl}}$ is the experimentally measured flow stress at ε_{pl} plastic strain. Furthermore, i is varied between [0, 70] with increments of 1. The mean value and standard deviation of the hardening law parameters resulting from this optimisation are presented in table 6.8. These values will be used for the stochastic description of the input parameters of the numerical model. Therefore, it is essential to verify whether the measurement uncertainty of the tensile experiments is less than the standard deviation of the hardening parameters. Since the measurement uncertainty is less than 1% (see table 6.6) and the scatter of the parameters in table 6.8 is larger than 1%, it is concluded that the tensile experiments are performed sufficiently accurate.

Table 6.8: Parameters of the material model for aluminium 2024-T351 in transverse direction.

Material parameter [Units]	E [GPa]	σ_i [MPa]	E_s [MPa]	σ_s [MPa]	γ [-]
	70.3 ±1.7	277 ±3.3	1520 ±29.4	107 ±2.6	174 ±7.3

In order to generate random values for the input parameters of the stochastic model, it is necessary to estimate or assume the distribution of the parameters of the material model and the correlation between the parameters must be determined. Here, a normal distribution is assumed for all parameters. The correctness of this assumption is verified using two goodness of fit hypothesis tests. One based on the Kolmogorov-Smirnov (K-S) test statistic and one based on the Anderson-Darling (A-D) test statistic, which are presented in subsection 3.3.2. Using a significance level of 5% no significant difference can be found between the empirical distribution function and the normal distribution function on basis of the K-S statistic as shown in table 6.9.

Table 6.9: Goodness of fit of the normal distribution to the material parameters.

Material	test	E	σ_i	E_s	σ_s	γ
2024 transverse	K-S	accepted	accepted	accepted	accepted	accepted
	A-D	rejected	accepted	accepted	accepted	accepted

Using the Anderson-Darling test statistic, the hypothesis test is only rejected for the Youngs modulus as presented in table 6.9. Since the A-D test statistic is sensitive for differences in the tails of the distribution while the K-S statistic is focused on the central part of the distribution, the difference must be present between the tails of the normal distribution function and the empirical distribution function. Even though a difference is found between the tail of the distribution of the Youngs modulus, it is decided to use the normal distribution for the parameters because it fits the central part of distribution well (see K-S test) and the tails of the distributions are of minor importance in this case. The assumed distributions are characterised by the mean value and the standard deviation listed in table 6.8. To express the correlation between the parameters of the material model that are calculated using the optimisation routine "fmin", the linear correlation coefficients are calculated with equation 3.14. The linear correlation matrix of aluminium 2024 in transverse direction is presented in table 6.10.

Table 6.10: Correlation matrix of the material parameters of 2024-T351 in transverse direction.

	E	σ_i	E_s	σ_s	γ
E	1.00	0.48	0.03	-0.17	0.22
σ_i	0.48	1.00	-0.08	-0.31	0.17
E_s	0.03	-0.08	1.00	-0.18	-0.03
σ_s	-0.17	-0.31	-0.18	1.00	0.79
γ	0.22	0.17	-0.03	0.79	1.00

To ensure that the random hardening curves based on these correlated material parameter distributions represent the measured hardening behaviour accurately, it is chosen to verify the stochastic description of the hardening behaviour. For this purpose, 100 random hardening curves are generated using the simple random sampling method presented in subsection 3.3.4. These randomly generated hardening curves and the experimentally measured hardening curves are presented in figure 6.21. Based on figure 6.21, it is concluded that the stochastic material model represents the measured stress - strain behaviour accurately, because the experimental results (open squares) and the numerical results (solid circles) match very well.

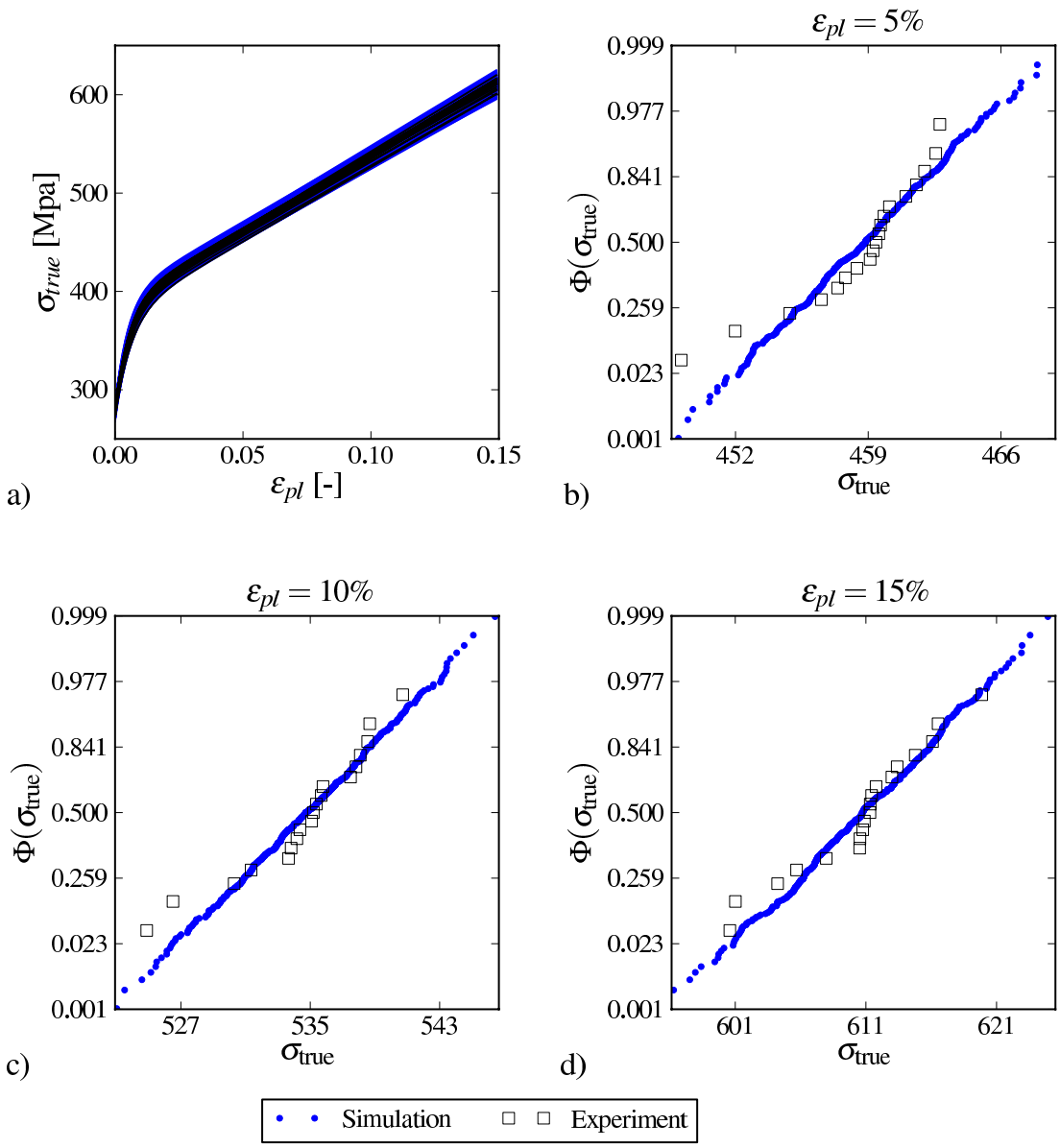


Figure 6.21: Distribution of stress-strain behaviour of 2024-T351 in transverse direction: a) True stress - true strain curves, b) Cumulative distribution of the true stress at 5% true plastic strain, c) Cumulative distribution of the true stress at 10% true plastic strain, d) Cumulative distribution of the true stress at 15% true plastic strain.

6.3.3 Results of the stochastic indentation simulation

Using the deterministic model and the stochastic parameter description presented in the previous subsections, the scatter of the indentation depth is calculated. For this purpose, 200 simulations are performed. The results of these 200 simulations are listed in table 6.11 by means of the mean value and the standard deviation of the indentation depth at several load levels.

Table 6.11: Mean value and standard deviation of the indentation depth of 200 indentation simulations at several load levels.

	Indentation depth	
Indentation load [N]	mean value [mm]	standard deviation [mm]
300	0.0315	0.0002
500	0.0476	0.0004
700	0.0629	0.0006
900	0.0778	0.0006
1100	0.0921	0.0007
1300	0.1061	0.0010
1500	0.1198	0.0009
1650	0.1301	0.0011
1850	0.1434	0.0011

6.4 Validation of the indentation model

To perform validation of the indentation model, the steps illustrated in figure 6.2 have to be executed. The first step of validation is discretisation to obtain the multivariate result space of the indentation curves during loading. This result space is used to determine the difference between the numerical and experimental indentation curves. To represent the indentation curve adequately, it is discretised into the following nine load levels: 300 N, 500 N, 700 N, 900 N, 1100 N, 1300 N, 1500 N, 1650 N and 1850 N.
The resulting indentation depths at these load levels are clearly correlated with each other due to the typical shape of the indentation curve, which is shown in figure 6.12. Using more discretisation steps leads to a more accurate representation of the indentation curve, but also leads to highly correlated results at more closely spaced load levels and thereby does not provide additional information that is significant. To demonstrate the effect of inappropriate discretisation on the measured distance between the numerical results and the experimental results as well as the decisions made during validation, the indentation loading curve is also discretised using only two load levels, namely 300 N and 1850 N. It is known in advance that only two load levels are insufficient to represent the nonlinear indentation curve accurately.
Next, the distance between the result spaces is initially measured using both the Image Based Distance Measure (IBDM) and the Mahalanobis distance. The Image Based Distance Measure and the Mahalanobis distance are used because these methods are efficient to detect differences between multivariate result spaces as explained in chapter 5. Both methods are used for the first iteration of the optimisation step of the validation routine and only the method with better performance is used to fully optimise and complete the iterations towards an agreement of the numerical and experimental results.
When the distance is measured using IBDM, it is necessary to specify two settings of the method as presented in figure 6.3, namely the kernel width and the number of pixels used to create the images. The number of pixels used for IBDM is specified in section 6.1, while the kernel width can be defined using either equation 3.28 or the measurement uncertainty. Here, both methods are used to demonstrate the influence of the kernel width on the measured distance and to determine which method is more suited for this problem. When the kernel width is defined using the measurement uncertainty, the standard deviation of the Gaussian kernel is equal to $1/1.96$ of the measurement

uncertainty, because the measurement uncertainties specified in table 6.4 represent the central 95% of the normally distributed measurement uncertainties. In case the Gaussian kernel width is defined using equation 3.28, the number of data points is equal to the number of data points in the smallest sample. The standard deviation is equal to the standard deviation of the sample with the greatest scatter. Comparing the standard deviations given in table 6.11 to those given in table 6.3 reveals that the experimental results contain most scatter, thus the scatter of the experimental results is used to define the kernel width. To avoid confusion among these two approaches to define the kernel, the Gaussian kernel based on equation 3.28 is named kernel k_{th} and the Gaussian kernel based on the measurement uncertainties is named kernel k_{unc}.
After computation of the distance between the numerical results and the experimental results, a randomisation hypothesis test is performed to determine whether this distance is significant. The randomisation hypothesis tests are performed using 999 randomisations and a significance level of 5% (see also section 5.2). Since a hypothesis test always depends on the size of the samples, the influence of the number of data points on validation is determined by varying the number of samples. For this purpose, subsets are drawn from the numerical and experimental results. The subsets contain the first 200, 50 or 20 data points of the complete data set. For further details concerning the randomisation hypothesis test, the reader is referred to section 6.1.
In the next subsections, the distance between the numerical results and the experimental results is determined using the methods listed above. Furthermore, the influence of the settings of IBDM on the measured distance and its significance level are investigated to determine which of these settings is best for validation.

6.4.1 Influence of the settings of IBDM on the measured distance

To investigate the influence of the kernel width on the distance, the distances measured using kernel k_{th} and k_{unc} based on the 9-dimensional result space are compared to each other and the distances based on the 2-dimensional result space are compared to each other. To visualise the 9-dimensional results space, two 2-dimensional projections of the result space are exemplarily shown in figure 6.22 whereas all 36 projections are included in appendix D. These figures show the 534 experimentally measured data points and the 200 simulated data points. The resulting measured distances are listed in table 6.12 for the previously defined result spaces and sample sizes.

Table 6.12: Distance between simulation and experiment determined with IBDM.

N_{sim}	N_{exp}	Image Based Distance Measure (ρ_m)			
		2-dimensional result space		9-dimensional result space	
		Kernel k_{th}	Kernel k_{unc}	Kernel k_{th}	Kernel k_{unc}
200	534	0.472	0.738	0.698	0.757
200	200	0.492	0.769	0.714	0.778
50	50	0.683	0.806	0.755	0.746
20	20	0.813	0.844	0.802	0.732

Concerning the distance between the 2-dimensional results (simulation and experiment) measured with IBDM, it is observed from table 6.12 that the correlation coefficient ρ_m based on kernel k_{th} is always smaller than the correlation coefficient ρ_m based on kernel k_{unc}. This is caused by the difference between the kernel width of kernel k_{th} and kernel k_{unc}. A small correlation

coefficient means that the distance is large while a correlation coefficient of 1.0 means that there is no difference between the samples. Since kernel k_{th} becomes smaller with increasing number of data points per sample, the difference between the kernels increases with increasing number of data points. Regarding the 9-dimensional result space, the following observations are made from table 6.12:

1. The correlation coefficient ρ_m based on kernel k_{th} is smaller than the correlation coefficient ρ_m based on kernel k_{unc} for $N \geq 200$.
2. The correlation coefficient ρ_m based on kernel k_{th} is larger than the correlation coefficient ρ_m based on kernel k_{unc} for $N \leq 50$.
3. The correlation coefficient ρ_m based on kernel k_{th} decreases with increasing number of experimental data points.
4. The correlation coefficient ρ_m based on kernel k_{unc} increases with increasing number of experimental data points up to 200 data points whereafter it decreases.

Based on observations 1 and 2, it can be concluded that the kernels are approximately equally wide for samples containing approximately 50 data points, because the distance is nearly identical for samples containing 50 data points. Therefore, it is recommended to use samples containing at least 50 data points. This amount of data points is also consistent with the necessary amount of data points based on the rule of thumb presented in [18]. Furthermore, it can be concluded from observation 4 that the scatter of the experimental results is larger than that of the numerical results, because the distance between simulation and experiment increases (the value of IBDM decreases with increasing distance) when more than 200 experimental measurements are used. This behaviour is not observed for kernel k_{th} (see observation 3), because the width of kernel k_{th} decreases with the number of data points by definition, which automatically decreases the correlation coefficient (increases the distance) between the experimental and numerical results.
To investigate the influence of the amount of discretisation steps used to approximate the indentation loading curve, the measured distance between the 2-dimensional results is compared to the distance between the 9-dimensional results. Comparing the correlation coefficients reported in table 6.12 for 20 data points reveals that the distance measured between the 2-dimensional results is smaller than the distance measured between the 9-dimensional results (smaller correlation coefficient) regardless of the used kernel. Consequently, it can be concluded that the difference at some intermediate loading levels must be greater than the distance at the two load levels analysed for this 2-dimensional case. Therefore, it is essential to also measure the distance between the indentation depths at intermediate loading. Since the 2-dimensional discretisation of the experimental and numerical indentation curves does not contain the results at intermediate load levels, it is insufficiently accurate to determine the distance between the indentation curves. Moreover, for large samples as well as small indentation depths, the kernel k_{th} is much smaller than the measurement uncertainty. Therefore, it is concluded that IBDM should be based on the measurement uncertainties (kernel k_{unc}) to measure a meaningful distance.

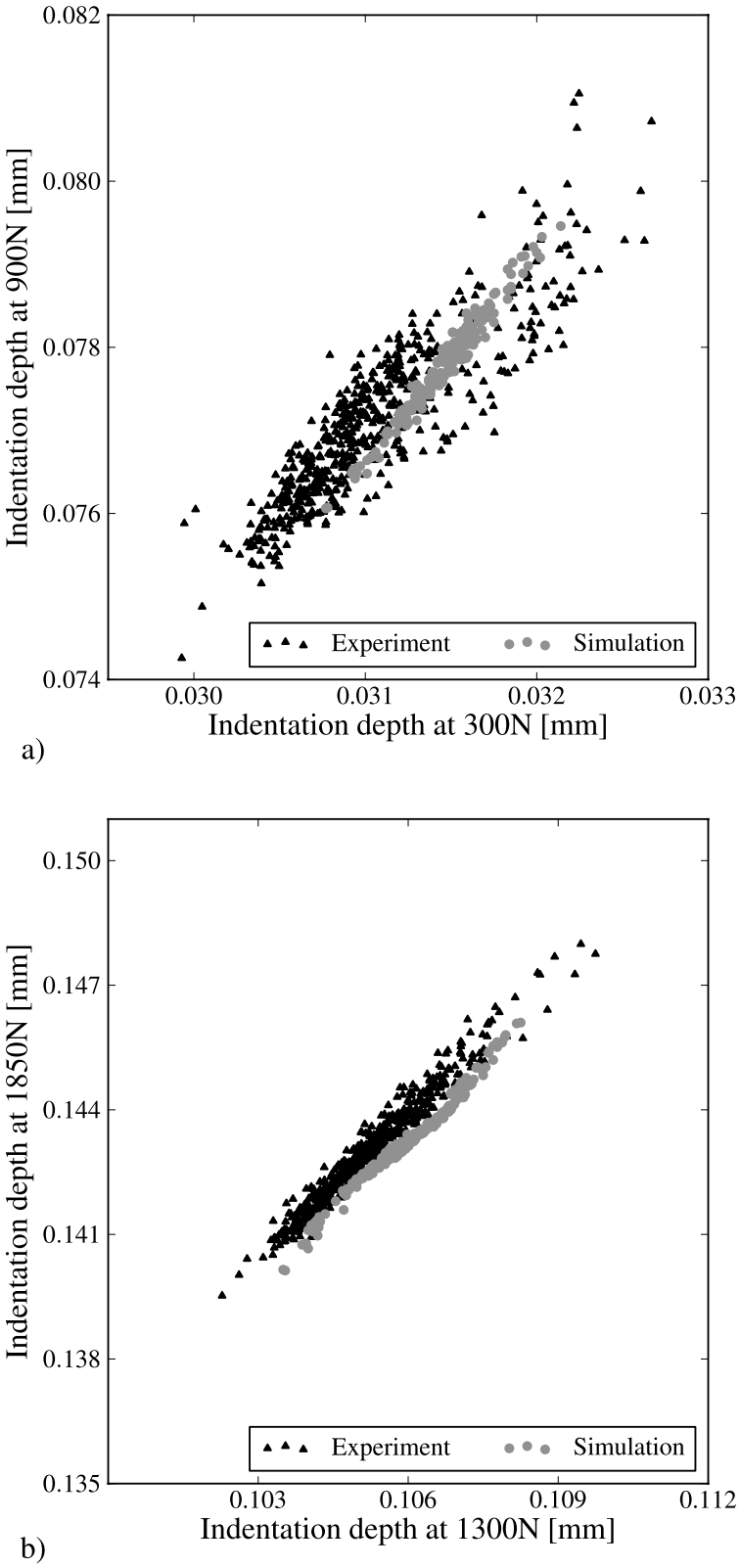


Figure 6.22: Two dimensional projections of the indentation depth of the 200 simulations and the 534 experiments at several load levels: a) Indentation depth at 300 N versus indentation depth at 900 N, b) Indentation depth at 1300 N versus indentation depth at 1850 N.

6.4.2 Mahalanobis distance between simulation and experiment

The distance between the simulated and experimental results is also measured with the Mahalanobis distance as presented in table 6.13. As can be seen in table 6.13, the Mahalanobis distance decreases with decreasing number of data points per sample. This tendency is in agreement with the tendency of the distances measured with IBDM based on kernel k_{th} (see table 6.12). Furthermore, it is observed that the difference between the 9-dimensional result space is always larger than the difference between the 2-dimensional result space. This observation confirms the earlier stated conclusion that the 2-dimensional discretised results space does not represent the indentation loading curve sufficiently accurate as expected.

Table 6.13: Mahalanobis distance between numerical and experimental results.

N_{sim}	N_{exp}	Mahalanobis distance	
		2-dimensional result space	9-dimensional result space
200	534	222.89	2920.59
200	200	155.38	2326.47
50	50	15.06	472.11
20	20	3.77	340.40

6.4.3 Significance level of the measured distances

To determine whether the measured distances are significant as well as to decide which distance measure is more appropriate for this case, it is required to perform randomisation hypothesis tests based on the null hypothesis $f_{\text{exp}}(\boldsymbol{x}) = f_{\text{sim}}(\boldsymbol{x})$. Further details concerning the randomisation hypothesis testing method can be found in section 6.1 and in subsection 3.4.4. Each randomisation hypothesis test results in one p-value ϕ which is compared to the significance level α to determine if the distance is significant. When the p-value ϕ is smaller than α, the null hypothesis is rejected. The smaller the p-value, the more confident one can be that the samples are different. The results of the hypothesis tests for all measured distances are presented in table 6.14. Based on table 6.14, the following observations are made:

1. The p-value of IBDM based on kernel k_{unc} is smaller than or equal to the corresponding p-value of IBDM based on kernel k_{th}.
2. All p-values of IBDM are smaller than 0.05
3. The p-value of the 2-dimensional Mahalanobis distance is larger than or equal to the corresponding p-value of the 9-dimensional Mahalanobis distance.
4. The p-value of the 2-dimensional Mahalanobis distance is larger than 0.05 when the samples contain 20 data points.
5. The p-value of the 9-dimensional Mahalanobis distance is smaller than 0.05.
6. The p-value of IBDM is smaller than the p-value of the Mahalanobis distance when the 2-dimensional samples contain 20 data points.

It can be concluded from observation 1 that the difference is detected easier with IBDM when based on kernel k_{unc}, because a larger p-value leads to a lower probability to reject the null hypothesis. Since all distances measured with IBDM resulted in a p-value smaller than the significance

Table 6.14: P-values of the distances measured between simulation and experiment with either IBDM or the Mahalanobis distance.

N_{sim}	N_{exp}	2-dimensional result space			9-dimensional result space		
		ϕ(IBDM)		$\phi({D_M}^2)$	ϕ(IBDM)		$\phi({D_M}^2)$
		Kernel k_{th}	Kernel k_{unc}		Kernel k_{th}	Kernel k_{unc}	
200	534	0.001	0.001	0.001	0.001	0.001	0.001
200	200	0.001	0.001	0.001	0.001	0.001	0.001
50	50	0.001	0.001	0.002	0.001	0.001	0.001
20	20	0.049	0.023	0.192	0.025	0.008	0.001

level of 0.05 (see observation 2), it can be concluded that the experimental results and the numerical results do not belong to the same distribution.
Concerning the Mahalanobis distance, it is concluded on basis of observation 4 that there is insufficient evidence to reject the null hypothesis using the 2-dimensional Mahalanobis distance for samples containing 20 data points, because the p-value is larger than 0.05. However, on basis of observation 5, it is concluded that there is sufficient evidence to reject the null hypothesis on basis of the 9-dimensional result space. Furthermore, it can be concluded from observation 3 that a 2-dimensional discretisation of the result space is insufficient to validate the indentation model, because the p-values of the Mahalanobis distance for samples containing 20 data points are significantly larger for a 2-dimensional discretisation compared to the 9-dimensional discretisation. Therefore, it is concluded on basis of observations 3, 4 and 5 that the difference is significant, i.e. the numerical results are not consistent with the experimental results.
Since IBDM and the Mahalanobis distance (normalised mean value difference) are both significant, at least a mean value difference must be present between the numerical results and the experimental results. Therefore, it is required to adjust parameters of the numerical model using the optimisation loop presented in figure 6.1. This optimisation loop uses one of the distance measure for the objective function. Thus, it is required to decide which distance measurement method is more appropriate for this case. On basis of observation 6 and observation 4, it is concluded that IBDM is more effective to detect the difference between the numerical and experimental results, because a possibly incorrect outcome of the hypothesis test is avoided. Consequently, it is chosen to use the Image Based Distance Measure based on kernel k_{unc} to optimise the simulation model, i.e. to minimise the distance between the numerical results and the experimental results.

6.5 Adjustment of the mean values of the input parameters of the numerical model

To improve the numerical model, some of the parametric input values of the numerical model have to be adjusted within their respective parameter range. From the unverified assumptions (the description of the local hardening behaviour), it is chosen to adjust the mean values of the hardening parameters first, because the mean hardening behaviour has the largest influence on the indentation depth. This is only true when the material under the indenter is significantly plastically deformed, which is the case for the performed experiments. Adjusting the mean value of the hardening parameters such that the distance between the numerical results and the experimental results becomes

minimal leads to an improved numerical model. This improved model has to be checked again for its validity.
To determine whether the optimisation routine yields meaningful results, the distance between simulation and experiments is measured again using a new set of experimental data. To avoid additional experiments, it is chosen to define subseries of the 534 experimental results. Up to four series of 20, 50 or 200 data points are drawn sequentially from the 534 measurements, where each measurement can be in one series only. The first series of experiments is used for the optimisation routine while the other series are used to verify the results of the optimisation routine. As explained in the previous section, the distance between these subseries of experimental results and the numerical results is measured using IBDM based on kernel k_{unc} using a 9-dimensional discretised result space. The measured distances between the subsamples of the experimental and numerical results are presented in table 6.15.

Table 6.15: Original distance between the numerical results and different samples taken from the experimental results.

Number of data points	IBDM based on kernel k_{unc}			
	Sample 1	Sample 2	Sample 3	Sample 4
200	0.778	0.692	-	-
50	0.746	0.929	0.723	0.309
20	0.732	0.665	0.903	0.768

6.5.1 Minimising the distance between simulation and experiment

To minimise the distance between the numerical results and the experimental results, an optimisation routine is used which is illustrated in figure 6.23. This optimisation routine changes the mean values of the parameters of the Voce hardening law σ_i, E_s, σ_s and γ within their respective ranges, which are presented in table 6.16.

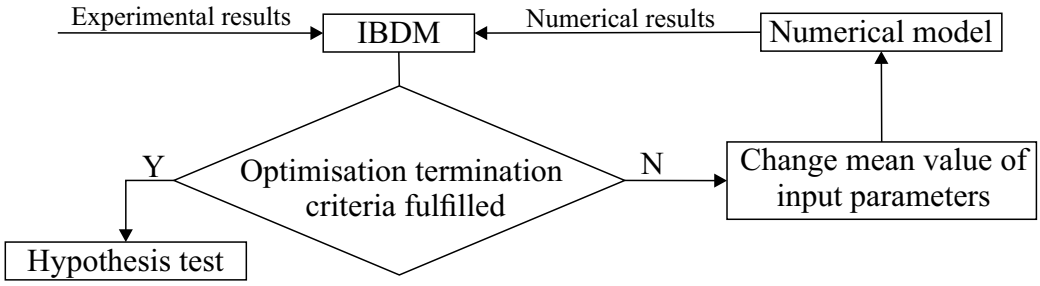


Figure 6.23: Flowchart of the optimisation loop used to adjust the mean values of the hardening parameters of the numerical model.

Table 6.16: Range of the optimisation parameters.

Parameter	σ_i [MPa]	E_s [MPa]	σ_s [MPa]	γ [-]
Minimum value	280	1000	95	160
Maximum value	320	1200	120	190

Since the derivatives of IBDM with respect to the input parameters are not available to the optimisation routine and a robust optimisation routine is desired, it is chosen to use the adaptive response surface method of the software OptiSLang 3.2 to minimise the difference between simulation and experiment. For details concerning the adaptive response surface method, the reader is referred to subsection 3.4.5. In the following, only the settings of the algorithm are presented. For this specific problem, the response surface of the adaptive response surface method is generated using the moving least square algorithm with the allowance to use previous support points (within the complete response space) to improve the possibly highly nonlinear response surface function. The optimisation loop is terminated when at least 3 iterations are performed and one of the following conditions is satisfied:

- the Image Based Distance Measure changes less than 0.1% compared to the previous iteration
- the input parameters change less than 0.1% compared to the previous iteration
- 150 designs are generated (sometimes necessary to limit the computation time)

To determine the optimal values of the optimisation parameters for 200, 50 and 20 data points, it would be required to perform three optimisations, which is very time consuming. However, in the used optimisation software, it is possible to reevaluate the previously performed optimisation for smaller samples using a new objective function without additional calculations. Consequently, a reevaluation is not a real optimisation and can only determine which of the evaluated designs would be the most optimal design. Since the changes in the objective function are rather small in this case, it is expected that it is possible to find an "optimum" near the true optimum without additional computational effort using the numerical efficient reevaluation routine. For further details about the reevaluation routine, the reader is referred to the manual of OptiSLang [67].
The results of the optimisation and reevaluation process are presented in table 6.17. It is obvious from table 6.17 that the optimal parameter set is identical for all sample sizes, which indicates that the samples represent the experimentally measured mean indentation behaviour approximately equally well. To visualise the achieved distance reduction, one projection of the 36 possible projections of the original numerical model and the optimised numerical model are presented in figure 6.24. All other projections of the optimised numerical model are included in appendix D. Comparing figure 6.24a (original numerical results, i.e. before optimisation) to figure 6.24b indicates that the distance between the numerical results and the experimental results has been reduced by adjustment of the numerical model.

Table 6.17: Optimal mean values of the parametric material model of aluminium 2024-T351.

	Adjusted model parameter			
Nr. data points	σ_i [MPa]	E_s [MPa]	σ_s [MPa]	γ [-]
200	297	1150	108	166
50	297	1150	108	166
20	297	1150	108	166

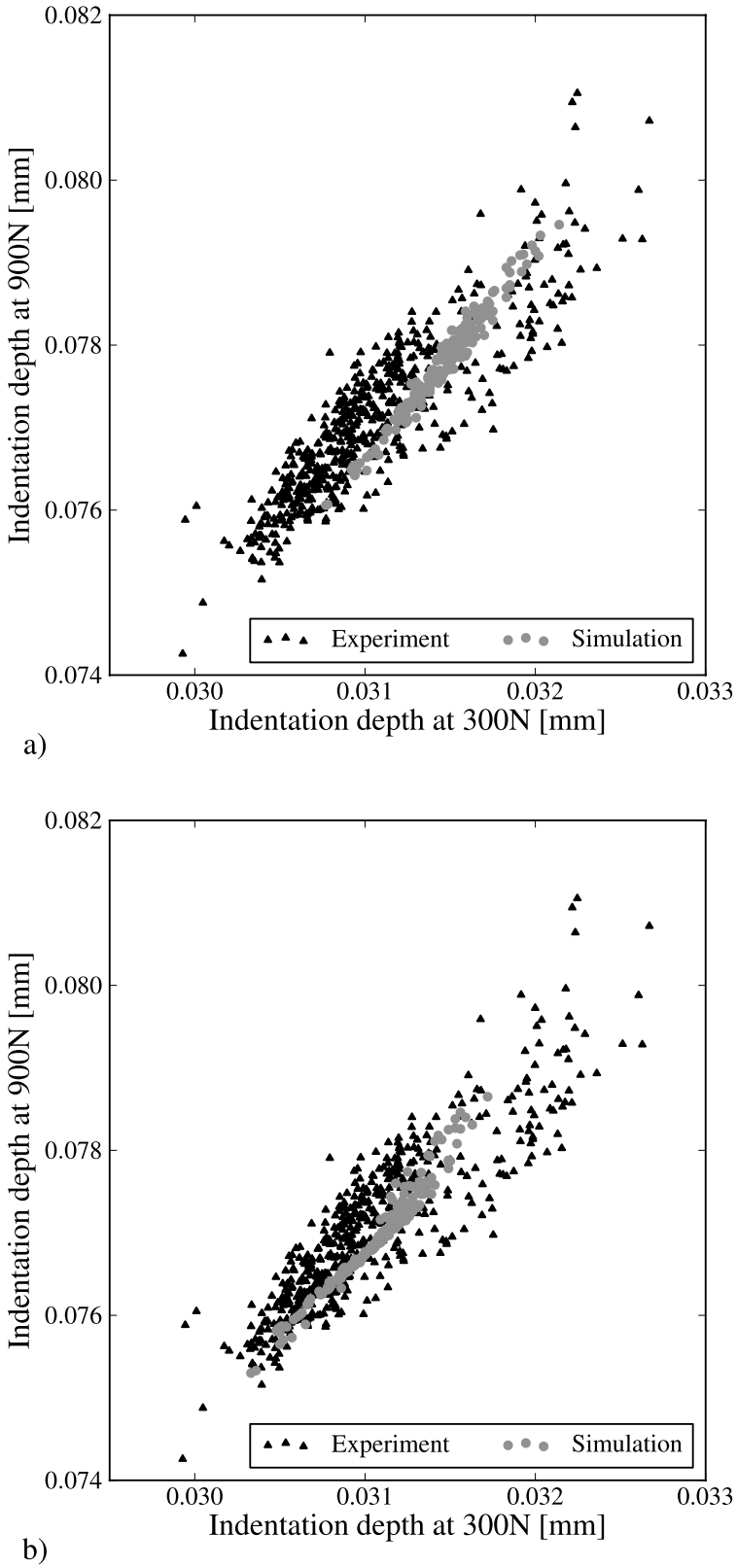


Figure 6.24: Two dimensional projection of the indentation depth at 300 N versus the indentation depth at 900 N of the 200 simulations and 534 experiments: a) Original numerical model versus experiment, b) Optimised numerical model versus experiment.

To ensure that the adjusted material hardening behaviour is still within a reasonable range from the macroscopic hardening behaviour, the identified hardening behaviour is compared to the mean hardening curves of aluminium 2024-t351 in transverse and longitudinal direction, which are shown in figure 6.25. It can be observed from figure 6.25 that the hardening behaviour described by the optimised parameters is within range of the macroscopically measured hardening behaviour. Therefore, it is concluded that the results of the optimisation routine are plausible.

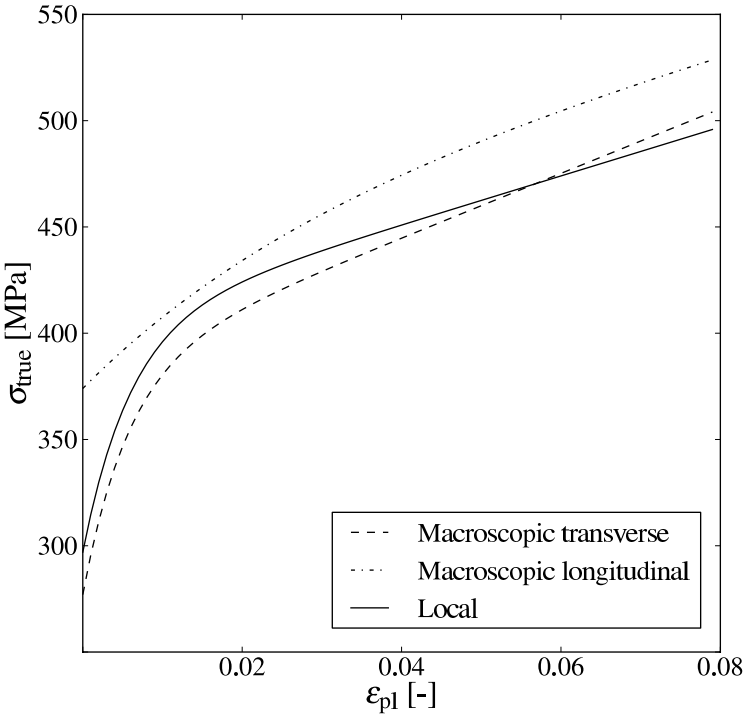


Figure 6.25: Macroscopic and local hardening behaviour of aluminium 2024-T351.

6.5.2 Measured distance between optimised numerical model and experiments

The measured distances between the experimental results and the results of the optimised numerical model are presented in table 6.18 together with the originally measured distances. Here, a larger value means a smaller distance. As mentioned before, experimental sample 1 is used to adjust the numerical model via the optimisation loop whereas sample 2 to 4 contain new experimental results which are used to verify that the optimisation loop produced meaningful results. From table 6.18, the following observations are made:

1. The distance between the results of the optimised numerical model and the experimental results has been reduced for samples containing 200 data points, because the correlation coefficient of IBDM increases.
2. The distance between the results of the optimised numerical model and the experimental results has been reduced for most samples containing 50 or 20 data points.
3. The distance between the results of the optimised numerical model and the experimental results has not been reduced for sample 2 containing 50 data points.

4. The distance between the results of the optimised numerical model and the experimental results has not been reduced for sample 3 containing 20 data points.

It could be concluded from observation 1 that the optimisation improved the numerical model, because the distance between the numerical results and experimental results has been reduced when the samples contain 200 data points. However, the opposing conclusion could be drawn from observation 3 and 4. To explain these two opposing conclusions, it is essential to realise that all samples are drawn by chance from the underlying distribution. Since all samples are drawn by chance, there is a chance that two samples match each other even though the underlying distributions are not identical. An example of such a sample could be sample 3 containing 20 data points, because the distance increases (decreasing correlation coefficient) for this sample while it decreases for all other samples containing an equal amount of data points. Since the probability to draw an unrepresentative subsample from the experimental results decreases with an increasing number of data points per sample, it can be concluded that observations 2, 3 and 4 are a consequence of randomly drawing subsamples. Therefore, it is concluded on basis of observations 1 and 2 that the model adjustment has improved the numerical model.

Table 6.18: The distance between the numerical results and the experimental results for different experimental samples based on IBDM.

Nr. data points	Numerical model	Sample 1	Sample 2	Sample 3	Sample 4
200	Original	0.778	0.692	-	-
	Optimised	0.944	0.902	-	-
50	Original	0.746	0.929	0.723	0.309
	Optimised	0.929	0.841	0.925	0.712
20	Original	0.732	0.665	0.903	0.768
	Optimised	0.888	0.902	0.639	0.923

6.5.3 Significance level of the measured distance

To determine whether the measured distances between the optimised numerical results and the experimental results are significant, it is necessary to perform randomisation hypothesis tests based on the null hypothesis $f_{\text{exp}}(\boldsymbol{x}) = f_{\text{sim}}(\boldsymbol{x})$. As before, each hypothesis test results in one p-value where a p-value smaller than 0.05 means that the null hypothesis can be rejected. The smaller the p-value, the more confident one can be that the null hypothesis is rejected correctly. The p-values of these hypothesis tests are presented in table 6.19. From this table the following is observed:

1. For samples containing 50 or 200 data points, all p-values based on the optimised numerical model are smaller than 0.05.
2. For samples containing 200 data points, the p-values based on the optimised numerical model are equal to the p-values based on the original numerical model.
3. For samples containing 20 data points, the p-values based on sample 1, 2 and 4 of the optimised numerical model containing 20 data points are larger than 0.05

Table 6.19: The p-values of the hypothesis tests based on IBDM for different samples.

Nr. data points	Numerical model	Sample 1	Sample 2	Sample 3	Sample 4
200	Original	0.001	0.001	-	-
	Optimised	0.001	0.001	-	-
50	Original	0.001	0.015	0.001	0.001
	Optimised	0.016	0.001	0.018	0.001
20	Original	0.008	0.001	0.191	0.002
	Optimised	0.185	0.079	0.002	0.177

Based on observation three, it is concluded that the distance between the numerical results and the experimental results is insignificant for this number of data points, because the null hypothesis $f_{\text{exp}}(\boldsymbol{x}) = f_{\text{sim}}(\boldsymbol{x})$ could not be rejected for 3 out of 4 samples. This does not mean there is no difference, because each hypothesis test has the possibility to make a type II error. When the amount of data points is increased, the probability to not reject the null hypothesis when the underlying distributions are different decreases. Consequently, the null hypothesis is rejected correctly more often for larger samples, which is supported by observations one and two.
For this work, a numerical model is called valid when either the distance is not stochastically significant or when the distance between any experimental data point and the nearest simulated data points is smaller than the combined uncertainty in the numerical model and the experiments. The first condition has been met for some samples containing 20 data points. However, the probability is not zero that the outcome of the hypothesis test is actually a type II error. Therefore, the second criterion is tested by comparing the minimum distances between the 534 experimental data points and their nearest neighbour of the 200 numerical data points to the uncertainty present in the numerical and experimental results.

6.5.4 Distance versus measurement uncertainty

To determine whether any numerical data point is in range of the experimental data points with respect to the uncertainty in the numerical results and the experimental results, the minimum distance between the numerical data points and the experimental data points is normalised using the combined uncertainty. Since the uncertainties are always expressed with respect to the result quantities, the combined uncertainty is obtained by summation of the numerical and experimental uncertainty as illustrated in figure 6.26. The experimental ellipse is defined using $1\,\sigma$ or $2\,\sigma$ of the measurement uncertainties, where the values of $2\,\sigma$ at several indentation depths are presented in table 6.4. The numerical error is equal to the numerical error due to size of the finite elements plus the error due to the boundary conditions. Using the values for these two errors presented in subsection 6.3.1, a value of 0.38% is obtained for the numerical error.
To calculate the normalised minimum euclidian distance between the experimental data points and the simulated data points, equation 6.3 is used,

$$d_{\text{unc},j} = \frac{\sum_{i=1}^{9} \left(\min\left(|\boldsymbol{x}_{\text{sim},i} - x_{\text{exp},ij}|\right)\right)^2}{\sum_{i=1}^{9} {u_i}^2} \tag{6.3}$$

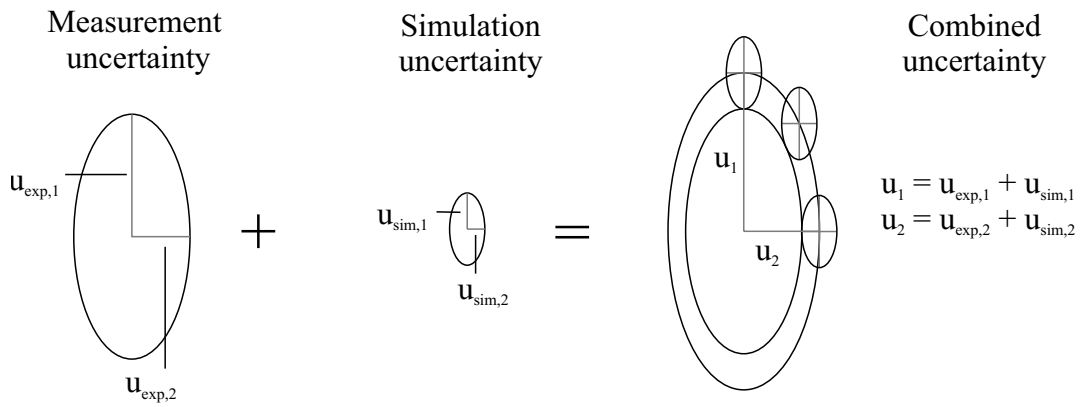


Figure 6.26: Definition of the combined uncertainty between numerical results and experimental results.

where $\boldsymbol{x}_{\mathrm{sim},i}$ is the result vector of the i-th result quantity of the numerical results, $x_{\mathrm{exp},ij}$ is the j-th data point of the i-th result quantity of the experimental results and u_i is the combined uncertainty of the i-th result quantity. The combined uncertainty is the uncertainty of the numerical model added to the experimental uncertainty. The discrete cumulative density function of the normalised distance based on the $1\,\sigma$ interval of the measurement uncertainties is presented in figure 6.27a whereas the discrete cumulative density function of the normalised distance based on the $2\ \sigma$ interval of the measurement uncertainties is presented in figure 6.27b. Here, $F(d_{\mathrm{unc}})$ is the discrete cumulative distribution function of d_{unc}. It is shown in figure 6.27 that not all experimental data points are within range of acceptance of the numerical results, because the curves contain normalised distances larger than 1.0. Nevertheless, the amount of numerical data points within range of acceptance of the experimental data points has increased significantly compared to the original numerical model as presented in figure 6.27a. Therefore, it is concluded that this optimisation can be interpreted as a kind of first step in the correct direction.

This conclusion is supported by the decreased distance measured with IBDM based on 200 data points (see table 6.18), but the p-values presented in table 6.19 indicate that there is still a significant difference between the numerical results and the experimental results, which is also supported by figure 6.27. This diagram shows that some experimental data points are not near any numerical data point, because the normalised distance is greater than 1.0. As apparent in figure 6.24, this difference is caused by a difference in the amount of scatter present in the results. Up to this point it was assumed that the scatter of the local material properties are identical to the scatter of the globally measured material properties, which are presented in table 6.8. However, the amount of scatter of local material properties is always larger than the scatter of the global material properties, because the global material properties represent the average of the local material properties. According to the central limit theorem [20, 40, 57, 58], an averaged material property (global) always exerts a smaller scatter than the local material property, because the larger deviations from the mean value are averaged out. Therefore, the amount of scatter of the local material input parameters should be scaled in accordance with the corresponding zone of influence activated during different types of experiments (tensile experiments, indentation experiments). This scaling is presented in the next section.

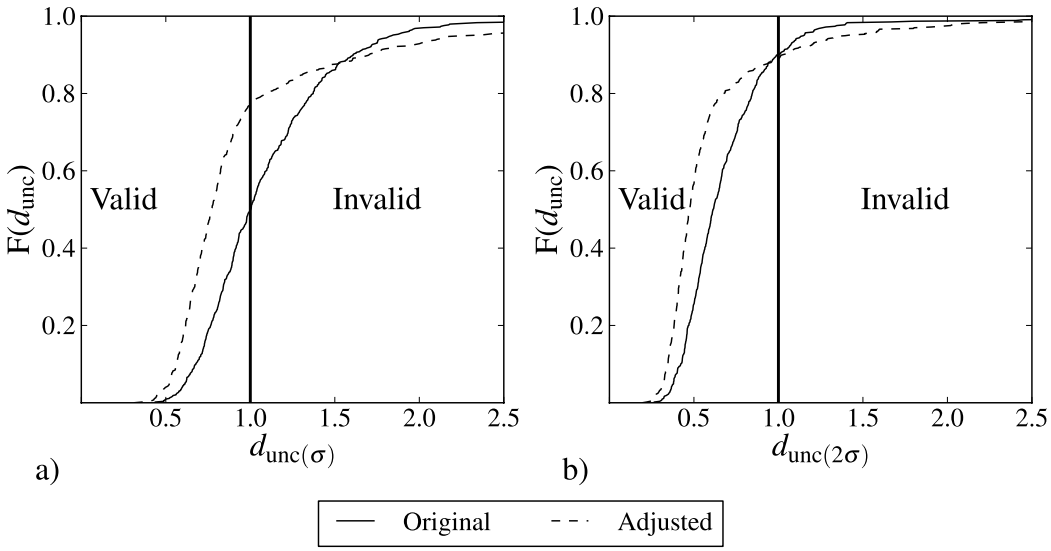


Figure 6.27: Discrete cumulative density function of the minimum distance between the experimental data points and the numerical data points, normalised using the combined uncertainty: a) Combined uncertainty based on the 1σ interval of the measurement uncertainties, b) Combined uncertainty based on the 2σ interval of the measurement uncertainties.

6.6 Adjustment of the variance of the input parameters of the numerical model

As recognised in the previous section, the amount of scatter of the experimental results is larger than the scatter of the numerical results. One likely source for this deviation is the fact that the local hardening behaviour exerts a larger scatter than the averaged global hardening behaviour taken from the tensile experiments. To investigate wether the remaining difference between the numerical model and the experiments could be caused by this effect, it is required to quantify this effect based on the zone of influence of the sphere indentation experiments and the tensile experiments. Here, the zone of influence is the domain where hardening and plasticisation take place.

6.6.1 Scaled variance of the hardening behaviour

Since the size of the zone of influence is not exactly known for both experiments, it will be estimated using the upper bound and the lower bound of the size of the zone of influence. The lower bound of the size of the zone of influence is defined by the averaged slip line area in the plastically deformed zone. For the upper bound, the size of the zone of influence is defined as the representative volume of plastically deformed material. In the following, the upper and the lower bound of the zone of influence are quantified for the tensile experiments and the indentation experiments.
The volumetric zone of influence of the indentation experiment is equal to the size of the plastic zone, which is calculated using the deterministic model based on the material parameters listed in table 6.17. The volume of the plastic zone in this finite element model is equal to $15.06\,\mathrm{mm}^3$ as shown in figure 6.28. For tensile specimen plasticisation occurs in the entire gauge length of the

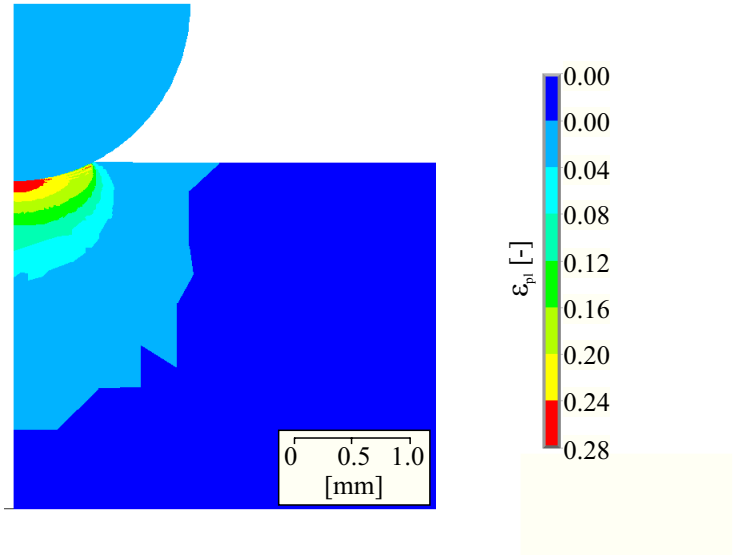


Figure 6.28: Finite element based illustartion of the plastic strain in aluminium 2024-T351 when a load of 2000 N is applied to the sphere indenter.

specimen when the yield stress is reached. However, the plastic strain at an equivalent loading level of the indentation experiments would be 8% [34], which causes necking in the tensile specimen. Since the zone where necking occurs is not exactly known, it is assumed that the maximum length of the specimen where necking occurs is equal to the thickness of the specimen. Hence, the zone of influence is equal to 6 x 6 x 12.5 = 450 mm^3. Therefore, the ratio of the zones of influence based on the upper boundary is equal to 29.88.
Next, the averaged slip line area of the tensile specimen and the sphere indentation are quantified to quantify the lower bound of the difference between both zones of influence. The averaged slip line area of a tensile specimen A_t is the cross sectional area under 45 degrees in thickness direction averaged over the parallel zone of the tensile specimen [30], which is calculated using equation 6.4,

$$A_t = \frac{1}{l_0}\int_{y=0}^{l_0} A_s(y)\,dy = \frac{1}{l_0}\int_{y=0}^{l_0} \sqrt{2}A\,dy \tag{6.4}$$

where $A_s(y)$ is the slip line area at position y of the specimen (see figure 6.29a) and l_0 is the gauge length of the tensile specimen. Since the cross-sectional area A is constant over the gauge length, A_t is equal to $\sqrt{2}A$ which is equal to 106 mm^2. To calculate the averaged slip line area for a sphere indentation, the plastic zone under the indenter is first simplified to be a half sphere [85, 86, 87]. Subsequently, the averaged slip line area of a sphere indentation experiment A_{SIE} is calculated using equation 6.5,

$$A_{\mathrm{SIE}} = \frac{1}{2r_p}\int_{x=-r_p}^{r_p}\int_{\theta=0}^{2\pi}\int_{y=-(x\cos(\phi)+\sqrt{R^2-x^2})}^{0} \frac{1}{\cos(\phi)}\partial y\partial\theta\partial x \tag{6.5}$$

where r_p is the radius of the plastic zone and ϕ is the bottom angle of the cone, which is illustrated in figure 6.29b. The angle ϕ of the cone at which the slip lines occur varies between 0 and 90 degrees due to the multiaxial stress state in the plastic zone. Therefore, it is necessary to consider all bottom angles from 0 to 90 degrees. Solving equation 6.5 for any angle ϕ between 0 and 90 degrees results in the same result, which is illustrated for 45 degrees in figure 6.29b, for 90 degrees in figure 6.29c and for 0 degrees in figure 6.29d. Assuming that the plastic zone is a half sphere,

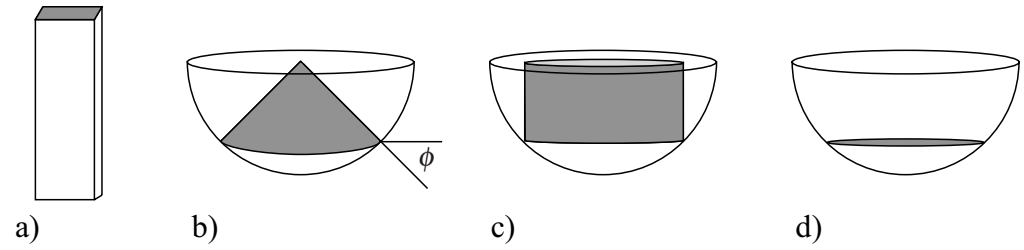


Figure 6.29: a) Slip line area under 45 degrees in a tensile specimen, b) The area of a possible slip line at $\phi = 45^\circ$ under a sphere indent, c) The area of a possible slip line at $\phi = 90^\circ$ under a sphere indent (a hollow cylinder), d) The area of a possible slip line at $\phi = 0^\circ$ under a sphere indent.

the radius of the plastic zone r_p is calculated from the volume of the plastic zone and it is equal to 1.93 mm. Hence, the averaged slip line area based on equation 6.5 is equal to 7.8 mm^2. Consequently, the ratio of the averaged slip line area (tensile experiment versus indentation experiments) is equal to 13.59.

By means of the central limit theorem, it is possible to scale properties using the ratio of the zone of influence if these properties are based on a summation. For stiffnesses, there exist models that can be described using a summation to calculate the global stiffness based on local stiffnesses, e.g. the model of Voigt-Reuss-Hill [38, 88, 89]. Since the variance of the stiffness scales identically for the model of Voigt and Reuss, there is no need to differentiate between the models of Voigt and Reuss. To calculate the stiffness according to the model of Voigt, equation 6.6 is used where [38, 88],

$$E_{\mathrm{g}} = \sum_{i=1}^{n} E_{l,i} \frac{A_i}{A_t}, \qquad A_t = \sum_{i=0}^{n} A_i \tag{6.6}$$

A_i, A_t and $E_{l,i}$ are respectively the size of the i-th local area, the total area and the local stiffness of the i-th local area as shown in figure 6.30. Assuming A_i is constant, equation 6.6 simplifies to equation 6.7,

$$E_{\mathrm{g}} = \sum_{i=1}^{n} \frac{E_l}{n} \tag{6.7}$$

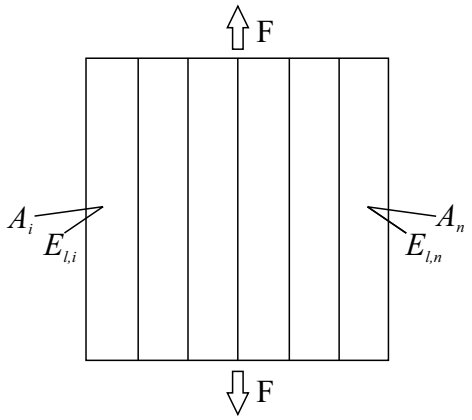


Figure 6.30: Illustration of the model of Voigt.

where E_l is the local stiffness and E_g the global stiffness. According to the central limit theorem, the variable E_g is normally distributed if sufficiently independent local stiffnesses E_l are averaged and E_l is an independent and identically distributed variable that is drawn from a distribution with expected values μ with and σ. More precisely, the central limit theorem states that the difference between the sample average E_g and its expected value μ when multiplied with $\sqrt{n}$ is approximately normally distributed with mean value zero and variance σ^2. Consequently, the variable E_g is close to normally distributed with mean value μ and variance σ^2/n regardless of the distribution of the variable E_l. This consideration only applies to stiffnesses or compliances. Hence, the hardening behaviour must be expressed as instantaneous stiffness. The instantaneous stiffness of the material is calculated by differentiating the flow stress (equation 6.1) with respect to the plastic strain as presented in equation 6.8.

$$\frac{\partial \sigma_y}{\partial \varepsilon_{pl}} = E_s + \sigma_s \gamma e^{-\gamma \varepsilon_{pl}} \tag{6.8}$$

Using the previously calculated lower and upper bound of the ratio of the zone of influence, the standard deviation of the global instantaneous hardening behaviour (tensile experiments) is $\sqrt{13.59}$ to $\sqrt{29.89}$ times smaller than the local measured hardening behaviour according to the central limit law. Here, the lower bound is calculated using the tensile-to-indentation ratio of average slip line area while the upper bound is calculated using the tensile-to-indentation ratio of the volumetric zone of influence. Scaling equation 6.8 with these factors on the left side and the right side of the equation gives the local standard deviation of the parameters of the Voce hardening law. Consequently, only E_s and σ_s are scaled as presented in table 6.20. Using the mean values of the hardening parameters given in table 6.17 and the scaled standard deviations of the hardening parameters given in table 6.20, 200 simulations are preformed to obtain the numerical results based on the scaled hardening behaviour.

Table 6.20: Original and scaled standard deviation of the Voce hardening law parameters for aluminium 2024-T351.

		Standard deviation of model parameter			
Scaling method	Scaling factor	σ_i [MPa]	E_s [MPa]	σ_s [MPa]	γ [-]
none	1.0	3.3	29.4	2.6	7.3
Average slip line area	$\sqrt{13.59}$	3.3	108.5	9.6	7.3
Volumetric zone of influence	$\sqrt{29.89}$	3.3	153.8	13.6	7.3

6.6.2 Distance between numerical results and experimental results

The distance between the numerical results and the experimental results is measured using IBDM for both scaled hardening models for up to 4 samples with 20, 50 or 200 data points per sample. The results are presented in table 6.21, where a value of 1.0 for IBDM indicates that no difference is present between simulation and experiment while 0.0 indicates that simulation and experiment are completely different. In addition, the results of the mean value optimised model presented in table 6.18 are included in table 6.21 to make it easier to compare the scaled numerical models to the optimised numerical model of section 6.5. From table 6.21, the following can be observed:

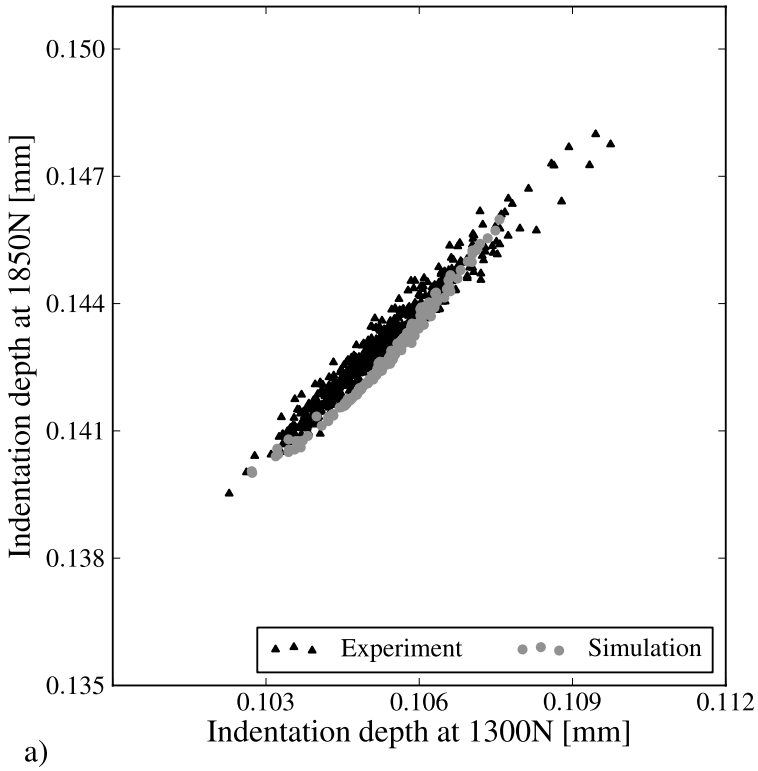
1. For samples containing 200 data points, scaling the hardening behaviour using the averaged slipe line area resulted in a decreased distance between the numerical results and the experimental results.

2. For samples containing 200 data points, scaling the hardening behaviour using the volumetric scaling method resulted in a larger distance between the numerical results and the experimental results compared to results based on the average slip line area.
3. For most samples containing 20 or 50 data points, scaling the hardening behaviour resulted in a larger distance between the numerical results and the experimental results.
4. For sample 4 containing 50 data points, scaling the hardening behaviour decreased the distance between the numerical results. In this case, the smallest distance is obtained using the average slip line area scale factor.
5. For sample 3 containing 20 data points, scaling the hardening behaviour decreased the distance between the numerical results. In this case, the smallest distance is obtained using the average slip line area scale factor.

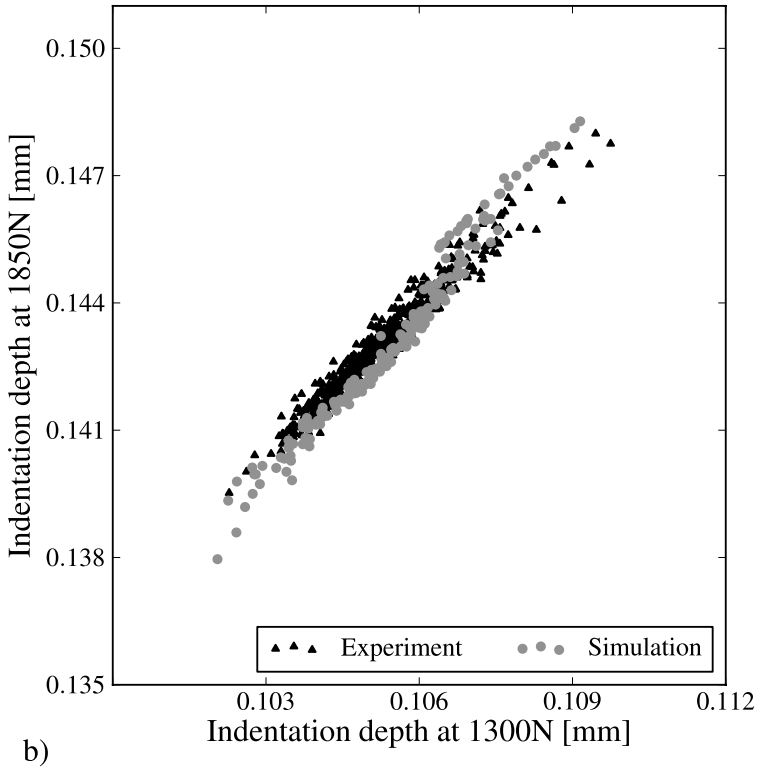
Based on observations 1 and 2 as well as figures 6.31 and 6.32, it is concluded that the local variance of the hardening behaviour is better approximated using the area method than the volumetric method because this scaling method resulted in the smallest distance between the numerical results and the experimental results. Furthermore, observation 3, 4 and 5 indicate that the scatter of the numerical results is still different from the experimentally observed scatter. As shown in figures 6.31 and 6.32, scaling hardening parameters changed the numerical cloud mainly in one direction without widening the numerical cloud regardless of the scaling method. Therefore, one could conclude that there is another difference between the numerical model and the experiments, which is still not considered in the validation model. To determine the plausibility of the above conclusion that there is another important difference, the null hypothesis $f_{\text{exp}}(\boldsymbol{x}) = f_{\text{sim}}(\boldsymbol{x})$ is tested using a randomisation hypothesis test. When there is another superior difference, the distance between the scaled numerical model and the experiments should be significantly different.

Table 6.21: Distance between the scaled numerical model and the experimental results according to IBDM for different numerical models and experimental samples.

Nr. data points	Numerical model	Sample 1	Sample 2	Sample 3	Sample 4
200	Optimised	0.944	0.902	-	-
	Area scaled	0.944	0.941	-	-
	Volume scaled	0.906	0.922	-	-
50	Optimised	0.929	0.841	0.925	0.712
	Area scaled	0.868	0.757	0.883	0.795
	Volume scaled	0.775	0.646	0.806	0.776
20	Optimised	0.888	0.902	0.639	0.923
	Area scaled	0.812	0.727	0.739	0.780
	Volume scaled	0.789	0.683	0.696	0.729



a)



b)

Figure 6.31: Two dimensional projection of the indentation depth at 1300 N versus the indentation depth at 1850 N of the 200 simulations and 534 experiments: a) Optimised numerical Model versus experiment, b) Numerical model of which the hardening parameters are scaled using the average slip line area versus experiment.

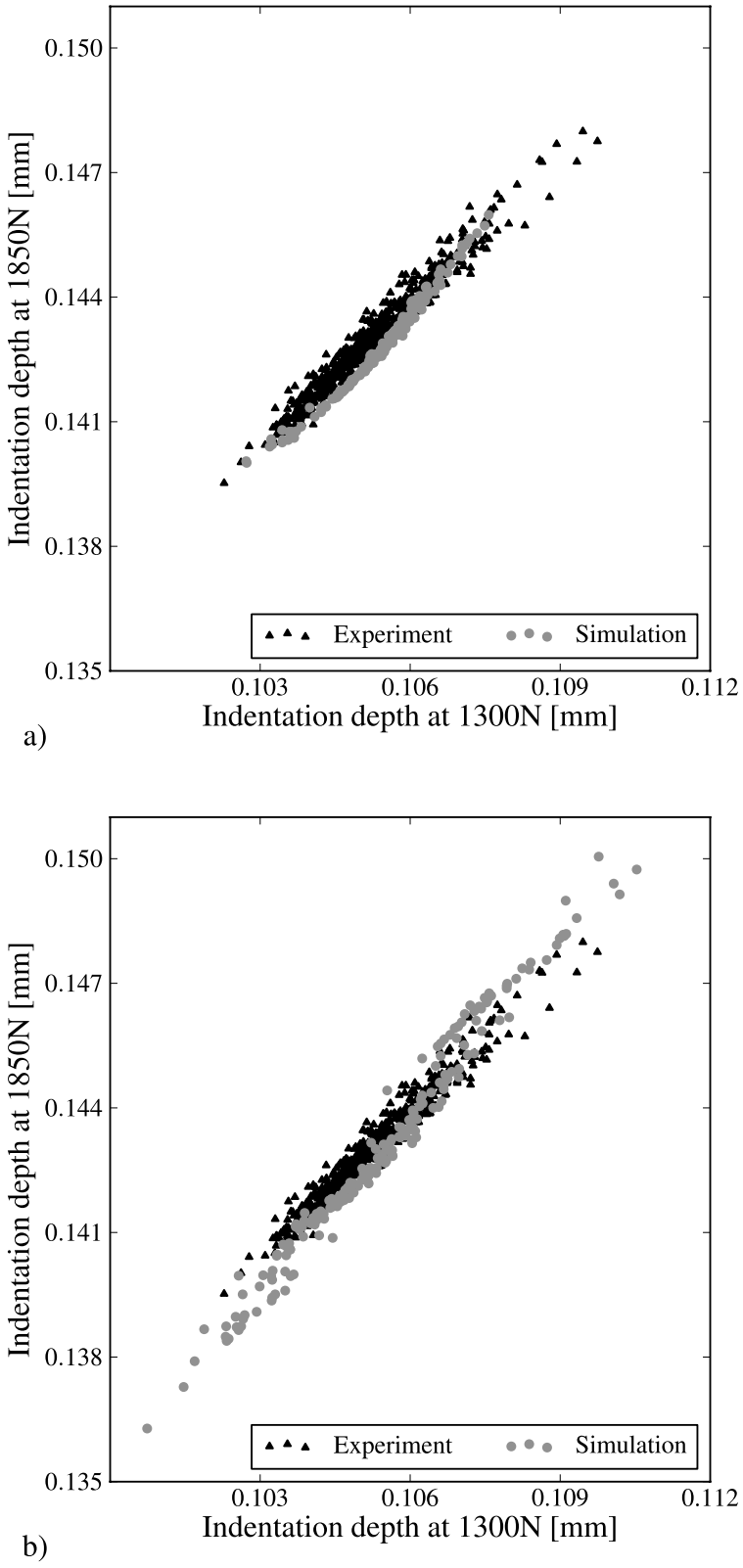


Figure 6.32: Two dimensional projection of the indentation depth at 1300 N versus the indentation depth at 1850 N of the 200 simulations and 534 experiments: a) Optimised numerical Model versus experiment, b) Numerical model of which the hardening parameters are scaled using the volume of the plastic zone versus experiment.

6.6.3 Significance of the distance

The p-values of the randomisation hypothesis tests are listed in table 6.22. From this table, it can be observed that the distance between the numerical results and the experimental results is significantly different for most samples regardless of the method used for scaling. The distance is only not significant for sample 1 containing 20 data points. Since most p-values are still smaller than 0.05, it must be concluded that there is a significant difference and that the influence of the scaled variance is most likely not as pronounced as some other difference between the numerical model and the experiments. Although, the distance is significant, it could be that the distance is smaller than the combined uncertainty in the experimental and numerical results. To ensure the distance is larger than the combined uncertainty, the minimum Euclidean distance between the experimental results and the simulated results is compared to the combined uncertainty, which consists of the measurement uncertainty and the numerical uncertainty. The results of this comparison are presented in the next subsection.

Table 6.22: The p-values of the hypothesis tests based on the distance (IBDM) between the numerical results and the experimental results for different numerical models and experimental samples.

Nr. data points	Numerical model	Sample 1	Sample 2	Sample 3	Sample 4
200	Optimised	0.001	0.001	-	-
	Optimised + area scaled	0.002	0.002	-	-
	Optimised + volume scaled	0.001	0.001	-	-
50	Optimised	0.016	0.001	0.018	0.001
	Optimised + area scaled	0.004	0.001	0.011	0.001
	Optimised + volume scaled	0.001	0.001	0.001	0.001
20	Optimised	0.185	0.079	0.002	0.177
	Optimised + area scaled	0.087	0.005	0.017	0.010
	Optimised + volume scaled	0.143	0.001	0.008	0.005

6.6.4 Distance versus measurement uncertainty

The discrete cumulative density function of the normalised distance based on the $1\,\sigma$ interval of the measurement uncertainty is presented in figure 6.33a whereas the discrete cumulative probability density function of the normalised distance based on the $2\,\sigma$ interval of the measurement uncertainty is presented in figure 6.33b. As shown in figure 6.33, not all experimental data points are within range of the numerical results after scaling of the variance of the hardening parameters. Moreover, the amount of experimental data points that are in the range of the numerical results has not significantly increased compared to the numerical model without scaling. This supports the conclusion presented in subsection 6.6.3, that the scaled variance has an inferior effect on the distance between the numerical results and the experimental results. Consequently, there is a different reason for the difference between the numerical results and the experimental results. To identify possible invalidated assumption, the assumptions of the numerical model are revised in the next section.

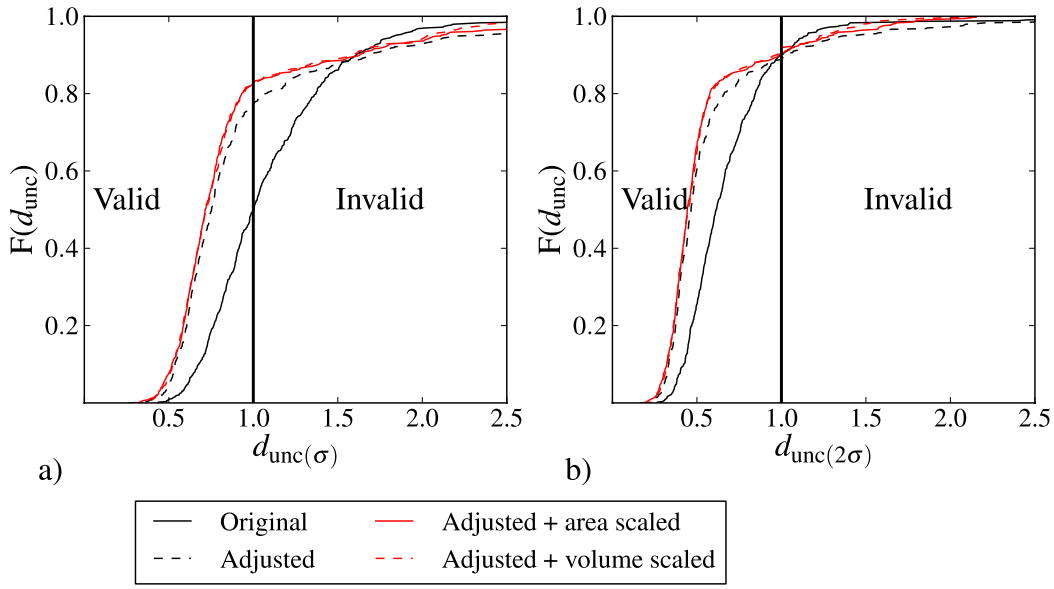


Figure 6.33: Minimum distance between the experimental data points and the numerical data points, normalised using the measurement uncertainty: a) Measurement uncertainty based on 1σ, b) Measurement uncertainty based on 2σ.

6.7 Revision of the assumptions of the numerical model

The assumptions concerning the boundary conditions, the contact friction, and the amount of finite elements are considered to be plausible, because their influence is significantly smaller than the difference between the numerical results and the experimental results. Furthermore, these assumptions would influence all indentations equally, which would result in a mean value shift.
Regarding the material model, some assumptions are made, which could have a significant influence on the numerical results. First of all, the material is assumed to be isotropic and governed by the Voce hardening law even though it is slightly anisotropic. However, identical indents can be created in different slightly anisotropic materials when their behaviour can be represented with the same equivalent isotropic material, because a sphere indenter cannot differ between the material behaviour along different directions due to its indenter form. Therefore, it is concluded that it is adequate to assume an isotropic representative material hardening behaviour to calculate the indentation behaviour. The difference in variance could be caused by inaccurate tensile experiments, but these experiments are performed sufficiently accurate and with a large number of samples to measure the variance of the stress-strain behaviour as shown in table 6.6. Also the Voce hardening law fits the experimental tensile curves well as presented in figure 6.19. Although all these assumptions are plausible, there is another issue with the description of the hardening behaviour based on the Voce Hardening law, namely the distribution of parameters of the hardening law. It is assumed that the hardening parameters are normally distributed with a correlation between each other on basis of the tensile experiments. Subsequently, the argument is raised that the variation of the local hardening behaviour is larger than that of the global hardening behaviour. Although it is possible to scale the variance using the central limit theorem from local to global, it is not possible to scale the variance from global to local without assuming the distribution function of the local properties, because the central limit theory states that a global property which is a summation of

many local properties is normally distributed regardless of the distribution of the local properties. Therefore, it was assumed that the local parameters are Gaussian distributed, which is most likely an incorrect assumption according to current validation results. Furthermore, the parameters σ_i and γ have not been scaled at all, and the correlation coefficients between the parameters have not been changed. All this indicates that the stochastic description of the local material behaviour is most likely incorrect.
Furthermore, all finite elements in the finite element model have the same material properties, which is incorrect because the elements near the indent are much smaller than the grain size of aluminium 2024. Therefore, it is required to use a constitutive crystal plasticity model and to vary the material properties between the elements. However, a simulation using the material properties of the grains of aluminium 2024 is not consistent with the intended use of the numerical model, namely identification of the representative local hardening behaviour from indentation experiments. Even though it is not consistent with the intended use of the numerical model, it might be required to investigate the microscopic hardening behaviour to model the representative local hardening behaviour. However, the goal of this work was to develop a validation method for multivariate results, which has been demonstrated under engineering conditions. Further investigation of the cause of the difference is therefore beyond the scope of this work.

7 Conclusions

The aim of this work was to develop a new validation method that is suitable to match multivariate numerical results with experimental results under the existence of scatter and uncertainty described. It is shown in this work that the newly developed Image Based Validation Method (IBVM) is a general applicable method to validate numerical models, which becomes increasingly valuable when the results deviate from normality and the results are subjected to uncertainty. Using this validation method, it is now possible to validate numerical models that generate multivariate arbitrarily distributed results using experimental results and uncertainties.
One of the key features of IBVM is its general applicability, due to its distance measure that quantifies the distributional difference between simulation and experiment using estimated probability density functions. Another key feature of IBVM is its capability to measure the distance between the numerical results and the experimental results based on the measurement uncertainty. It is found that the distance between simulation and experiment can be determined better when the underlying distributions are estimated using the measurement uncertainties instead of a theoretical derived estimator. However, this is only true when the samples contain sufficient data points to accurately estimate the underlying distribution of the samples. In case no measurement uncertainty is known, it is still possible to measure the difference between two estimated probability density functions using a theoretical Gaussian kernel.
To determine whether the Image Based Validation Method is an effective method to validate numerical models, the effectiveness of its Image Based Distance Measure (IBDM) has been determined and compared to the effectivity of the typically used distance measures, i.e. the Mahalanobis distance and the Box-M statistic. It is shown in this work that IBDM is at least as effective as the Mahalanobis distance to detect a mean value difference between non-normally distributed results using bivariate lognormally distributed results. However, IBDM is less effective than the Mahalanobis distance to detect a mean value difference between normally distributed samples. Furthermore, it has been shown that IBDM can be as effective as the Box-M to detect a difference in scatter between the samples when this difference is moderately large or when both samples contain sufficient data points.
Obviously, a difference between simulation and experiment in reality is most likely a combination of several differences. Using a combination of a rotational difference with a mean value difference as an example of such a difference, it is found that IBDM can be more effective than the Box-M statistic and the Mahalanobis distance. Especially when the results (experiment and simulation) contain many data points, it is found that IBDM is at least as effective as the Box-M statistic and the Mahalanobis distance. Furthermore, it has been shown using the sphere indentation example that IBDM is more effective than the Mahalanobis distance to detect a difference in scatter and mean value for moderately large samples. A similar behaviour is observed for normally distributed samples.
Overall, it can be concluded that IBDM is an effective method to measure the distance between two samples for two reasons. First, it can detect the difference between multivariate non-normally dis-

tributed samples at least as effective as the Mahalanobis distance. Second, in contrary to typically used distance measures, IBDM is capable to measure any difference between multivariate arbitrarily distributed data. Since numerical and experimental results are often non-normally distributed and contain (measurement) uncertainties, it can be concluded that the Image Based Validation Method is an adequate method to determine whether numerical results deviate from experimental results.

8 Outlook

The developed method can be used in future research to validate numerical models of which the results are multivariate and arbitrarily distributed without assuming a distribution. Furthermore, using this method, IBVM, it is possible to include the measurement uncertainty and the numerical uncertainty in its distance measure.
The effectivity of the developed validation method has been demonstrated for bivariate normal and bivariate lognormally distributed samples by benchmarking the method versus typically used methods, i.e. the Mahalanobis distance and the Box-M statistic. Furthermore, it has been shown that IBDM detected a difference between the numerical sphere indentation results and the experimental results where the Mahalanobis distance did not detect the difference. To gain further insight in the effectivity of IBVM, the benchmark test should be expanded by investigating the effectiveness of IBDM to detect differences between several other non-normally distributed samples, e.g. bimodal bivariate distributions. Also the influence of the measurement uncertainty and the numerical (simulation) uncertainty on the detectability of the distance has not been fully explored, because the difference between the theoretical kernel and the uncertainty based kernel is only investigated using several samples. To better understand the influence of the uncertainties on the effectivity of IBDM, a possible next step would be to systematically investigate the advantages of using IBVM based on the measurement uncertainty over IBDM based on a theoretical Gaussian kernel. For example, by means of comparing the operation characteristics of both methods.
To demonstrate the engineering relevance of IBVM, the sphere indentation problem has been chosen due to its abilities to generate multivariate experimental results. However, due to the large amount of assumptions and the measurement uncertainty on the experimental side, it was not possible to show that the numerical model is valid, i.e. a good representation of the experimental results. A possible next step would be to begin validation at the material model to determine whether the assumptions made concerning the local material behaviour of the material model are correct for sphere indentation modelling.
Finally, the Image Based Validation Method has been created for multivariate arbitrarily distributed results. But, it can also be used to compare numerical result fields to experimental result fields. For example, determining the resemblance between a numerically calculated strain field and an experimental strain fields measured using optical measurement systems.

9 Summary

During product development, engineers and scientist use a variety of numerical models to determine how assemblies, components or parts behave. To describe the behaviour of these products, multiple performance relevant quantities are used such as weight, stiffness, lifetime, efficiency and energy consumption. These performance quantities can be estimated using experiments or numerical models. However, it is often not possible to obtain all performance quantities using one approach. Therefore, it is necessary to combine the experimental and the numerical approach to obtain the performance quantities of interest. Consequently, it is necessary to ensure that both approaches represent reality accurately. To determine whether the numerical results differ from reality, validation is performed. During validation it is determined whether the distance between the numerical results and reality is significant, where reality is represented by the experimental results.

In context of validation, it is necessary to point out that experimental results scatter due to variations in the production process and the operation conditions to which the product is subjected. Therefore, it would be appropriate to compare the stochastic experimental results to stochastic numerical results. Since these results are generally multivariate and arbitrarily distributed, it is required to use a validation method that is suitable for arbitrarily distributed multivariate results. However, currently no validation method exists to compare such results.

In this work, a method is developed to validate numerical models using arbitrarily distributed multivariate results. To quantify the shape difference between the samples without assuming a distribution, the underlying distributions of the experimental and numerical results are estimated based on the results. Furthermore, the measurement uncertainties and the numerical uncertainties can be used explicitly in the distance measure for additional information. To determine whether the numerical model is valid, it is determined whether the numerical model is significantly different from the experimental results using a hypothesis test. Furthermore, it is determined whether the distance between the numerical results and the experimental results is larger than the uncertainties present in the numerical and experimental results.

To investigate whether the developed validation method is more effective than the typically used methods for multivariate problems, benchmark tests are performed. Since these distance measures were developed for multivariate normally distributed data, the benchmarks are performed using normally distributed data and a test data set that represents non-normally distributed data. Using these benchmark tests, it is evidenced that the developed method is more effective to determine the distance between the numerical and the experimental results than the typically used distance measures. Furthermore, it is demonstrated by means of an example, that it is meaningful to use the developed validation method for engineering problems.

Using the developed method, it is now possible to validate stochastic numerical models without assuming distributions for the experimental and numerical results. In this method, it is also possible to incorporate the measurement uncertainties and the simulation uncertainties in its distance measure.

A Probability density difference determination

The difference between two discrete estimated probability functions can be determined using the linear correlation coefficient, the rank correlation coefficient and the overlap coefficient according to Manders. The differences between these methods are investigated using the set of samples shown in table A.1. Of these samples a probability density function is estimated using the bivariate gaussian kernel and the balloon estimator to investigate the effect of a continuous and a discrete probability density function.

Table A.1: Samples to compare different image comparison methods

Sample nr./Data point nr.	Variable 1	Variable 2
1/1	1.2	3.4
1/2	1.6	2.4
1/3	1.9	3.1
2/4	1.3	3.2
2/5	1.4	2.6
2/6	1.7	2.9

The first difference investigated is the sensitivity of the correlation coefficients for scaling of the image object (zoom). If the object remains identical and the zoom factor is changed in both images equally, while assuming the discretisation is still adequate, the correlation between the images should be approximately equal. Using the balloon estimator, shown in equation 3.20 where $c_d r^d$ is equal to πr^2, the influence of zoom at the correlation coefficients is investigated if the distribution estimator is discrete. For the data presented in table A.1, a radius of 0.4 is chosen. To investigate the influence difference between a continuous kernel estimator and a discrete estimator (the balloon estimator), a bivariate kernel estimator with a chosen standard deviation equal to 0.15 and 0.3 for respectively variable 1 and variable 2. Furthermore, the discretisation of the probability density estimate is calculated using 500 x 500 discretisation steps to ensure an accurate discretisation. These settings do not influence the correlation coefficients significantly if the discretisation error remains insignificant and the distribution estimation is not under smoothed.
The difference between the set of estimated probability density functions is determined using the linear correlation (ρ_0), rank correlation (ρ_r), and the correlation coefficient according to Manders (ρ_m). When comparing the linear correlation coefficient and the correlation coefficient according to Manders presented in table A.2 of this example, there is no significant difference between these measurements when a large domain is selected because the mean density value converges to zero for infinitely large domains. When zooming in on the object of interest (the probability density function), the linear correlation coefficient changes significantly compared to the zoomed out situation. This dependency is undesired for validation since the object of interest should be in focus not the background of the object. Comparing the linear correlation coefficient and the rank correlation coefficient shows that the rank correlation method cannot find the difference because it is too

focussed on the background. Moreover, using the balloon estimator multiple discretisation steps with an absolute density value of zero are present, which makes a rank determination impossible. Since the linear and the rank correlation coefficient are ineffective in determining the difference between two objects of interest(without background influence), the overlap coefficient according to Manders is chosen as distance measurement method.

Table A.2: Correlation coefficients for different domains

Discretisation domain	Method	ρ_m	ρ_0	ρ_r
(-4.5, 7.5], (-3.0, 9.0]	Balloon	0.61	0.60	0.68
(0.0, 3.0], (1.5, 4.5]	Balloon	0.61	0.55	0.63
(-4.5, 7.5], (-3.0, 9.0]	Kernel	0.70	0.70	0.95
(0.0, 3.0], (1.5, 4.5]	Kernel	0.70	0.64	0.96

B Optimal kernel width

Assuming the kernel $K(t)$ is a radial symmetrical probability density function and that the unknown density f has bounded and continuous second derivatives, the kernel width can be determined objectively using the mean integrated square error. It is equal to the summation of the integrated square bias and the integrated variance. Using the multidimensional Taylor's theorem, the bias error and variance can be estimated as shown in equations B.2 and B.3. For mathematical simplification, the following constants C_1 and C_2 are defined by:

$$C_1 = \int {t_1}^2 K(t) dt C_2 = \int K(t)^2 \mathrm{d}t \tag{B.1}$$

$$bias_h(\mathbf{x}) \approx \frac{1}{2} h^2 C_1 \nabla^2 f(\mathbf{x}) \tag{B.2}$$

$$var\hat{f}(\mathbf{x}) \approx n^{-1} h^{-d} C_2 f(\mathbf{x}) \tag{B.3}$$

Hence the mean integrated square error is equal to:

$$MISE(\hat{f}) = \frac{1}{4} h^4 {C_1}^2 \int \nabla^2 f(\mathbf{x}) \mathrm{d}\mathbf{x} + n^{-1} h^{-d} C_2 \tag{B.4}$$

Minimising the mean integrated square error provides the optimal kernel width, which is presented in equation B.5.

$$h^{d+4} = \frac{d C_2}{{C_1}^2 n^1 \int (\nabla^2 f(\mathbf{x}))^2 \mathrm{d}\mathbf{x}} \tag{B.5}$$

To determine a numerical value for the optimal kernel width, it is mandatory to choose the probability density function of the kernel and assume a probability density function for the data. This is required to determine the second derivatives of the probability density function. Assuming the data is taken from a standard normal distribution, the optimal kernel width is equal to:

$$h^{d+4} = \frac{d C_2 (2\sqrt{\pi})^2}{{C_1}^2 n^1 (\frac{1}{2} d + \frac{1}{4} d^2)} \tag{B.6}$$

If a gaussian kernel is used to estimate the probability density function of the data, the optimal kernel width is equal to:

$$\sigma_{\mathrm{opt}} = \sigma n^{-1/(d+4)} \tag{B.7}$$

C Sample data points related to table 4.3

Table C.1: Normally distributed sample data points to determine the influence of discretisation on distance measurements

Data point number	Sample 1	Sample 2
1	(4.84 , 8.17)	(4.96 , 7.45)
2	(4.87 , 7.66)	(5.25 , 7.90)
3	(4.72 , 7.90)	(5.02 , 8.13)
4	(5.30 , 7.82)	(4.81 , 7.57)
5	(4.96 , 8.21)	(4.84 , 8.57)
6	(5.03 , 7.62)	(4.78 , 8.10)
7	(4.64 , 8.25)	(5.12 , 7.81)
8	(4.97 , 7.71)	(4.92 , 8.32)
9	(5.40 , 8.05)	(5.00 , 8.13)
10	(5.08 , 8.09)	(4.51 , 7.90)
11	(5.41 , 8.05)	(4.79 , 8.19)
12	(5.01 , 8.38)	(4.89 , 8.23)
13	(5.18 , 7.90)	(4.50 , 7.92)
14	(5.22 , 8.12)	(4.83 , 8.03)
15	(4.50 , 8.34)	(4.51 , 8.21)
16	(5.00 , 7.77)	(5.10 , 7.92)
17	(5.03 , 8.22)	(4.76 , 7.91)
18	(5.02 , 8.26)	(5.42 , 7.38)
19	(5.17 , 8.08)	(4.96 , 8.05)
20	(5.02 , 7.81)	(5.06 , 7.94)
21	(5.46 , 8.34)	(4.92 , 8.04)
22	(5.00 , 8.10)	(4.55 , 8.11)
23	(5.30 , 8.03)	(5.09 , 7.79)
24	(4.88 , 8.34)	(4.77 , 7.98)
25	(4.66 , 8.41)	(4.97 , 7.86)
26	(5.66 , 7.71)	(5.34 , 7.37)
27	(5.00 , 8.06)	(4.75 , 8.23)
28	(5.66 , 7.81)	(4.95 , 8.47)
29	(5.11 , 8.33)	(4.84 , 8.07)
30	(5.18 , 8.05)	(4.67 , 7.49)
31	(5.12 , 8.20)	(4.72 , 7.87)
32	(5.10 , 8.17)	(5.04 , 8.42)
33	(4.97 , 7.93)	(4.92 , 7.79)
34	(4.77 , 8.11)	(4.67 , 8.17)
35	(5.49 , 8.17)	(5.49 , 7.97)
36	(4.78 , 8.02)	(5.10 , 7.73)
37	(4.78 , 7.99)	(5.40 , 8.07)
38	(5.19 , 8.14)	(4.80 , 8.24)
39	(5.41 , 8.30)	(5.11 , 7.96)
40	(5.35 , 8.21)	(5.18 , 7.92)
41	(5.12 , 7.88)	(5.10 , 8.03)
42	(5.12 , 7.97)	(5.01 , 7.99)
43	(4.65 , 8.21)	(4.85 , 7.96)
44	(5.18 , 8.11)	(4.82 , 8.09)
45	(4.89 , 8.16)	(4.78 , 7.79)

Table C.2: Lognormally distributed sample data points to determine the influence of discretisation on distance measurements

Data point number	Sample 1	Sample 2
1	(4.66 , 4.51)	(4.13 , 9.22)
2	(4.84 , 8.84)	(7.19 , 8.15)
3	(3.82 , 7.14)	(5.01 , 9.68)
4	(4.71 , 6.75)	(3.47 , 11.57)
5	(5.07 , 11.78)	(5.34 , 9.16)
6	(5.40 , 7.44)	(7.29 , 10.38)
7	(6.21 , 7.00)	(4.28 , 5.54)
8	(3.00 , 7.06)	(5.24 , 8.48)
9	(7.31 , 4.21)	(4.95 , 9.98)
10	(6.77 , 4.16)	(9.34 , 5.80)
11	(4.49 , 6.30)	(4.73 , 7.21)
12	(5.79 , 7.19)	(6.88 , 9.57)
13	(4.95 , 8.82)	(3.68 , 7.08)
14	(3.94 , 9.39)	(7.25 , 8.17)
15	(4.65 , 8.13)	(5.71 , 8.42)
16	(3.78 , 9.18)	(4.84 , 8.26)
17	(3.49 , 9.16)	(3.88 , 8.62)
18	(5.35 , 7.99)	(5.48 , 6.89)
19	(4.04 , 5.06)	(6.53 , 6.49)
20	(4.35 , 9.69)	(4.92 , 11.26)
21	(5.14 , 7.29)	(4.94 , 6.41)
22	(4.63 , 12.37)	(9.26 , 6.43)
23	(7.89 , 7.54)	(7.88 , 9.21)
24	(4.91 , 7.69)	(5.48 , 7.51)
25	(4.16 , 13.59)	(4.66 , 9.56)
26	(2.97 , 7.17)	(5.80 , 7.02)
27	(4.48 , 8.08)	(7.65 , 10.84)
28	(4.16 , 8.34)	(5.43 , 10.70)
29	(5.35 , 5.92)	(3.89 , 7.91)
30	(4.19 , 7.49)	(3.45 , 9.53)
31	(3.92 , 8.56)	(4.99 , 5.33)
32	(4.09 , 7.96)	(6.07 , 8.73)
33	(3.11 , 8.66)	(4.86 , 8.55)
34	(3.50 , 4.69)	(5.79 , 8.15)
35	(7.20 , 8.34)	(3.90 , 7.67)
36	(4.05 , 8.48)	(5.79 , 8.63)
37	(5.46 , 6.43)	(3.40 , 9.96)
38	(2.99 , 9.58)	(2.99 , 10.89)
39	(5.30 , 6.33)	(6.54 , 8.02)
40	(3.70 , 6.86)	(5.43 , 9.49)
41	(4.00 , 9.82)	(5.84 , 8.88)
42	(3.92 , 6.29)	(4.35 , 9.10)
43	(4.48 , 10.66)	(4.72 , 5.85)
44	(5.37 , 7.18)	(4.85 , 6.16)
45	(3.87 , 7.57)	(4.32 , 10.87)

D Projections of the sphere indentation results

This appendix includes the results of the original numerical model and the results of the optimised numerical model, which will be compared to the original results in section 6.5. These figures illustrate the difference between the experimental results and the numerical results. The experimental results and the numerical results of the original numerical model (see section 6.4) are presented on the left hand side while the experimental results and the results of the optimised numerical model (see section 6.5) are shown on the right hand side.

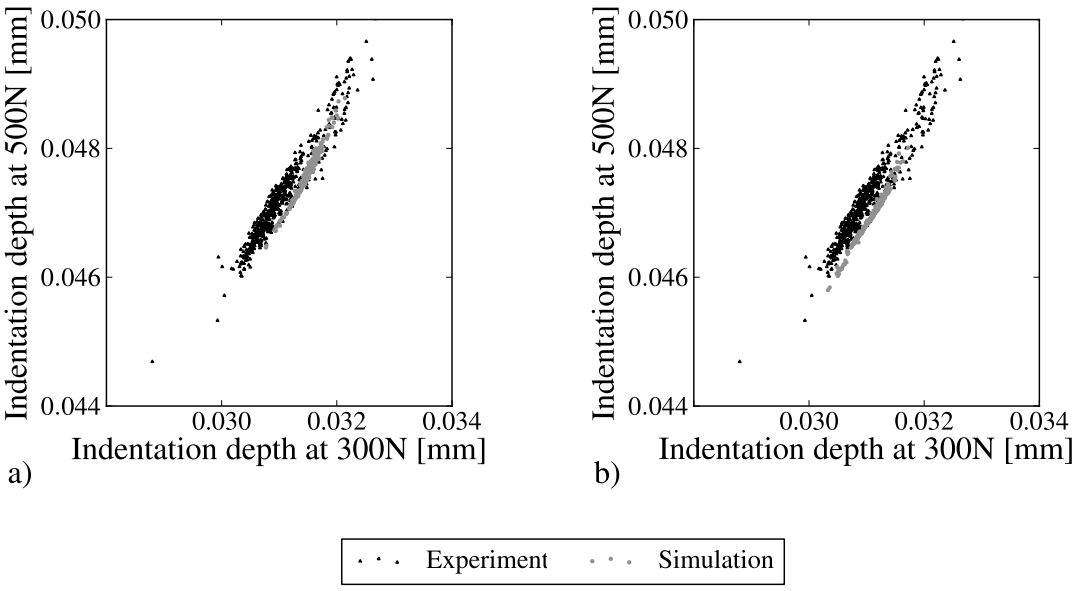


Figure D.1: The indentation depth at 300 N versus the indentation depth at 500 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

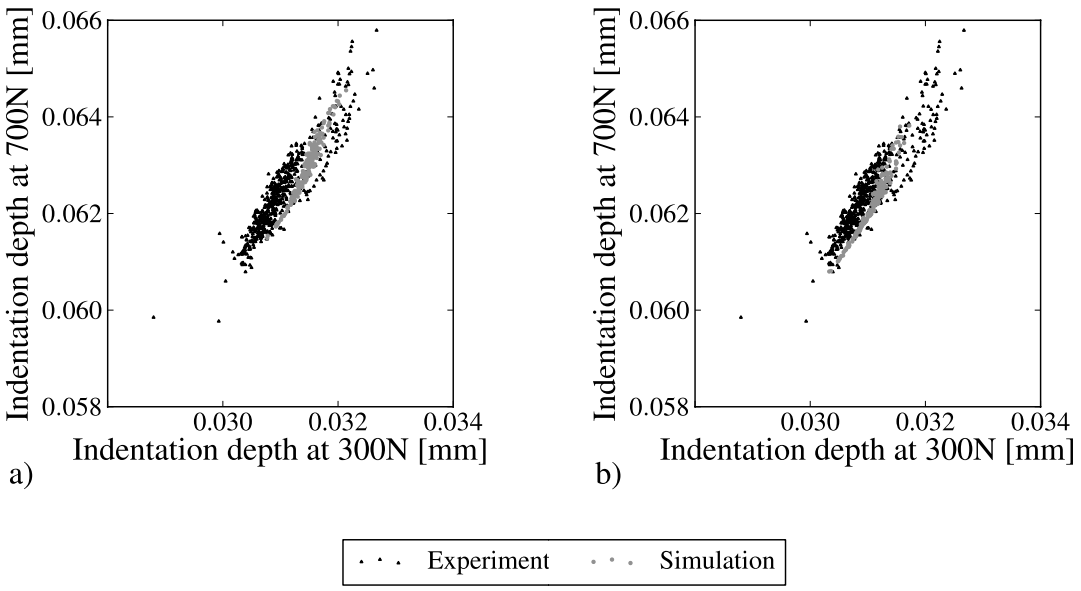


Figure D.2: The indentation depth at 300 N versus the indentation depth at 700 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

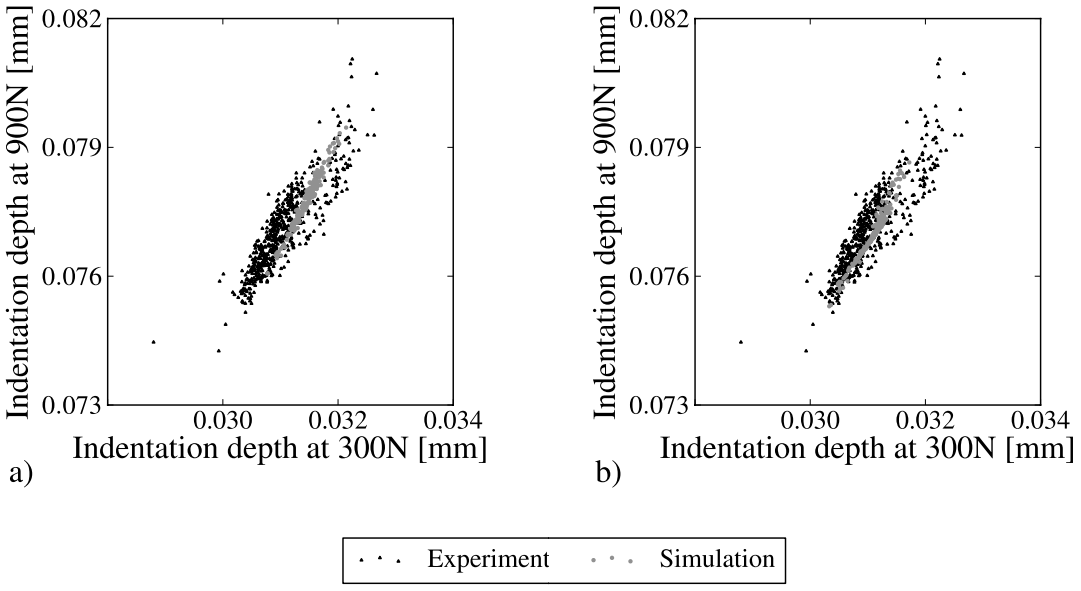


Figure D.3: The indentation depth at 300 N versus the indentation depth at 900 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

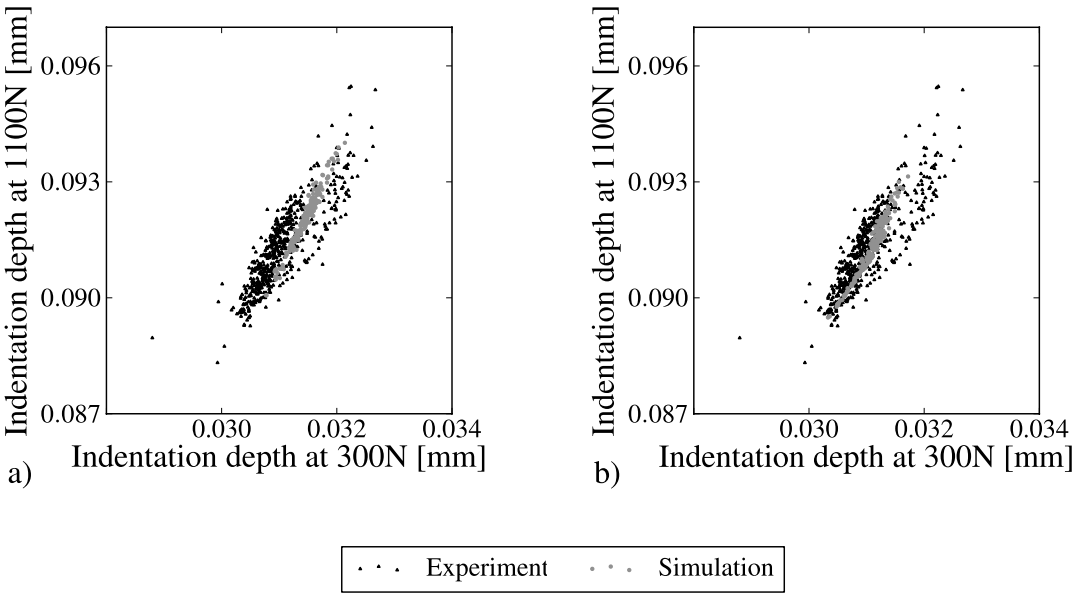


Figure D.4: The indentation depth at 300 N versus the indentation depth at 1100 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

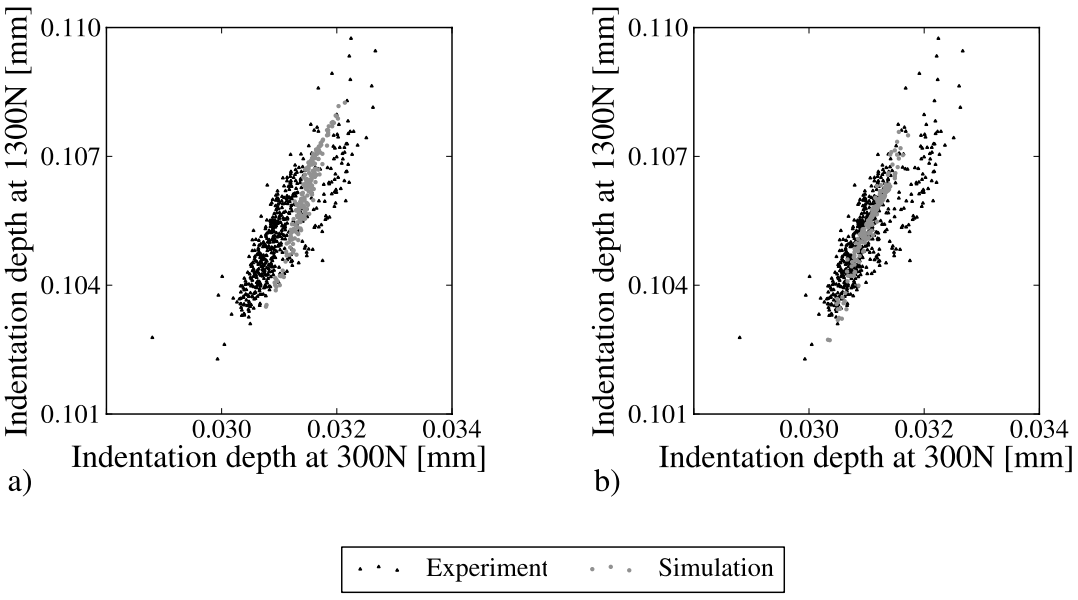


Figure D.5: The indentation depth at 300 N versus the indentation depth at 1300 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

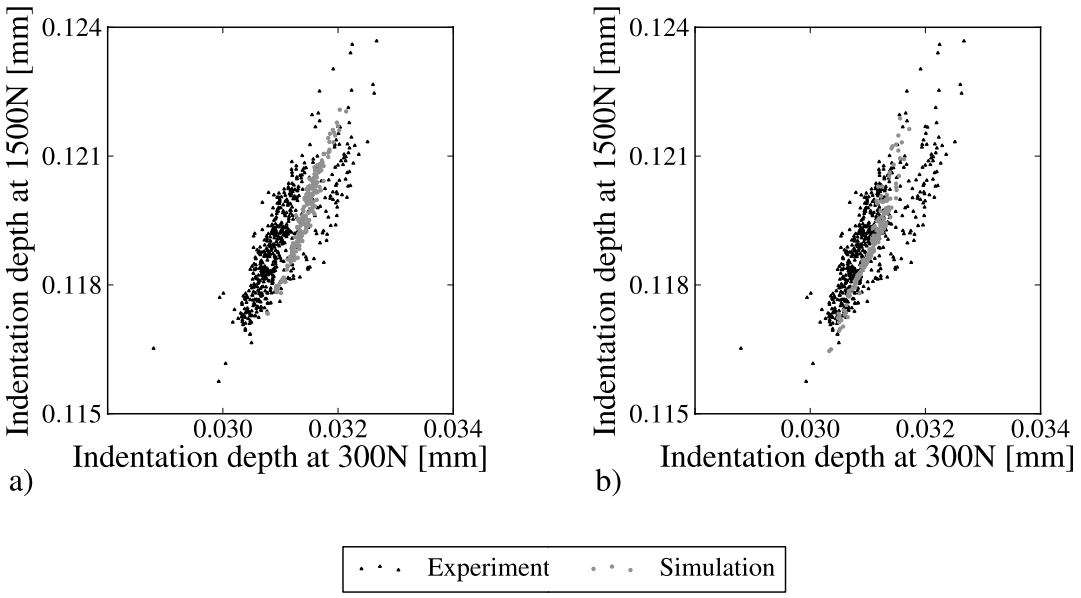


Figure D.6: The indentation depth at 300 N versus the indentation depth at 1500 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

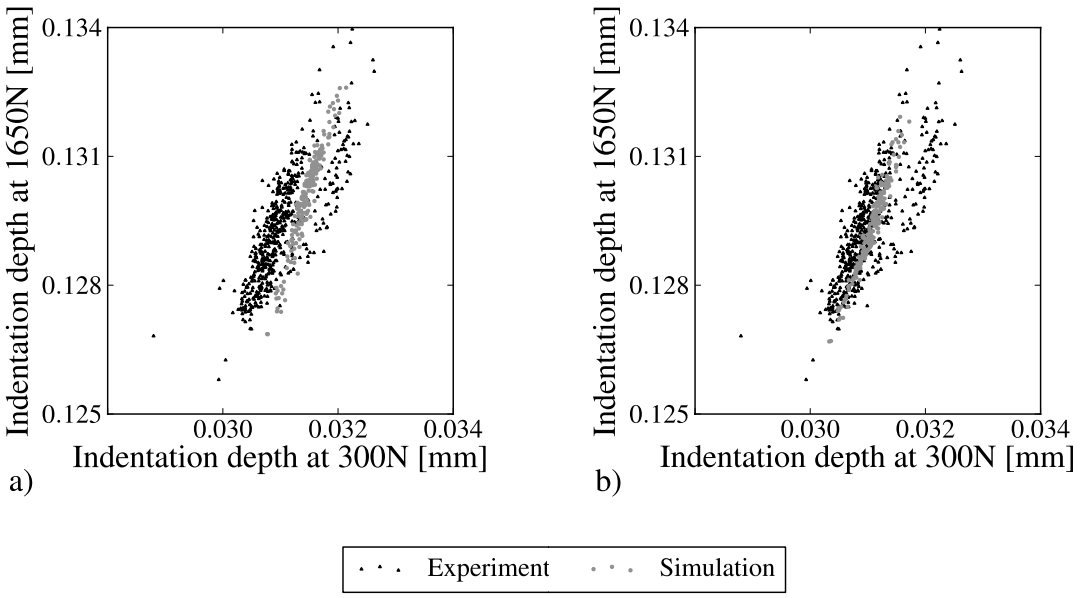


Figure D.7: The indentation depth at 300 N versus the indentation depth at 1650 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

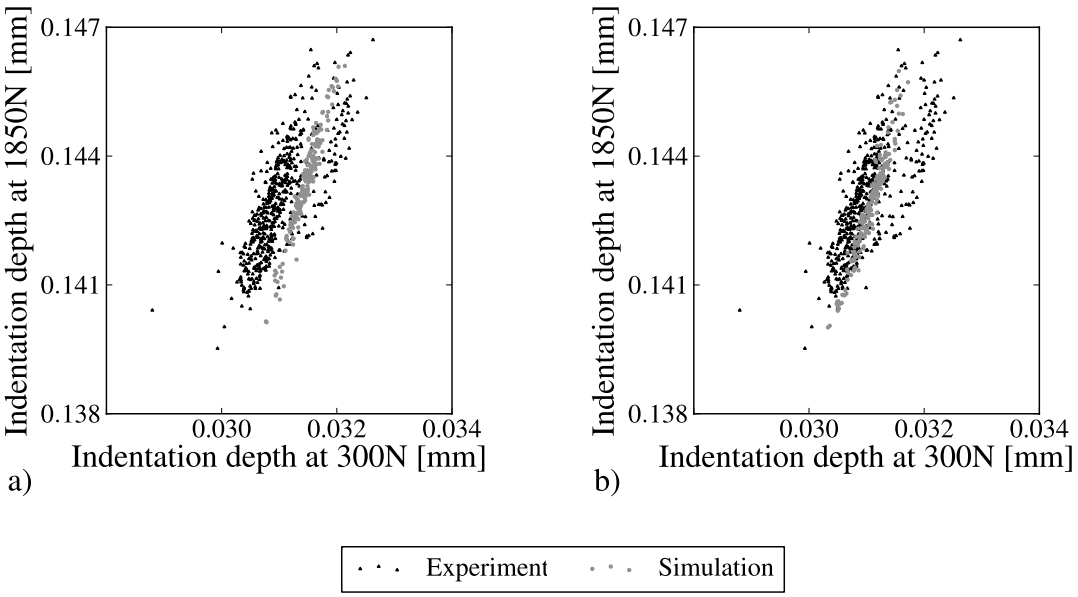


Figure D.8: The indentation depth at 300 N versus the indentation depth at 1850 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

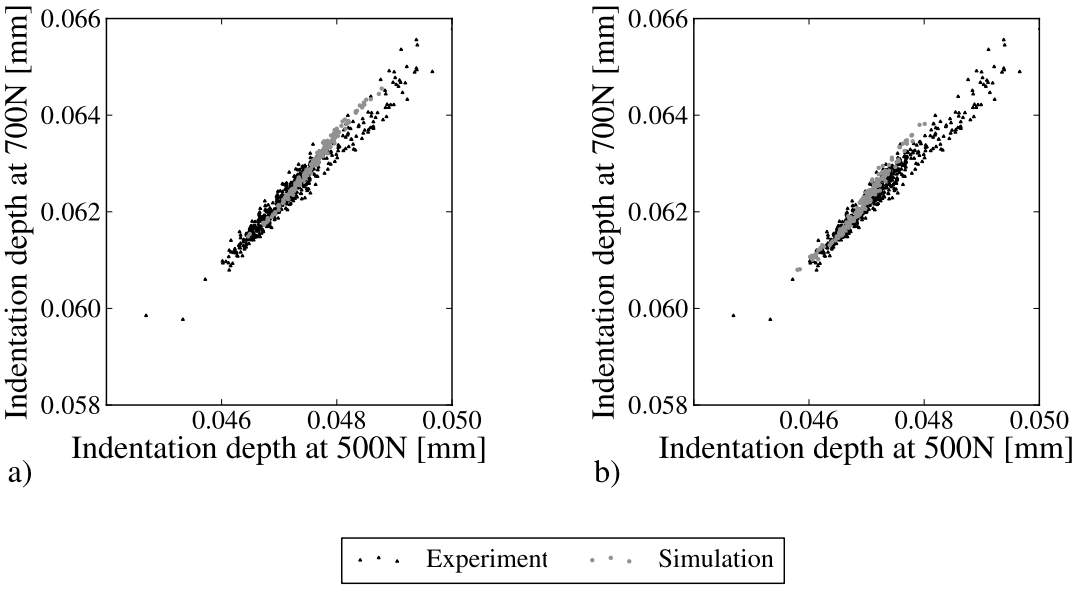


Figure D.9: The indentation depth at 500 N versus the indentation depth at 700 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

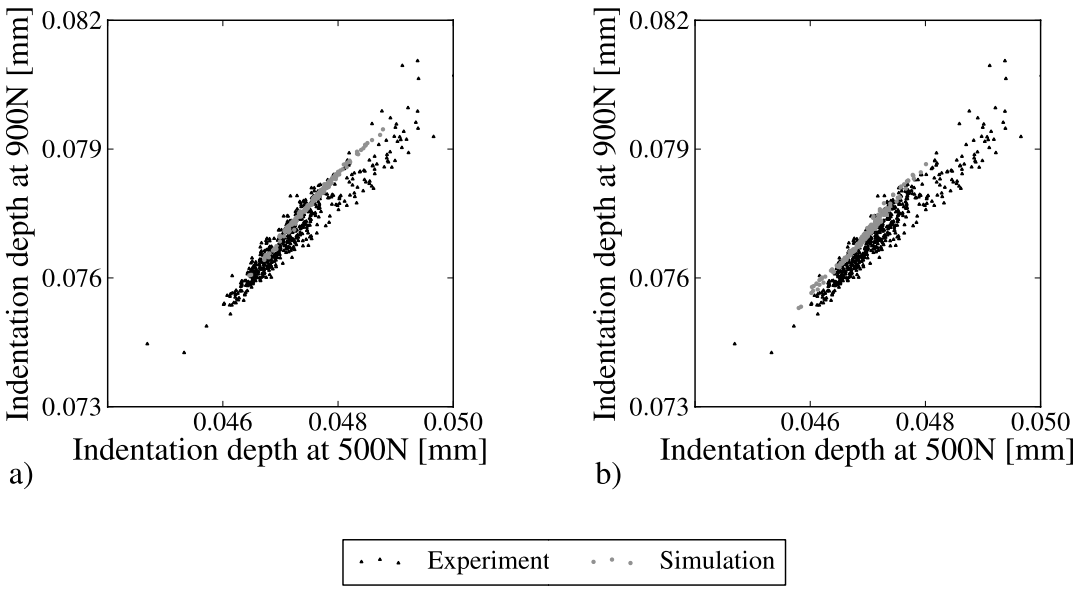


Figure D.10: The indentation depth at 500 N versus the indentation depth at 900 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

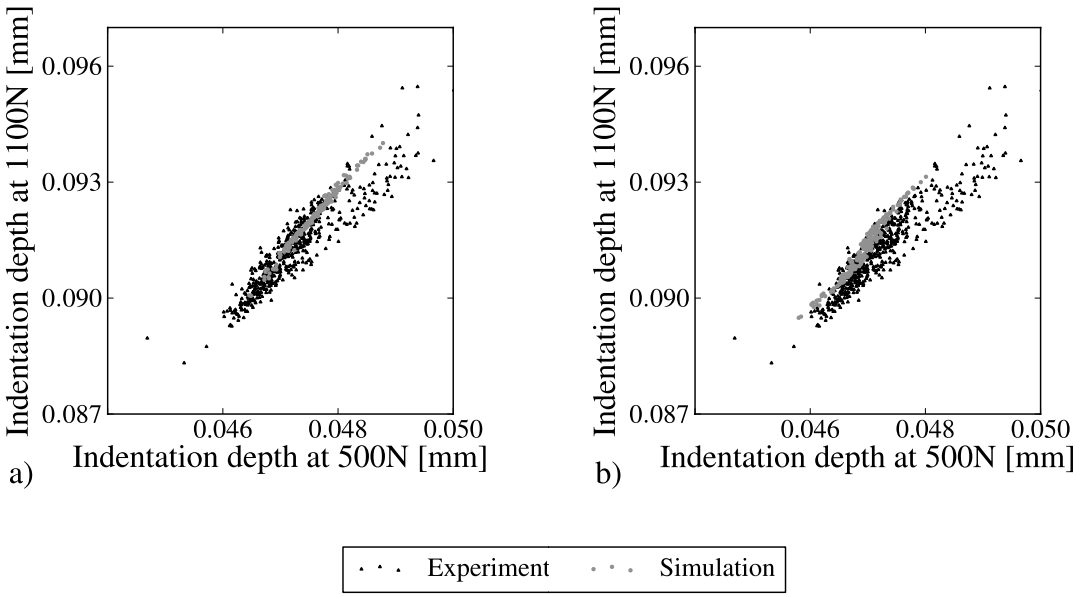


Figure D.11: The indentation depth at 500 N versus the indentation depth at 1100 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

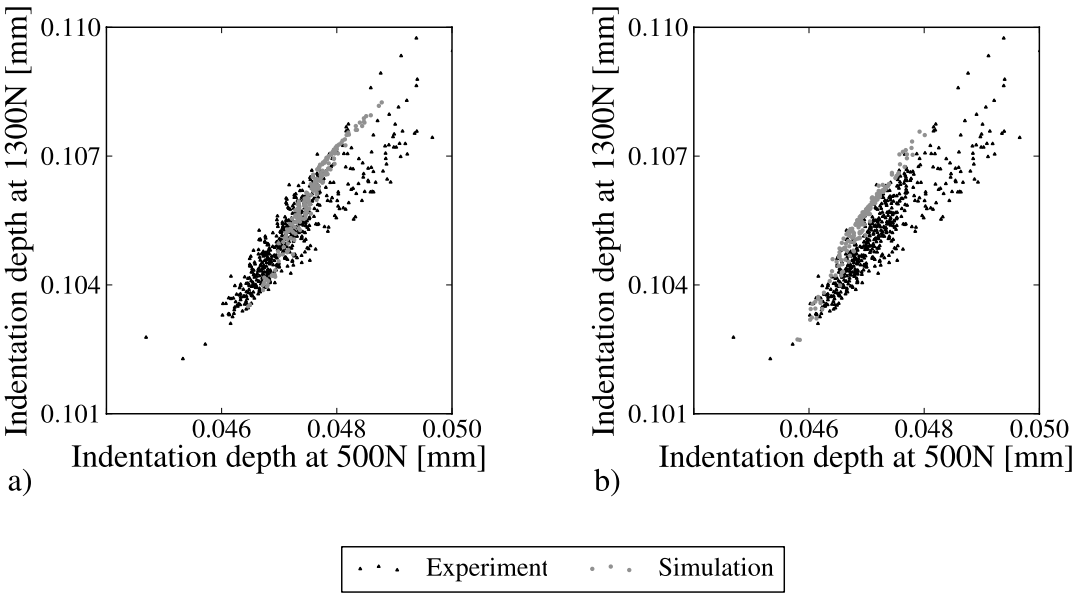


Figure D.12: The indentation depth at 500 N versus the indentation depth at 1300 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

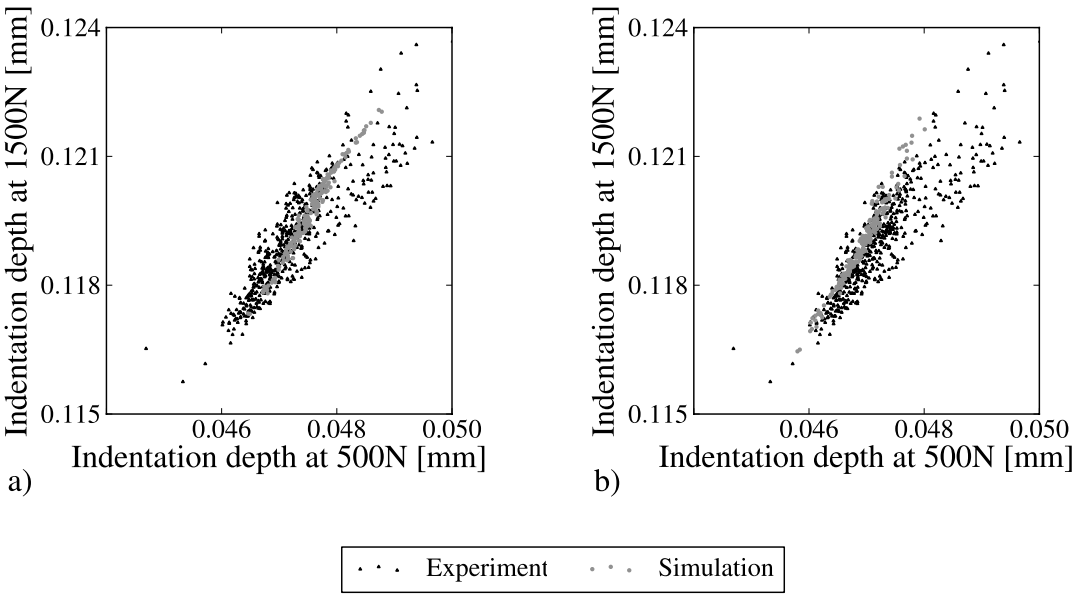


Figure D.13: The indentation depth at 500 N versus the indentation depth at 1500 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

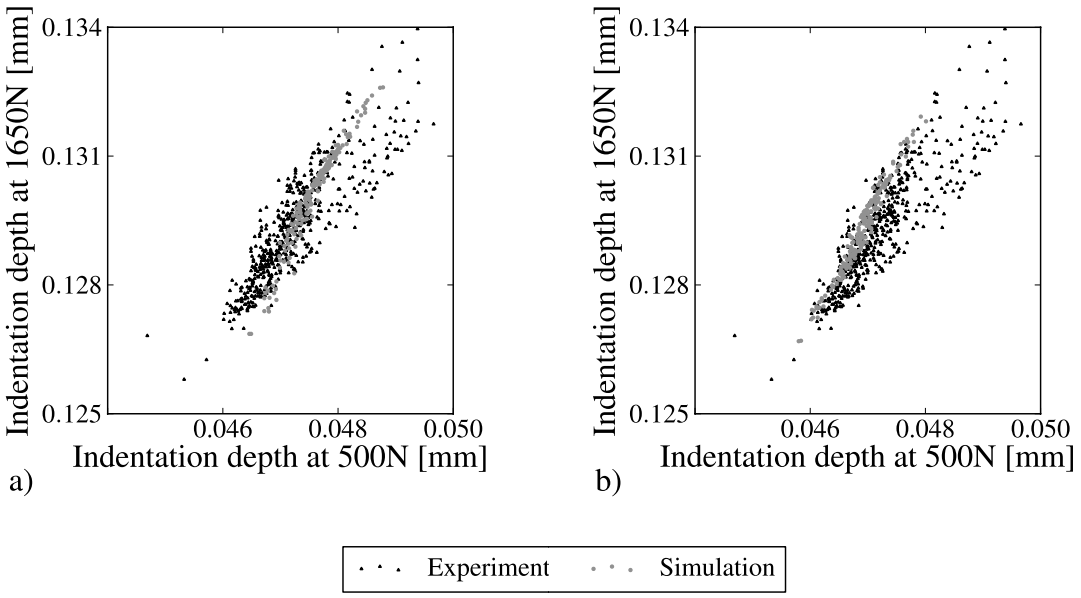


Figure D.14: The indentation depth at 500 N versus the indentation depth at 1650 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

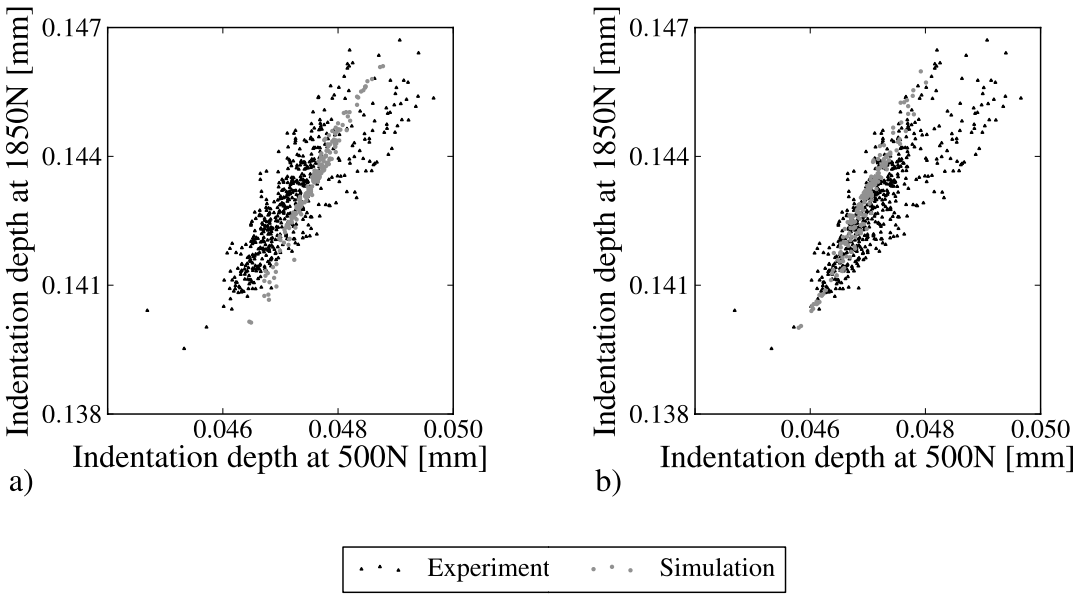


Figure D.15: The indentation depth at 500 N versus the indentation depth at 1850 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

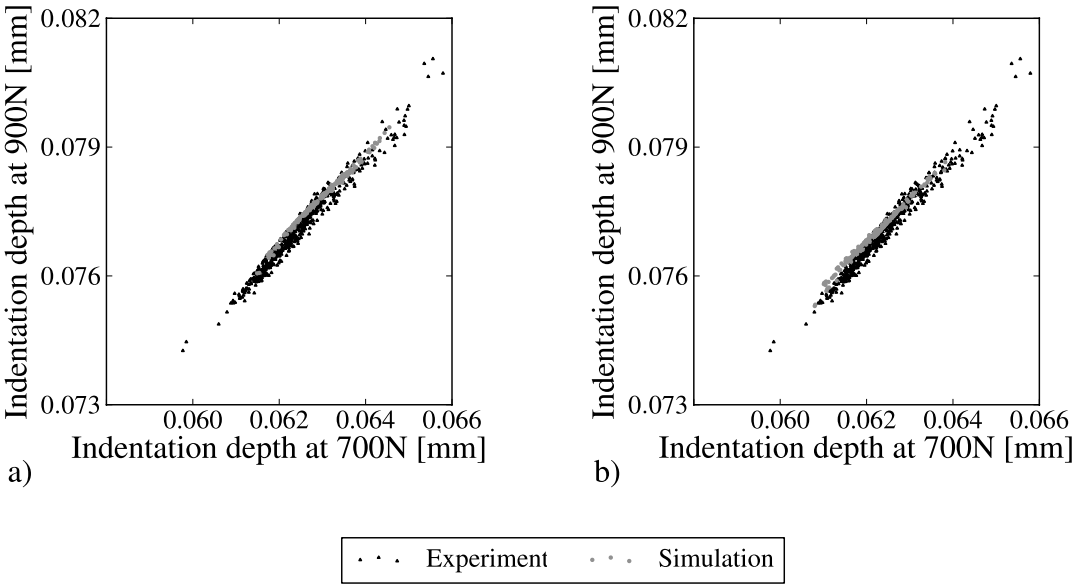


Figure D.16: The indentation depth at 700 N versus the indentation depth at 900 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

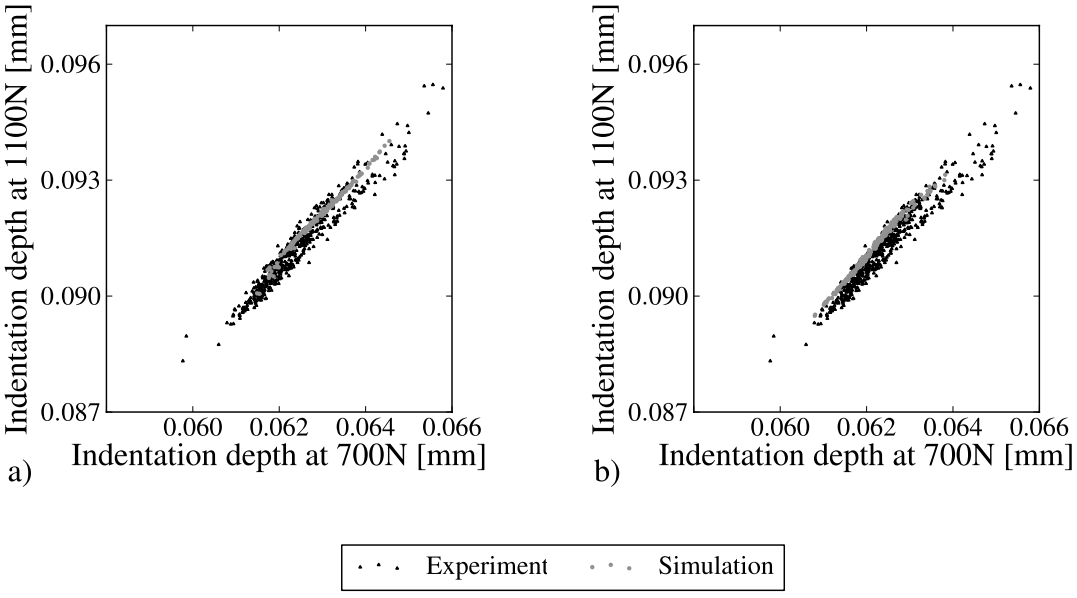


Figure D.17: The indentation depth at 700 N versus the indentation depth at 1100 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

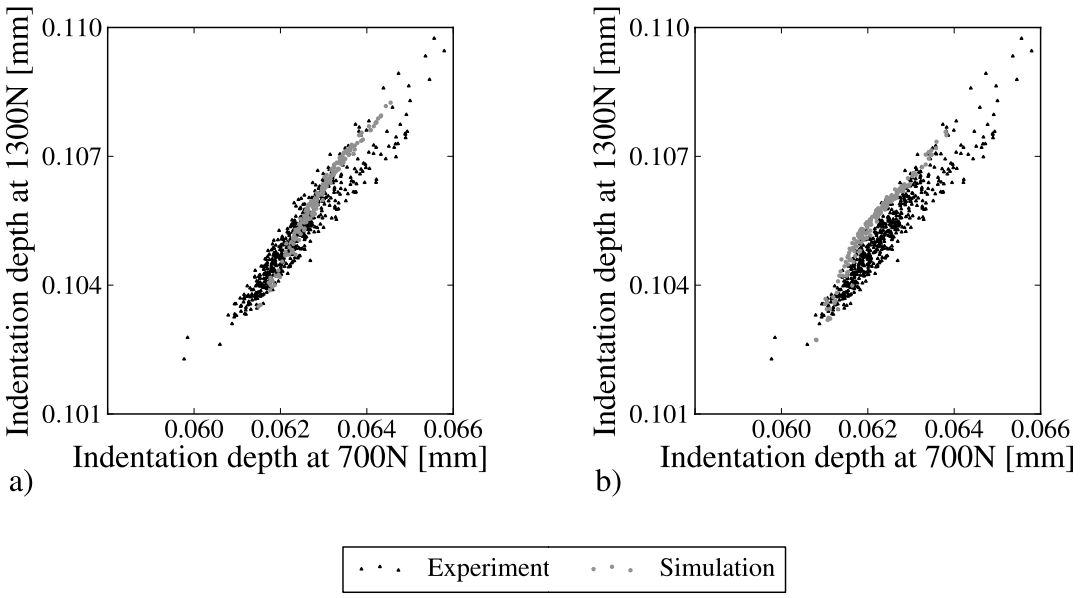


Figure D.18: The indentation depth at 700 N versus the indentation depth at 1300 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

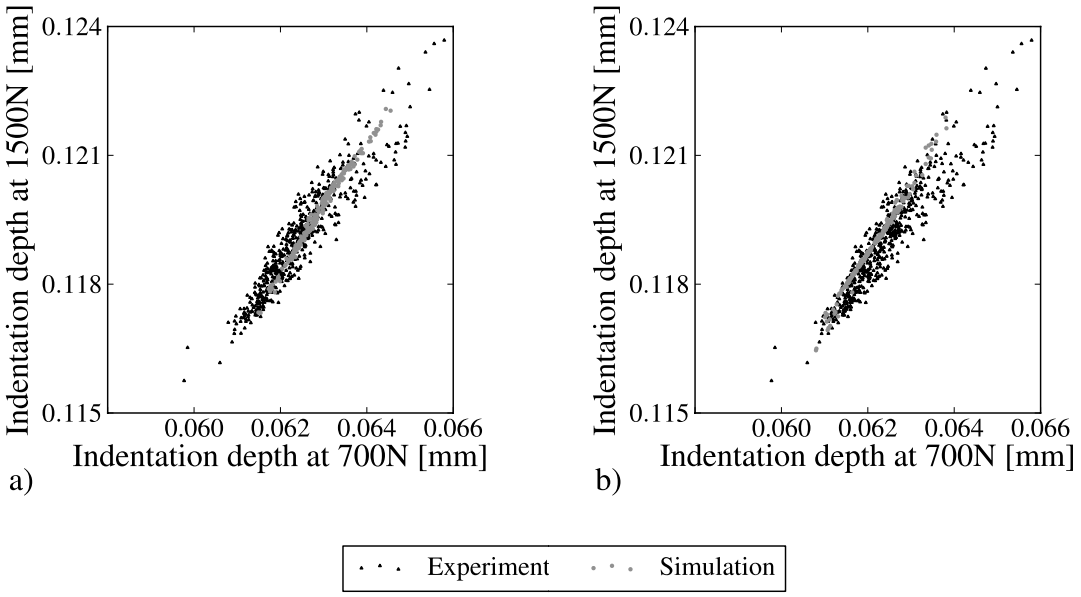


Figure D.19: The indentation depth at 700 N versus the indentation depth at 1500 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

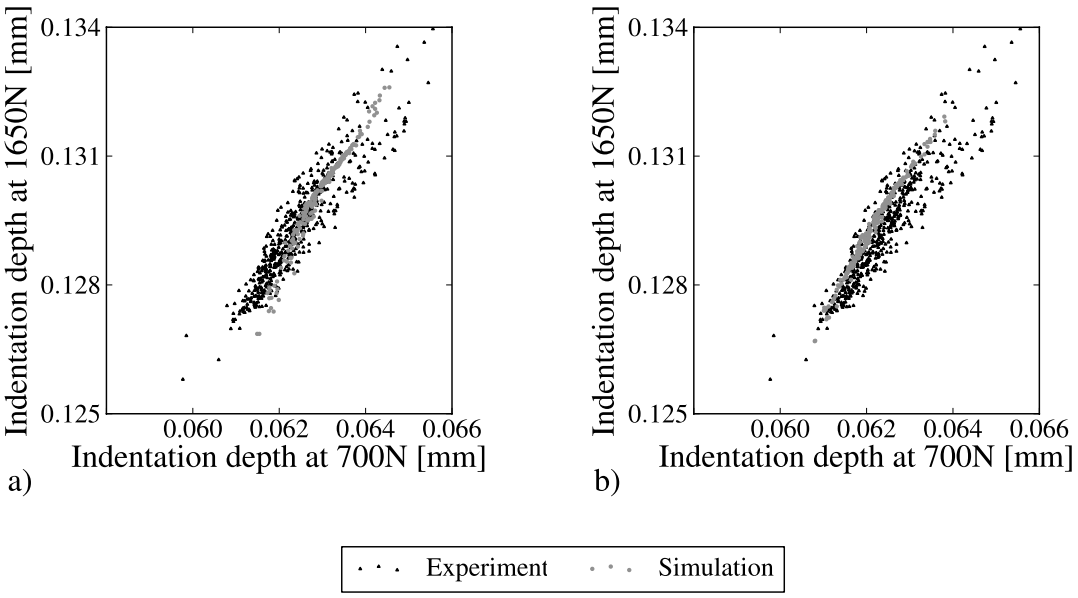


Figure D.20: The indentation depth at 700 N versus the indentation depth at 1650 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

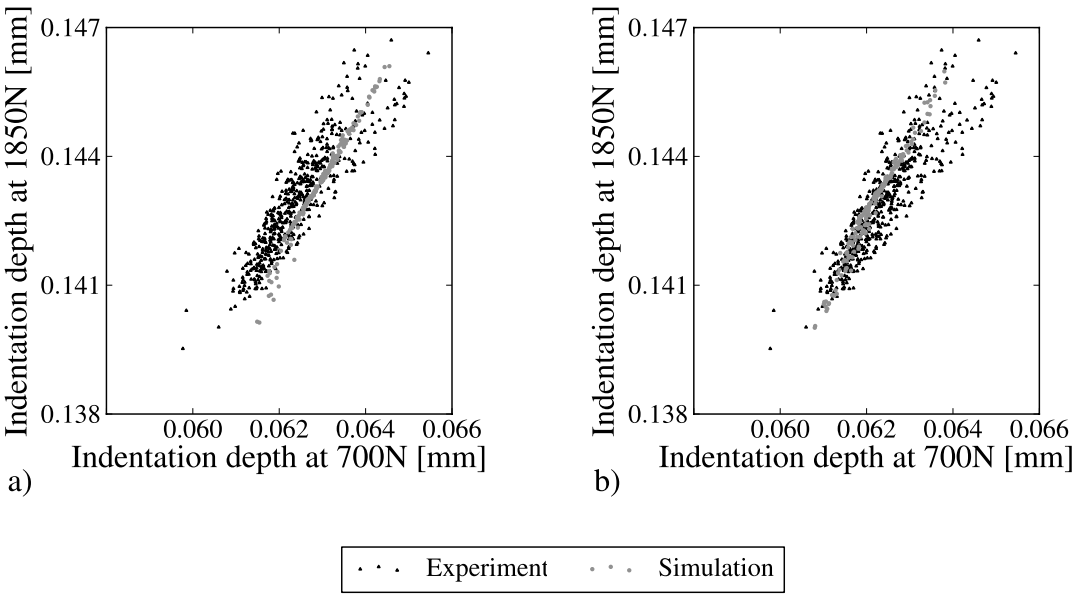


Figure D.21: The indentation depth at 700 N versus the indentation depth at 1850 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

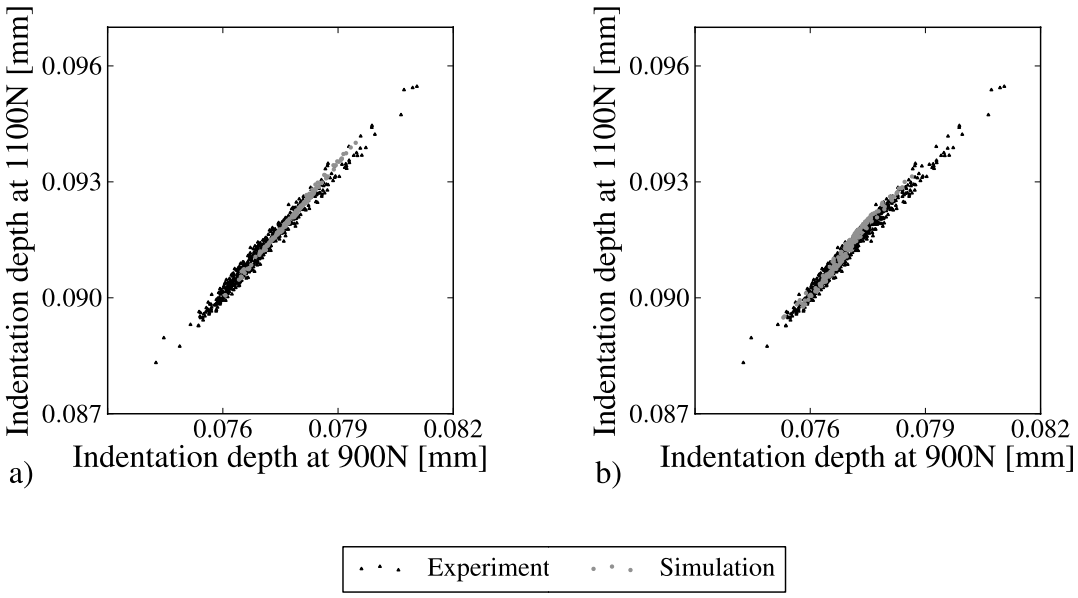


Figure D.22: The indentation depth at 900 N versus the indentation depth at 1100 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

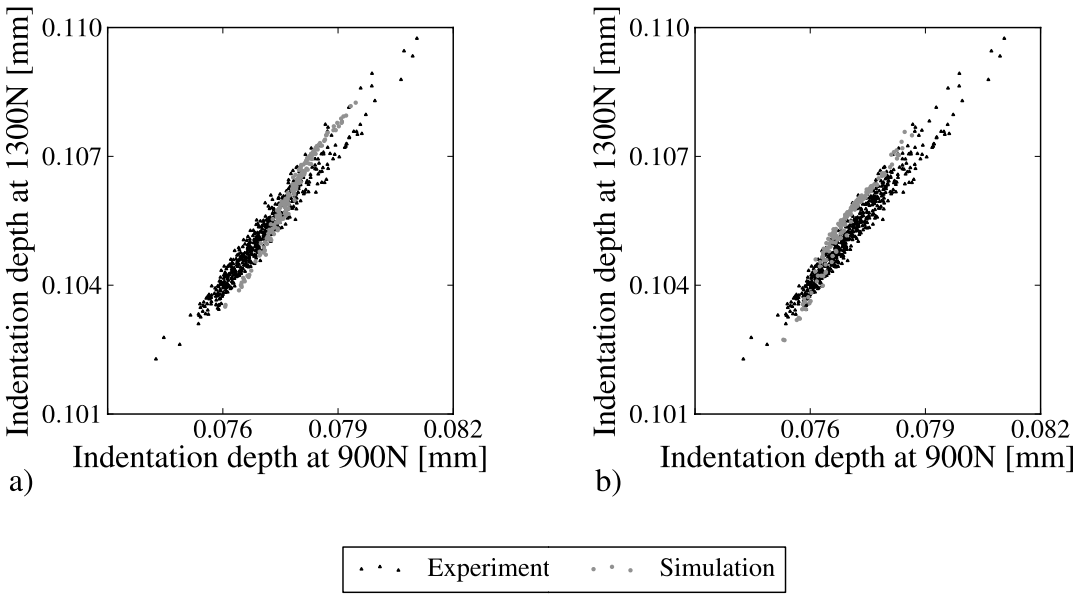


Figure D.23: The indentation depth at 900 N versus the indentation depth at 1300 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

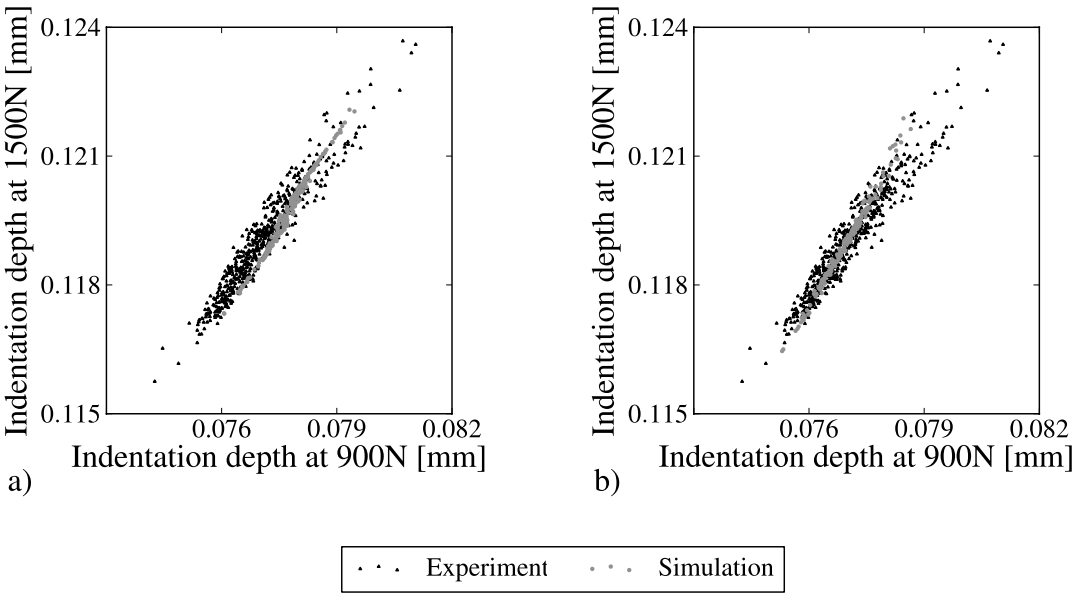


Figure D.24: The indentation depth at 900 N versus the indentation depth at 1500 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

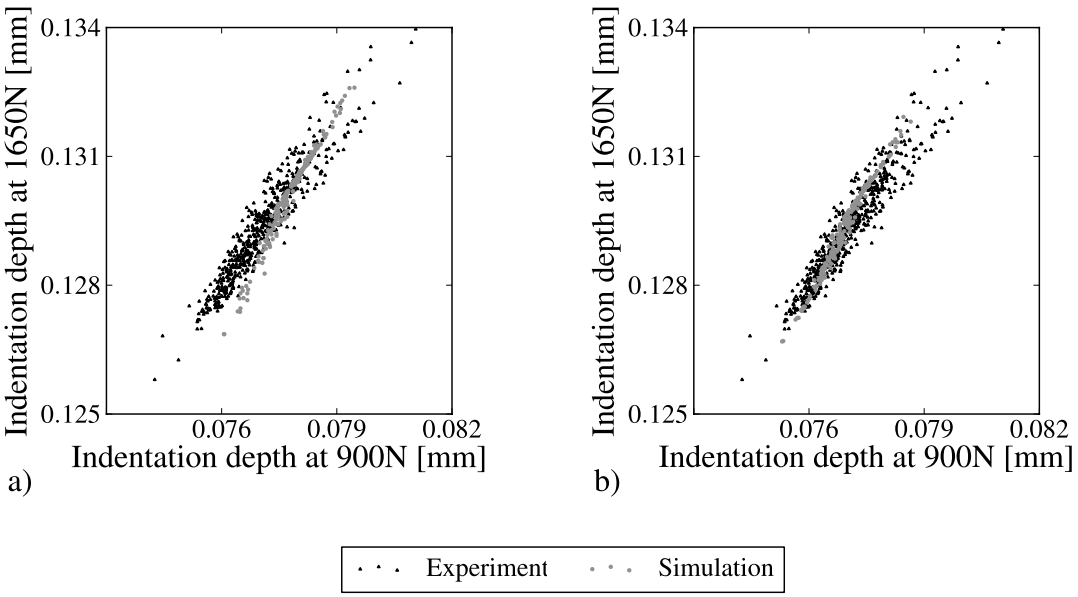


Figure D.25: The indentation depth at 900 N versus the indentation depth at 1650 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

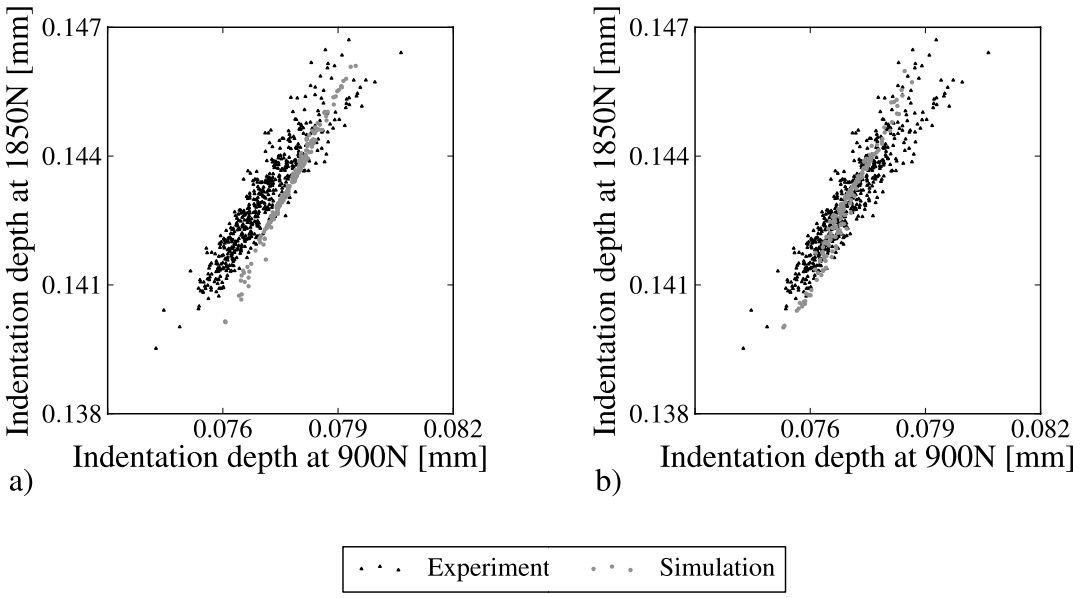


Figure D.26: The indentation depth at 900 N versus the indentation depth at 1850 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

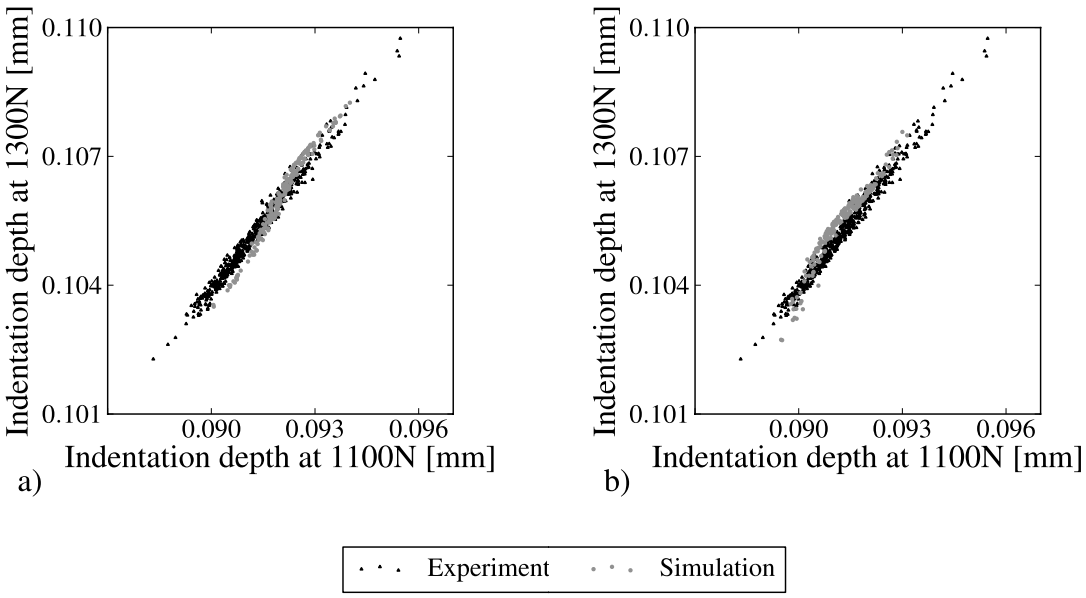


Figure D.27: The indentation depth at 1100 N versus the indentation depth at 1300 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

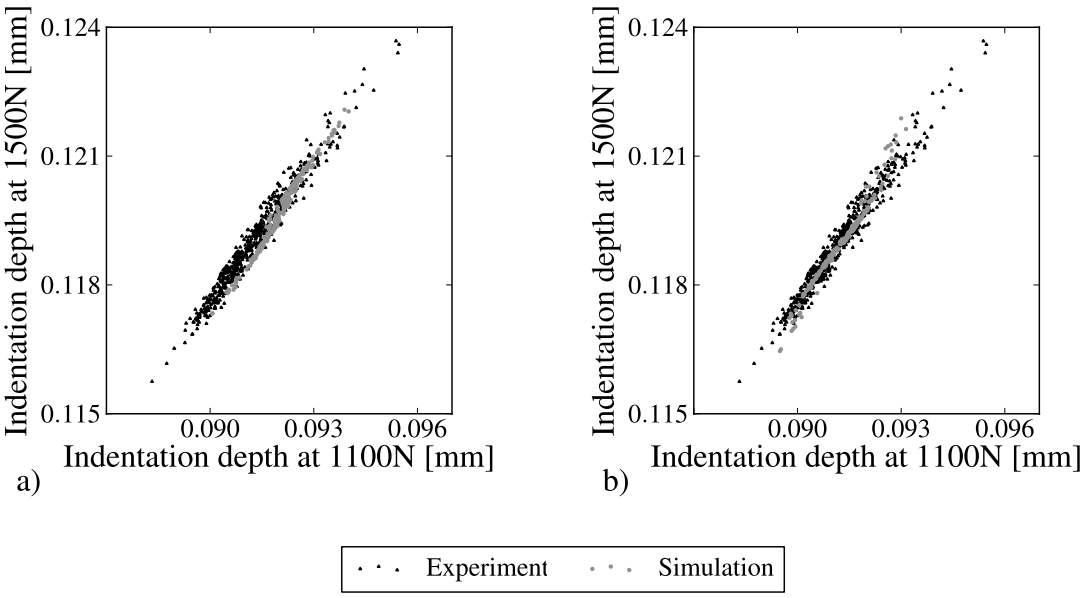


Figure D.28: The indentation depth at 1100 N versus the indentation depth at 1500 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

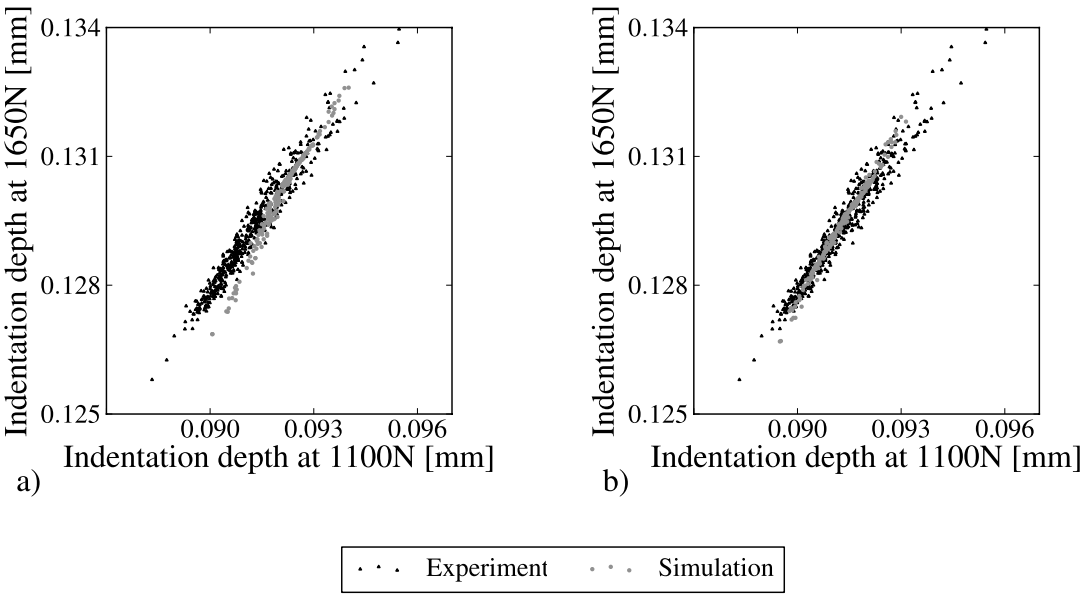


Figure D.29: The indentation depth at 1100 N versus the indentation depth at 1650 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

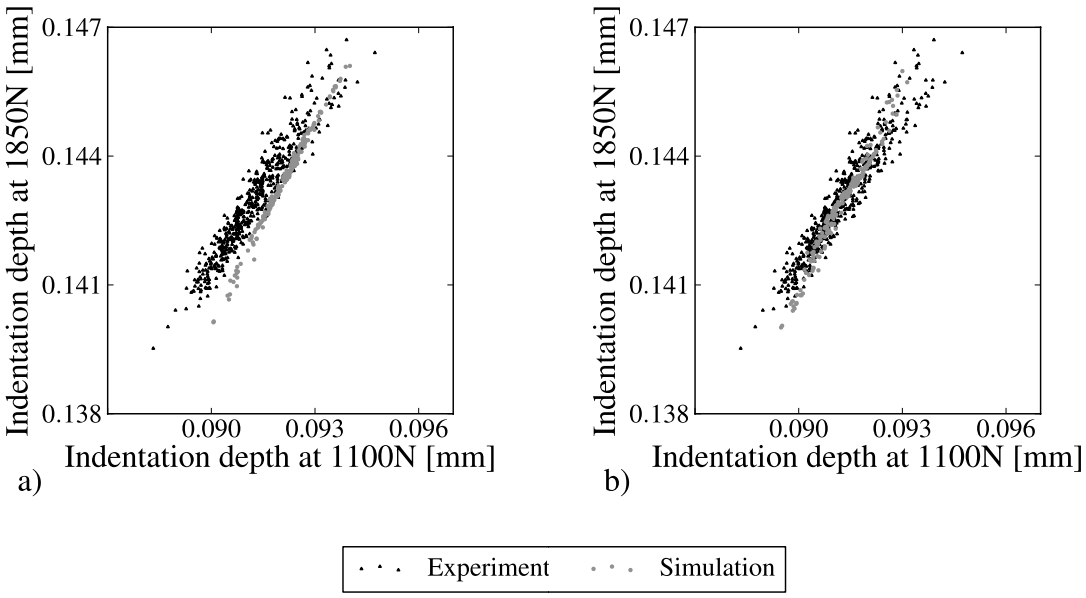


Figure D.30: The indentation depth at 1100 N versus the indentation depth at 1850 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

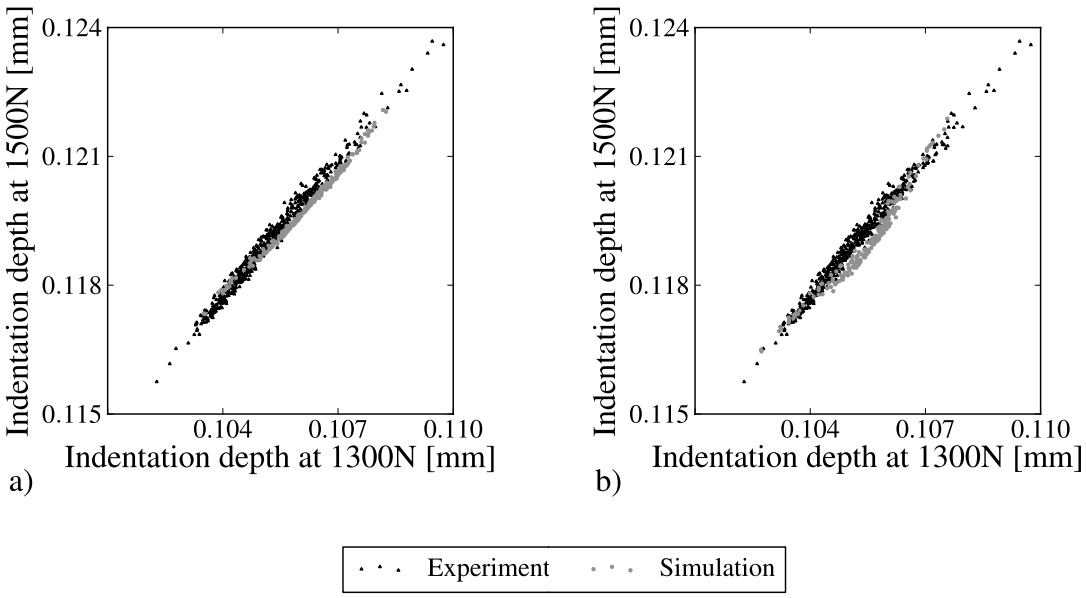


Figure D.31: The indentation depth at 1300 N versus the indentation depth at 1500 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

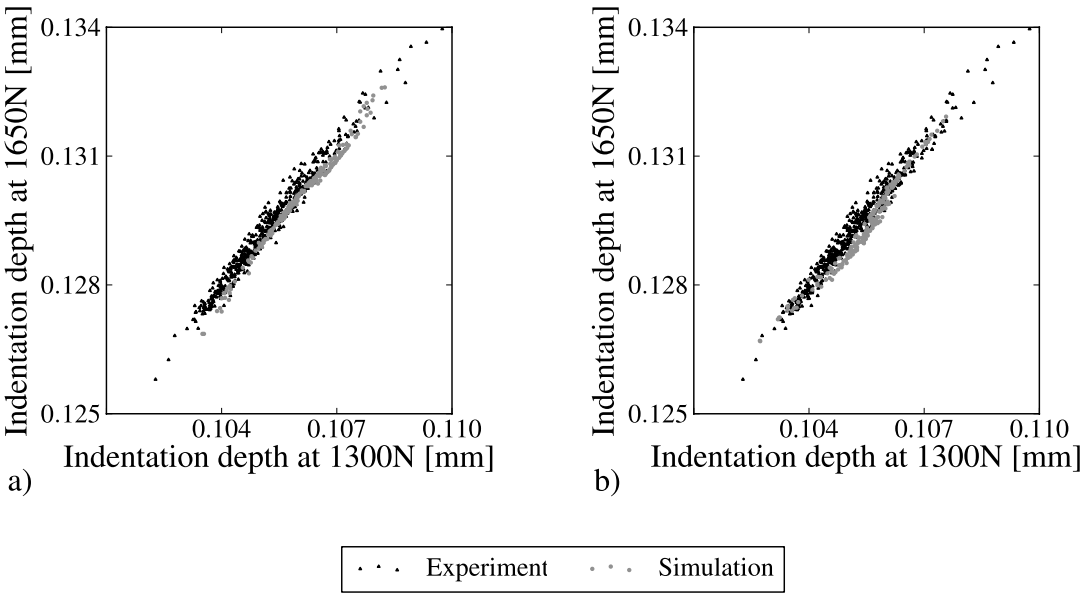


Figure D.32: The indentation depth at 1300 N versus the indentation depth at 1650 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

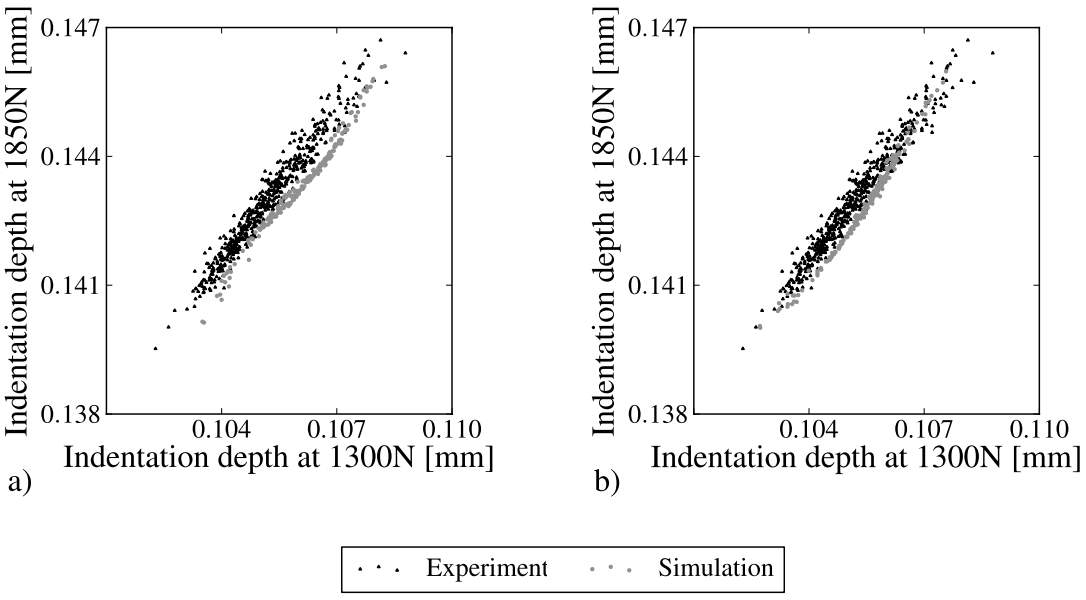


Figure D.33: The indentation depth at 1300 N versus the indentation depth at 1850 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

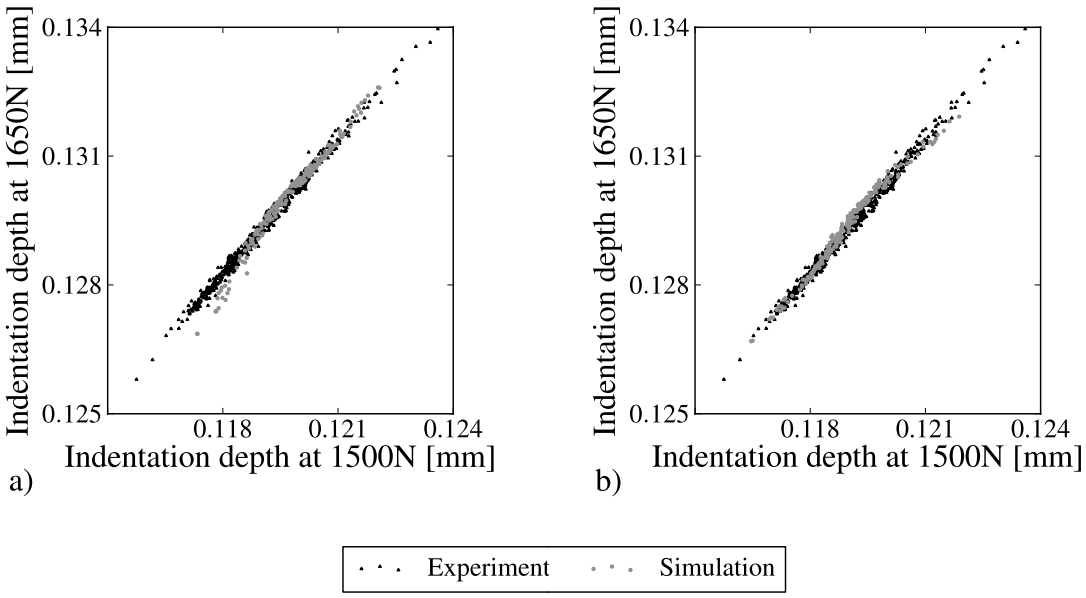


Figure D.34: The indentation depth at 1500 N versus the indentation depth at 1650 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

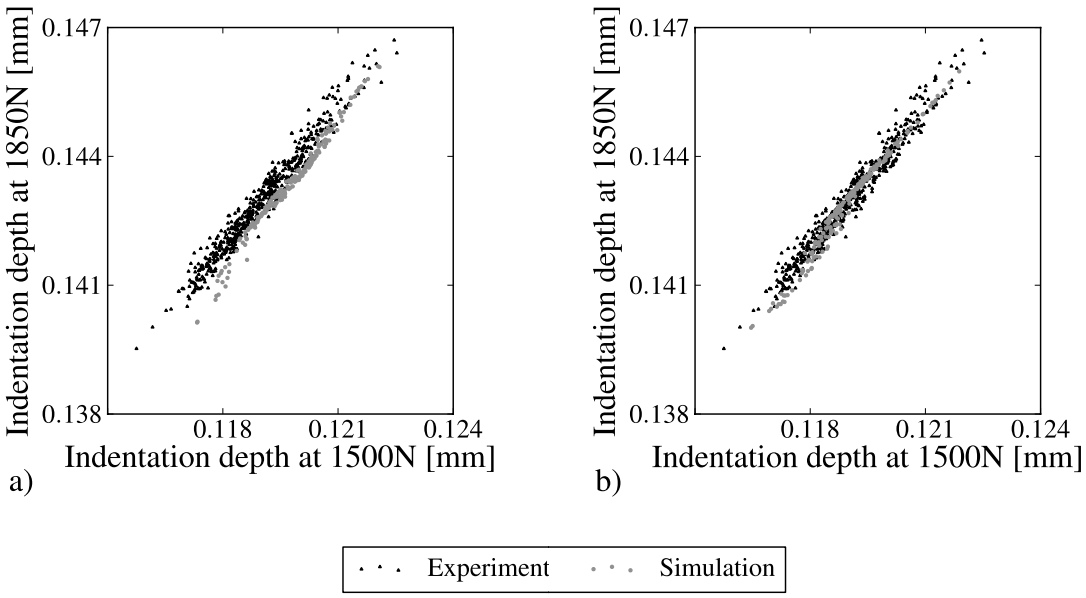


Figure D.35: The indentation depth at 1500 N versus the indentation depth at 1850 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

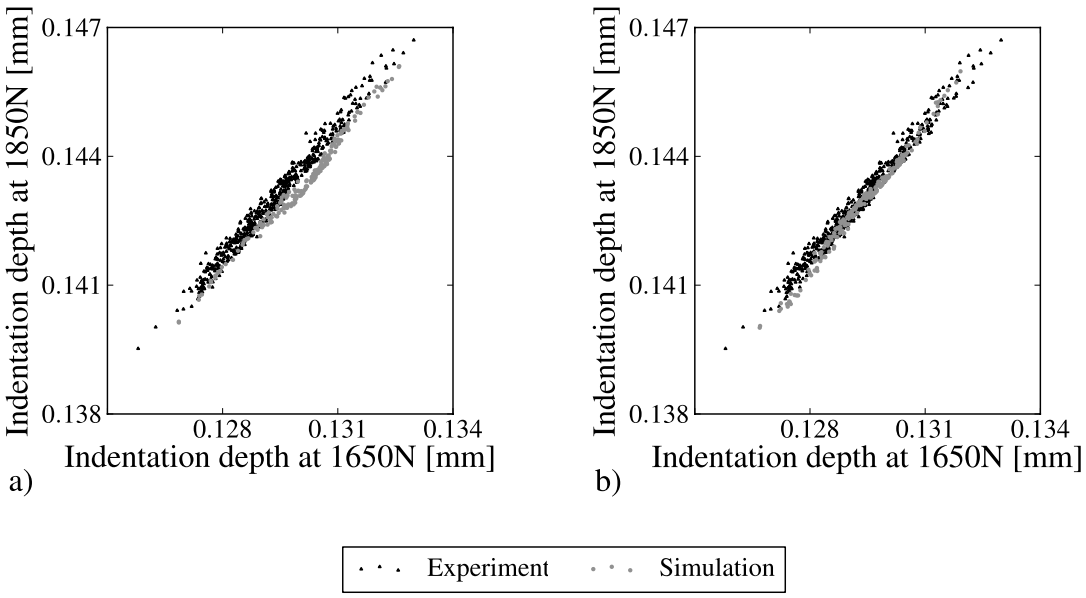


Figure D.36: The indentation depth at 1650 N versus the indentation depth at 1850 N: a) Difference between the 534 experimental results and the 200 numerical results of the original numerical model, b) Difference between the 534 experimental results and the 200 numerical results of the optimised numerical model.

Bibliography

[1] ASME and NAFEMS. What is verification and validation?

[2] ASME. *Guide for Verification and Validation in Computational Solid Mechanics*, ASME V&V 10. The American Society of Mechanical Engineers, 2006.

[3] K.-J. Bathe. *Finite Element Procedures*. Prentice Hall, Pearson Education Inc., 2006.

[4] J. Fish and T. Belytschko. *A First Course in Finite Elements*. Wiley, Chichester, 2009.

[5] J. Marczyk. Validation of fe models using experimental data and monte carlo simulation.

[6] Dod modeling and simulation (m&s) verification, validation, and accreditation (vv&a). Instruction 5000.61, Department of Defense, 2009.

[7] M. Fortier, A. Karl, P. Koch, R. Nayak, R. Rebba, and S.Tsai. *Stochastics and its Role in Robust Design*. NAFEMS, 2012.

[8] W.L. Oberkampf and C.J. Roy. *Verification and Validation in Scientific Computing*. Cambridge University Press, Cambridge New York, 2010.

[9] Documentation of verification, validation, and accreditation (vv&a) for models and simulations. Technical Report MIL-STD-3022, Department of Defense, 2008.

[10] AIAA. *Guilde for the Verification and Validation of Computational Fluid Dynamics Simulation*, AIAA G-077-1998. American Institute of Aeronautics and Astronautics, 1998.

[11] T.H. Naylor and J.M Finger. Verification of computer simulation models. *Management Science*, 14(2):92–101, 1967.

[12] U.B. Mehta. Credible computational fluid dynamics simulations. *AIAA Journal*, 36-5:665–667, 1998.

[13] J. Marczyk. Recent trends in fem.

[14] R. Rebba, S. Huang, Y. Liu, and S. Mahadevan. Statistical validation of simulation models. *International Journal Materials and Product Technologie*, x(x):xxx–xxx, 200x.

[15] B.W. Silverman. *Density Estimation for Statisticas and Data Analysis (Monograph on Statistics and Applied Probability 26)*. Chapman & hall, London New York, 1986.

[16] P.A. Tukey and J.W. Tukey. Data-driven view selection; agglomeration and sharpening. In V. Barnett, editor, *Interpreting Multivariate Data*. Wiley, Chidester, 1981.

[17] C.J. Roy and W.L. Oberkampf. A complete framework for verification, validation, and uncertainty quantification in scientific computing (invited). *AIAA*, 2010-124, 2010.

[18] I. Doltsinis, F. Rau, and M. Werner. Analysis of random systems. In I. Doltsinis, editor, *Stochastic Analysis of Multivariate Systems in Computational Mechanics and Engineering*, pages 9–160. International Center for Numerical Methods in Engineering, Barcelona, Spain, 1999.

[19] A.H-S. Ang and W.H. Tang. *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering*. Wiley, 2007.

[20] R.E. Walpole. *Probability and Statistics for Engineers and Scientists*. Macmillan Publishing Company, New York, 1985.

[21] M.A. Stephens. Tests based on edf statistics. In R.B. D'Agostino and M.A. Stephens, editors, *Goodness-of-fit techniques*, volume 68 of *Statistics: a Series Textbooks and Monographs*, pages 97–194. Marcel Dekker, New York, 1986.

[22] R.B. D'Agostino. Tests for the normal distribution. In R.B. D'Agostino and M.A. Stephens, editors, *Goodness-of-fit techniques*, volume 68 of *Statistics: a Series Textbooks and Monographs*, pages 367–420. Marcel Dekker, New York, 1986.

[23] S. Ferson, W.L. Oberkampf, and L. Ginzburg. Model validation and predictive capability for the thermal challenge problem. *Computer methods in applied mechanics and engineering*, 197:2408–2430, 2008.

[24] P.W. Mielke and K.J. Berry. *Permutation Methods A Distance Function Approach*. Springer Series in Statistics. Springer, 2007.

[25] S.F. Wojtkiewicz, M.S. Eldred, R.V. Field, A. Urbina, and J.R. Red-Horse. Uncertainty quantification in large computational engineering models. In *AIAA 2001-1455*, 2001.

[26] A.H.C. van Maldegem and C.J. Beekum. *Leerboek fysische meettechniek*. Educatieve Partners Nederland, Houten, 1999.

[27] H.W. Coleman and W.G. Steele. *Experimentation and Uncertainty Analysis for Engineers*. Wiley-Interscience, New York Chichester Weinheim Brisbane Singapore Toronto, 1999.

[28] Din 50125 prüfung metalischer werkstoffe: Zugproben. In *DIN-Taschenbuch: Materialprüfnormen für metalische Werkstoffe 1*, pages 107–114. Beuth Verlag, Berlin Köln, 1990.

[29] Din 50145 prüfung metalischer werkstoffe: Zugversuch. In *DIN-Taschenbuch: Materialprüfnormen für metalische Werkstoffe 1*, pages 128–143. Beuth Verlag, Berlin Köln, 1990.

[30] G. Gottstein. *Physical Foundations of Materials Science*. Springer, Berlin Heidelberg, 2004.

[31] W.C. Young and R.G. Budynas. *Roark´s formulas for stress and strain*. McGraw-Hill, Singapore, 2002.

[32] W.C. Oliver and G.M. Pharr. An improvemed technique for determining hardness and elastic modulus using load and displacement sencing indentation experiments. *Journal of Materials Research*, 7(6):1564–1583, 1992.

[33] DIN EN ISO 14577-1. Instrumentierte eindringprüfung zur bestimmung der härte und anderer werkstoffparameter. 2003.

[34] D. Tabor. The hardness of solids. *Review of Physics in Technology*, 1(3):145–178, 1970.

[35] E.G. Herbert, G.M. Pharr, W.C. Oliver, B.N. Lucas, and J.L. Hay. On the measurement of stress-strain curves by spherical indentation. *Thin Solid Films*, 398-399:331–335, 2001.

[36] S. Pathak, D. Stojakovic, R. Doherty, and S.R. Kalidindi. Importance of surface preparation on the nano-indentation stress-strain curves measured in metals. *Journal of Materials Research*, 24(3):1142–1155, 2009.

[37] I. Peyrot, P.-O. Bouchard, R. Ghisleni, and J. Michler. Determination of plastic properties of metals by instrumented indentation using a stochastic optimization algorithm. *Journal of Material Research*, 24(3):936–947, 2009.

[38] C. Dresbach. *Ermittelung lokaler mechanische Kennwerte mikroelektronischer Drahtkontaktierungen*. PhD thesis, Martin-Luther-Universität Halle-Wittenberg, 2010.

[39] S. Reh. *Zuverlaessigkeitsanalyse komplexer Bauteile mit Hilfe der stochastischen Finite-Elemente-Methode*. PhD thesis, University Karlsruhe, 1993.

[40] R.A. Johnson. *Miller & Freuds Probability & statistics for engineers*. Prentice Hall, London Sydney Toronto Mexico New Delhi Tokyo Singapore Rio De Janeiro, 1994.

[41] O. Ditlevens and H.O. Madsen. *Structural Reliability Methods*. Wiley, Chichester, 2005.

[42] H.O. Madsen, S. Krenk, and N.C. Lind. *Methods of Structural Safety*. Dover Publications, New York, 2006.

[43] J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods*. Methuen's Monograph on Applied Probability and Statistics. Chapman and Hall, London, 1965.

[44] P.-L. Liu and A.D. Kiureghian. Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, 1(2):105–112, 1986.

[45] M.D. McKay, R.J. Beckman, and W.J. Conover. A comparison of three methods for selecting values on input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[46] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P.Flannery. *Numerical Recipies in C. The Art of Scientific Computing*. Cambridge University Press, Cambridge New York Oakleigh, 1999.

[47] D.C. Lay. *Linear Algebra and Its Applications*. Adison-Wesley World Student Series. Pearson Education, Boston San Francisco New York London Toronto Tokyo Singapore Madrid Mexico City Paris Cape Town Hong Kong Montreal, 2003.

[48] D. Shen and Z. Lu. Computation of correlation coefficient and its confidence interval in sas. In *SUGI 31 Proceedings: Paper 170-31*, 2006.

[49] R.D. Blevins. *Formulas for natural frequency and mode shape*. Van Nostrand Reinhold Company, New York, 1979.

[50] R.C. Hibbeler. *Sterkteleer voor technici*. Academic Service, Schoonhoven, 2000.

[51] D.W. Scott. *Multivariate density estimation: theory, practice, and visualization*. Wiley series in probability and mathematical statistics. Wiley-Interscience, New York Chichester Brisbane Toronto Singapore, 1992.

[52] E. Schechtman and M. Sherman. The two sample t-test witha known ratio of variaces. *Statistical Methodology*, 4:508–514, 2007.

[53] E.M.M. Manders, F.J. Verbeek, and J.A. Aten. Measurement of co-localization of objects in dual-colour confocal images. *Journal of Microscopy*, 169(3):375–382, 1993.

[54] T. Collins. Colocalisation. McMaster Biophotonics Facility, Dept. Biochemistry and Biomedical Sciences.

[55] V. Zinchuk, O. Zinchuk, and T. Okada. Quantitative colocalization analysis of multicolor confocal immunofluorescence microscopy images: Pushing pixels to explore biological phenomena. *Japan Society Histochemistry and Cytochemistry*, 40(4):101–111, 2007.

[56] F.P. Cordelieres and S. Bolte. Jacop v2.0: improving the user experience with co-localization studies.

[57] F. Mosteller, R.E.K. Rourke, and G.B. Thomas. *Probability and Statistics*. Addison Wesley Publishing Company, Reading, Massachusetts Palo Alto London Don Mills Ontario, 1961.

[58] M.H. Rheinfurth and L.W. Howell. Probability and statistics in aerospace engineering. Technical Report NASA / TP-1998-207194, NASA, 1998.

[59] I.-K. Yeo and R. Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959, 2000.

[60] S. Weisberg. Yeo-johnson power transformations. www.stat.umn.edu/arc/yjpower.pdf.

[61] G.E.P. Box and D.R. Cox. Analysis of transformations. *Journal or the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252, 1964.

[62] S. Thompson. On the distribution of type ii errors in hypothesis testing. *Applied Mathematics*, 2:189–195, 2011.

[63] M.G. Natrella. *Experimental Statistics*, volume 91 of *NBS Handbook*. National Bureau of Standards, Washington, 1966.

[64] A. Justel, D. Pena, and R. Zamar. A multivariate kolmogorov-smirnov test of goodness of fit. *Statistics & Probability Letters*, 35:251–259, 1997.

[65] E.W. Noreen. *Computer Intensive Methods for Testing Hypotheses: An Introduction*. Wiley-Interscience Publication, New York, 1989.

[66] M.J. Anderson. A new method for non-parametric multivariate analysis. *Austral Ecology*, 26:32–46, 2001.

[67] Dynardo Dynamic Software and Engineering GmbH. Optislang documentation. Version 3.2.1.

[68] J. Schijve. *Fatigue of Structures and Materials*. Kluwer Academic Publishers, Dordrecht, 2001.

[69] R.C. Reed. *The superalloys Fundamentals and Applications*. Cambridge University Press, Cambridge New York Melbourne Madrid Cape Town Singapore Sao Paulo, 2008.

[70] S.P. Millard and N.K. Neerchal. *Environmental Statistics with S-Plus*. CRC applied environmental statistics series. CRC Press, 2000.

[71] C.E. Land. An evaluation of approximate confidence interval estimation methods for lognormal means. *Technometrics*, 14(1):145–158, 1972.

[72] T.B. Parkin, S.T. Chester, and J.A. Robinson. Calculating confidence intervals for the mean of a lognormally distributed variable. *Soil Science Society of America Journal*, 54:321–326, 1990.

[73] J.H. Steiger. Beyond the f test: Effect size confidence intervals and test of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, 9(2):164–182, 2004.

[74] H.-J. Bargel and G. Schulze. *Werkstoffkunde*. Springer, Berlin Heidelberg, 2000.

[75] M. Riehle and E. Simmchen. *Grundlagen der Werkstofftechnik*. Deutscher Verlag fuer Grundstoffindustrie, Stuttgart, 2000.

[76] J.C. Wyant. White light interferometry. In *Proceedings of SPIE*, volume 4737, 2002.

[77] P. Hariharan. *Basics of Interferometry*. Academic Service, Burlington San Diego, 2007.

[78] National Research Company. Properties of diamond. Retrieved from http://www. nationalresearchcompany.com/ technical-data.

[79] B. Clausen, M.B. Prime, S. Kabra, D.W. Brown, P. Pagliaro, P. Blacklund, S. Shah, and E. Criss. Residual stress and plastic anisotropy in indented 2024-t351 aluminium disks. In *Proceedings of the SEM Annual Conference*, Albuquerque New Mexico, 2009.

[80] Din 1319 teil3 grundbegriffe der meßtechnik: Begriffe für die meßunsicherheit und für die beurteilung von meßgeräten und meßeinrichtungen. In *DIN-Taschenbuch: Materialprüfnormen für metalische Werkstoffe 1*, pages 13–26. Beuth Verlag, 1990.

[81] ANSYS. *ANSYS Mechanical User's Guide*. Version 14.0, 2013.

[82] ANSYS. *ANSYS Mechanical APDL Command Reference*. Version 14.0, 2013.

[83] M. Mata, O. Casals, and J. Alcal*á*. The plastic zone size in indentation experiments: The analogy with the expansion of a spherical cavity. *International Journal of Solids and Structures*, 43:5994–6013, 2006.

[84] ANSYS. *ANSYS Structural Analyis Guide*. Version 14.0, 2013.

[85] R. Hill. *The Mathematical Theory of Plasticity*. Oxford Classic Texts in the Physical Science. Oxford university press, Oxford New York, 1998.

[86] K.L. Johnson. *Contact Mechanics*. Cambridge University Press, Cambridge, 1999.

[87] A.C. Fischer-Cripps. *Nanoindentation*. Springer, New York Dordrecht Heidelberg London, 2011.

[88] H.J. Bunge, R. Kiewel, Th. Reinert, and L. Fritsche. Elastic properties of polycrystals - influence of texture and stereology. *Journal of the Mechanics and Physics of Solids*, 48:29–66, 2000.

[89] R.J. Young and P.A. Lovell. *Introduction to polymers*. CRC Press, Boca Raton London New York Washington D.C., 1991.

Summary

During product development, engineers and scientist use a variety of numerical models to determine how assemblies, components or parts behave. To describe the behaviour of these products, multiple performance relevant quantities are used such as weight, stiffness, lifetime, efficiency and energy consumption. These performance quantities can be estimated using experiments or numerical models. However, it is often not possible to obtain all performance quantities using one approach. Therefore, it is necessary to combine the experimental and the numerical approach to obtain the performance quantities of interest. Consequently, it is necessary to ensure that both approaches represent reality accurately. To determine whether the numerical results differ from reality, validation is performed. During validation it is determined whether the distance between the numerical results and reality is significant, where reality is represented by the experimental results.

In context of validation, it is necessary to point out that experimental results scatter due to variations in the production process and the operation conditions to which the product is subjected. Therefore, it would be appropriate to compare the stochastic experimental results to stochastic numerical results. Since these results are generally multivariate and arbitrarily distributed, it is required to use a validation method that is suitable for arbitrarily distributed multivariate results. However, currently no validation method exists to compare such results.

In this work, a method is developed to validate numerical models using arbitrarily distributed multivariate results. To quantify the shape difference between the samples without assuming a distribution function, the underlying distributions of the experimental and numerical results are estimated based on the results. Furthermore, the measurement uncertainties and the numerical uncertainties can be used explicitly in the distance measure for additional information. To determine whether the numerical model is valid, it is determined if the numerical model is significantly different from the experimental results using a hypothesis test. Furthermore, it is determined whether the distance between the numerical results and the experimental results is larger than the uncertainties present in the numerical and experimental results.

To investigate if the developed validation method is more effective than the typically used methods for multivariate problems, benchmark tests are performed. Since these distance measures were developed for multivariate normally distributed data, the benchmarks are performed using normally distributed data and a test data set that represents non-normally distributed data. Using these benchmark tests, it is shown that the developed method is more effective to determine the distance between the numerical and the experimental results than the typically used distance measures. Furthermore, it is demonstrated that it is meaningful to use the developed validation method for engineering problems by an example of a spherical indentation model.

Using the developed method, it is now possible to validate stochastic numerical models without assuming distribution functions for the experimental and numerical results. It is also possible to incorporate the measurement uncertainties and the simulation uncertainties in the distance measure of this method.

Kurzfassung

Während der Produktentwicklung verwenden Ingenieure und Wissenschaftler eine Vielfalt an numerischen Modellen um vorherzusagen, wie sich Systeme, Baugruppen und Bauteile verhalten. Um das Verhalten dieser Produkte zu beschreiben werden mehrere relevante Eigenschaften verwendet, zum Beispiel: Masse, Steifigkeit, Lebensdauer, Effizienz und Energieverbrauch. Diese Eigenschaften können numerisch oder experimentell bestimmt werden. Es ist aber meistens nicht möglich alle relevanten Eigenschaften mittels einer dieser Methoden zu bestimmen. Deswegen ist es notwendig, die numerische und die experimentelle Methode zu kombinieren um die relevanten Eigenschaften zu bestimmen. Hierfür müssen Experiment und Simulation die Realität gut abbilden. Um festzustellen, ob das numerische Modell der Realität entspricht, wird eine Validierung durchgeführt. Hierbei werden die numerischen Ergebnisse mit der Realität verglichen, die durch die experimentellen Ergebnisse dargestellt wird.
Im Zusammenhang mit Validierung ist es notwendig zu erwähnen, dass die experimentellen Ergebnisse durch Variationen in der Produktion und in den Betriebsbedingungen streuen. Da die experimentellen Ergebnisse schwanken, ist es sinnvoll, stochastische experimentelle Ergebnisse mit stochastischen numerischen Ergebnissen zu vergleichen. Hinzu kommt, dass die experimentellen und die numerischen Ergebnisse im Allgemeinen multivariat und beliebig verteilt sind, deswegen muss die Validierungsmethode für multivariate beliebig verteilte Ergebnisse geeignet sein. Derzeit gibt es keine Methode für diesen allgemeinen Fall.
In der vorliegenden Arbeit wurde eine Methode entwickelt, welche numerische Modelle mit Hilfe von beliebig verteilten, multivariaten Ergebnissen validieren kann. Um den Unterschied zwischen den numerischen und den experimentellen Ergebnissen mittels eines Abstandsmaßes zu quantifizieren, werden die jeweiligen Verteilungen, basierend auf den vorhandenen Ergebnissen, mit geeigneten Methoden geschätzt. Hierbei ist es möglich, die Messungenauigkeit und die Unsicherheit in der Simulation explizit bei der Bestimmung des Abstands zwischen Experiment und Simulation zu berücksichtigen. Um festzustellen ob das numerische Modell valide ist, wird mit Hilfe eines Hypothese-Tests bestimmt, ob der Abstand zwischen Simulation und Experiment signifikant ist und ob der Abstand zwischen den numerischen und den experimentellen Ergebnissen größer ist als die Unsicherheiten der jeweiligen Ergebnisse.
Um zu untersuchen, ob die neu entwickelte Validierungsmethode effektiver ist als die typischerweise verwendeten Methoden für mehrdimensionale Probleme, wurden Benchmark-Tests durchgeführt. Da diese typischen Methoden für normalverteilte Größen entwickelt wurden, werden die Benchmark-Tests an normalverteilten Daten und an einem Datensatz, der nicht normalverteilt ist, durchgeführt. Mittels dieser Benchmark-Tests wird gezeigt, dass die entwickelte Methode in einigen Fällen den Abstand zwischen den numerischen und experimentellen Ergebnissen besser bestimmen kann als die typischerweise verwendeten Abstandsmaße. Darüber hinaus wird am Beispiel eines Kugeleindruckmodells gezeigt, dass die neu entwickelte Validierungsmethode im Rahmen von Ingenieurfragestellungen sinnvoll angewendet werden kann.
Mittels der neu entwickelten Validierungsmethode ist es jetzt möglich, stochastische numerische Modelle mit Hilfe von beliebig verteilten, mehrdimensionalen, experimentellen und numerischen Ergebnissen zu validieren. Mit dieser Methode ist es auch möglich die Messunsicherheit und die Simulationsunsicherheit bei der Bestimmung des Abstands zwischen Modell und Experiment direkt zu berücksichtigen.